

# Selective Constraints Determine the Time Dependency of Molecular Rates for Human Nuclear Genomes

Sankar Subramanian\* and David M. Lambert\*

Environmental Futures Centre, Griffith University, Nathan, Queensland, Australia

\*Corresponding authors: E-mail: s.subramanian@griffith.edu.au; d.lambert@griffith.edu.au.

Accepted: October 6, 2012

## Abstract

In contrast to molecular rates for neutral mitochondrial sequences, rates for constrained sites (including nonsynonymous sites, D-loop, and RNA) in the mitochondrial genome are known to vary with the time frame used for their estimation. Here, we examined this issue for the nuclear genomes using single-nucleotide polymorphisms (SNPs) from six complete human genomes of individuals belonging to different populations. We observed a strong time-dependent distribution of nonsynonymous SNPs (nSNPs) in highly constrained genes. Typically, the proportion of young nSNPs specific to a single population was found to be up to three times higher than that of the ancient nSNPs shared between diverse human populations. In contrast, this trend disappeared, and a uniform distribution of young and old nSNPs was observed in genes under relaxed selective constraints. This suggests that because mutations in constrained genes are highly deleterious, they are removed over time, resulting in a relative overabundance of young nSNPs. In contrast, mutations in genes under relaxed constraints are nearly neutral, which leads to similar proportions of young and old SNPs. These results could be useful to researchers aiming to select appropriate genes or genomic regions for estimating evolutionary rates and species or population divergence times.

**Key words:** rates of evolution, natural selection, time dependency, deleterious polymorphisms, population genetic theory.

## Introduction

The rate of molecular evolution is a fundamental parameter in genetics and evolutionary biology. Although rates that are estimated using different timescales and methods are expected to be similar, recently studies suggested otherwise. For example, studies based on pedigree analyses have recorded higher rates of evolution compared with those estimated using phylogenetics methods (Parsons et al. 1997; Howell et al. 2003; Denver et al. 2004; Haag-Liautard et al. 2007; Millar et al. 2008). Comparative genomic studies have shown that molecular rates estimated using short timescales, for instance, those based on intraspecific data, are much higher than those estimated using interspecies data (Garcia-Moreno 2004; Ho et al. 2005, 2007; BurrIDGE et al. 2008; Subramanian and Lambert 2011). A number of reasons have been suggested for this discrepancy. However, most of these involve biases or errors in estimation such as calibration errors, saturation effects on nucleotide positions, and phylogenetic and demographic model misspecification (Emerson 2007; Bandelt 2008; Debruyne and Poinar 2009; Henn et al. 2009; Ho et al. 2011). The higher intraspecific rates have also been attributed to artifacts such as sequencing errors,

postmortem damage (in the case of ancient DNA), and ascertainment bias (reviewed by Ho et al. [2011]).

The major biological factor that produces a time-dependent pattern of molecular rates appears to be purifying selection (Endicott and Ho 2008; Subramanian et al. 2009; Ho et al. 2011; Subramanian and Lambert 2011). At short timescales, slightly deleterious polymorphisms will segregate in populations, and this standing variation results in a higher diversity. This is evident from a number of recent studies that have shown an overabundance of low-frequency nonsynonymous single-nucleotide polymorphism (nSNPs) in human protein-coding genes (Li et al. 2010; Nelson et al. 2012; Subramanian 2012; Tennessen et al. 2012). Thus, an elevated molecular rate is observed for the data from within a population or for closely related populations (Kimura 1983). In contrast, deleterious polymorphisms are removed over time, which results in reduced divergence between species or distantly related populations.

Note that throughout this article, the rate of evolution refers to the observed or estimated divergence divided by calibration time. It does not denote a mutation rate, and hence, we do not intend to suggest that mutation rate varies with time.

Time-dependent rates have been typically reported for the D-loop region of the mitochondrial genome (Parsons et al. 1997; Lambert et al. 2002; Howell et al. 2003; Ho et al. 2005; Hay et al. 2008; Henn et al. 2009). Because the D-loop is hypervariable, the saturation effects on nucleotide sites confound the effects of purifying selection in this region. Hence, later studies (Endicott and Ho 2008; Subramanian et al. 2009; Subramanian and Lambert 2011) used coding genes of mitochondria to examine the time-dependency effect, because these are less prone to estimation errors. These studies showed that rates of molecular evolution at synonymous sites of protein-coding genes are similar across different timescales. In contrast, these studies found a strong time-dependent rate of evolution at nonsynonymous sites. Because synonymous sites are free from selection and nonsynonymous positions are under selective constraints, the observed pattern clearly points to the effects of purifying selection as predicted theoretically (Kimura 1983). This is also evident from the time-dependent decline in the ratio of divergences/diversities at nonsynonymous and synonymous sites ( $dN/dS$  or  $pN/pS$ ) for protein-coding genes of human mitochondrial genes (Subramanian 2009) and nuclear genes of virus (Holmes 2003) and bacteria (Rocha et al. 2006).

Studies on the nuclear genomes of vertebrates did not reveal a clear-cut pattern of time dependency. Although earlier studies suggested a time-dependent pattern in microsatellites (Zivotovsky et al. 2004, 2006) of the human nuclear genome, later studies based on next-generation sequencing did not identify such a pattern (Xue et al. 2009; Sun et al. 2012). This was further confirmed by a number of genome-wide studies, which showed that the evolutionary rates estimated using complete nuclear genomes of human pedigrees were similar to those estimated using human–chimpanzee species comparison (Conrad et al. 2010; Roach et al. 2010; Kong et al. 2012). However, as discussed earlier, time-dependent variation is expected only for constrained genomic regions. Therefore, in this study, we examined protein-coding genes in the human nuclear genome. It is well known that different nuclear genes are under different levels of selective constraint, depending on the relative importance of their functions. Hence, it would be interesting to examine the time-dependent pattern of evolutionary rates in genes, under various magnitudes of selection pressure. Therefore, we grouped nuclear genes of the human genome into four categories based on the intensities of selection pressure on them. We examined the pattern of time dependency using synonymous SNPs (sSNPs) and nSNPs from the complete genomes of six humans belonging to different populations.

## Materials and Methods

### DNA Sequence and Polymorphism Data

Protein-coding sequence alignments for humans (build 36) and chimpanzee (build 2) were obtained (for 16,750 known

genes) from the University of California–Santa Cruz (UCSC) genome bioinformatics (<http://genome.ucsc.edu>). Polymorphism data from six complete genomes belonging to a Khoisan (Schuster et al. 2010), two Yorubans (Bentley et al. 2008; Schuster et al. 2010), a European (Levy et al. 2007), a Chinese (Wang et al. 2008), and a Korean person (Kim et al. 2009) were obtained from UCSC genome bioinformatics and PSU Bioinformatics (<http://main.genome-browser.bx.psu.edu/>) data repositories. sSNPs and nSNPs of these genomes were determined using the chromosomal coordinates and gene boundary information. The number of sSNPs and nSNPs from each genome is given in table 1. The reference human genome was used to determine the ancestral state of each SNP. However, using the chimpanzee genome to orient the direction of SNP produced similar results (data not shown).

### Divergence and Diversity Estimation

The numbers of synonymous and nonsynonymous positions and substitutions in each gene were calculated using the codeml program of the software PAML (Yang 2007). Pairwise evolutionary distances at synonymous and nonsynonymous sites for the human–chimpanzee pair were estimated using the Jukes–Cantor method. sSNPs and nSNPs were grouped based on the pattern of sharing, as described in the results (Subramanian 2012). To estimate the proportion of differences ( $pS$  or  $pN$ ), the number of sSNPs or nSNPs belonging to each group was divided by the total number of synonymous or nonsynonymous positions in the genome. The binomial variance was used to compute the standard error.

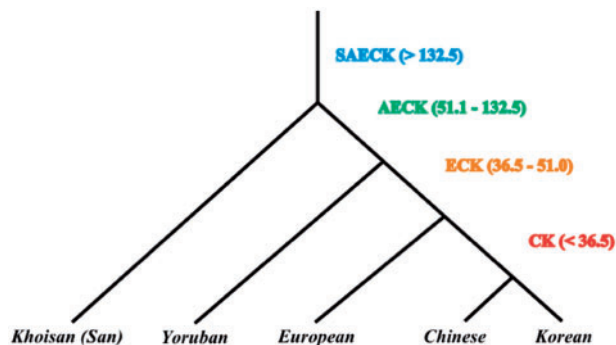
## Results and Discussion

To examine nucleotide diversity at various temporal or evolutionary depths, we obtained SNP data from five complete human genomes belonging to a Chinese, a Korean, a European, a Yoruban (West African), and a Khoisan, an ancient African lineage. The phylogenetic relationship between the genomes of these humans (Tishkoff et al. 2009) is shown in figure 1. sSNPs and nSNPs were grouped based on the pattern of sharing between these genomes. If an SNP is shared between the Khoisan and any other genome, it was considered to be the oldest (see SAECK in fig. 1). If an SNP is shared between European (or an Asian) and Yoruban genomes and not shared by the Khoisan, it was considered to be common to Yorubans (or Africans) and non-Africans (AECK). Similarly, if an SNP is shared between a European and an Asian but not shared by Yoruban or Khoisan, it was considered to be ancestral to Eurasians (ECK). Finally, if an SNP is present in the Chinese or Korean genomes and not shared with any other genome, it was considered to be specific to Asians (CK). These grouping assumed that convergent or parallel mutations are rare. A recent study estimated the population divergence times of 4.5 (1–8), 36.5 (26–47), 51.0 (38–64), and 132.5 (108–157) Kyr for Chinese–Korean, Asian–European,

**Table 1**

Number of SNPs and Estimates of Ratios for each Branch of the Human Tree

Branch on the Tree	Nonsynonymous SNPs (A)	Synonymous SNPs (S)	A/S	pN/pS (SE)
<b>dN/dS = 0–0.1</b>				
SAECK	423	2,309	0.18	0.067 (0.0035)
AECK	137	534	0.26	0.093 (0.0089)
ECK	122	371	0.33	0.120 (0.0125)
CK	558	1,049	0.53	0.194 (0.0101)
<b>dN/dS = 0.1–0.2</b>				
SAECK	721	1,552	0.46	0.169 (0.0076)
AECK	168	327	0.51	0.187 (0.0177)
ECK	169	235	0.72	0.262 (0.0264)
CK	571	702	0.81	0.296 (0.0167)
<b>dN/dS = 0.2–0.6</b>				
SAECK	2,406	2,695	0.89	0.325 (0.0091)
AECK	626	636	0.98	0.358 (0.0202)
ECK	461	434	1.06	0.387 (0.0259)
CK	1,571	1,188	1.32	0.481 (0.0185)
<b>dN/dS = 0.6–1.0</b>				
SAECK	875	571	1.53	0.558 (0.0300)
AECK	244	165	1.48	0.538 (0.0542)
ECK	160	96	1.67	0.607 (0.0783)
CK	506	286	1.77	0.644 (0.0476)



**Fig. 1.**—Population phylogeny of the human genomes. The sharing pattern of SNPs used in this study is illustrated. Abbreviation denotes SNPs specific to Asians (CK), those shared between Europeans and Asians (ECK), Yorubans and Eurasians (AECK), and Khoisan and other genomes (SAECK). The relative age of SNPs (Kyr) obtained using the population divergence times estimated by a previous study (Gronau et al. 2011) are given in parenthesis.

African–Eurasian, and Khoisan–other humans, respectively (Gronau et al. 2011). Hence, these times provide relative ages for the SNPs as <36.5, 36.5–51.0, 51.1–132.5, and >132.5 for CK, ECK, AECK, and SAECK, respectively.

We then grouped human genes based on the level of selective constraint on them. To quantify this, we used dN/dS ratio estimated from the human–chimp divergence at nonsynonymous- (dN) and synonymous sites (dS). Because synonymous sites are free from selection, this ratio reveals the

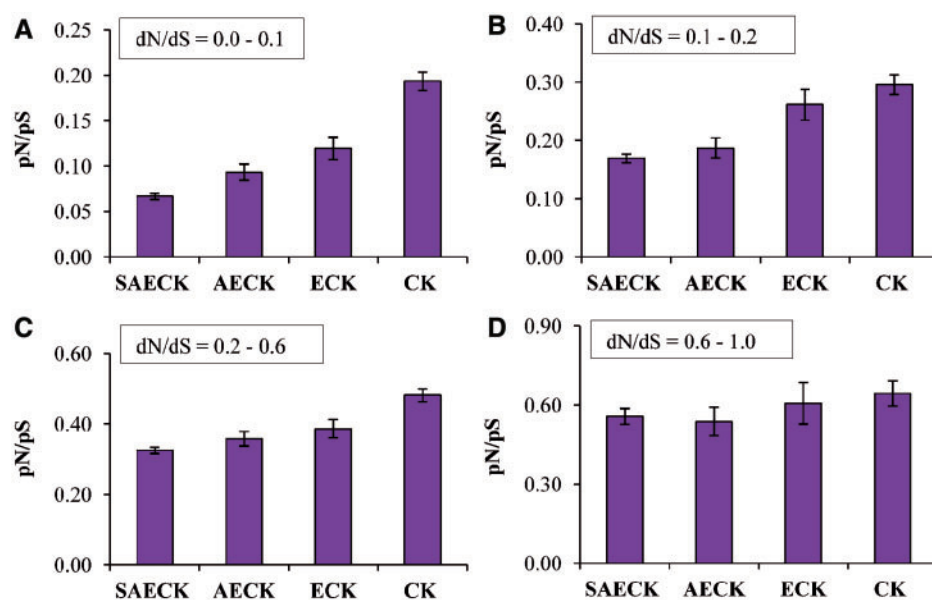
extent of selective constraint on amino acids. For each group of genes, we estimated the ratio (pN/pS) of nSNPs (pN) to sSNPs (pS) per site using the polymorphisms shared between different branches of the human population tree (table 1). First, we examined the distribution of SNPs in the highly constrained human genes. This revealed a clear negative relationship between the age of SNPs (based on the extent of sharing) and pN/pS ratios (fig. 2A). For example, the pN/pS ratio of the Asian-specific SNPs (CK) was 2.9 times higher than that estimated using the SNPs shared with the ancient Khoisan genome (SAECK). This is a perfect example for time dependency, as the latter (>132.5 Kyr) is more than three times older than the former (<36.5 Kyr).

Interestingly, the distribution of pN/pS of genes gradually changes with decreasing selection pressures (fig. 2A–D). The distribution observed for highly constrained genes was strongly right skewed (fig. 2A). In contrast, it is largely uniform for the genes under relaxed selective constraints (fig. 2D) where the difference between pN/pS ratios of young and old SNPs disappeared (Z test,  $P > 0.12$ ).

In addition to using dN/dS ratios, we also grouped human genes based on a different measure namely the Genome Evolutionary Rate Profiling (GERP) score described previously (Cooper et al. 2005). A GERP score is determined using multiple sequence alignments of orthologous genes and the level of selective constraints on each position. The score is quantified using the deficit of substitutions compared with a neutral site. We estimated the average GERP score for each gene and grouped genes based on their mean GERP scores. The results based on average GERP score (supplementary fig. S1, Supplementary Material online) were very similar to those showed in figure 2. For instance, constrained genes with a mean GERP score > 3.5 showed a clear-cut pattern of time dependency, which was absent for genes under relaxed selective constraints (average GERP score < 0.5).

Theoretical studies suggested that deleterious mutations contribute to the diversity of a population, but they are prevented from reaching higher frequencies and are purged over time (Kimura 1983). The results of this study strongly support this prediction. Because mutations in constrained genes are highly deleterious in nature, they do not typically spread through populations. As they are removed over time, we observe a very strong time-dependent effect on these genes (fig. 2A). In contrast, mutations on genes under relaxed selective constraints are typically only mildly harmful to humans and, therefore, segregate in human population for a long time. Particularly mutations in the most rapidly evolving genes with a dN/dS > 0.6 (fig. 2D) are almost neutral with negligible effects on the fitness of humans. This results in the similarity of pN/pS ratios across various timescales.

Previous studies suggested a population bottleneck in European and Asian populations (Marth et al. 2004; Li and Durbin 2011), and thus, a much smaller effective population size is expected for non-African populations than Yoruban



**Fig. 2.**—The ratio of nonsynonymous- to synonymous SNPs per site ( $pN/pS$ ) estimated for each of the branches shown in figure 1. The results shown are using the genes with  $dN/dS$  ratio (estimated for the human–chimpanzee orthologous pair) of (A)  $<0.1$ , (B)  $0.1–0.2$ , (C)  $0.2–0.6$ , and (D)  $>0.6$ . Error bars are the standard error of the mean. In (D), the differences in  $pN/pS$  ratios between any two categories were not statistically significant (Z test,  $P > 0.12$ ).

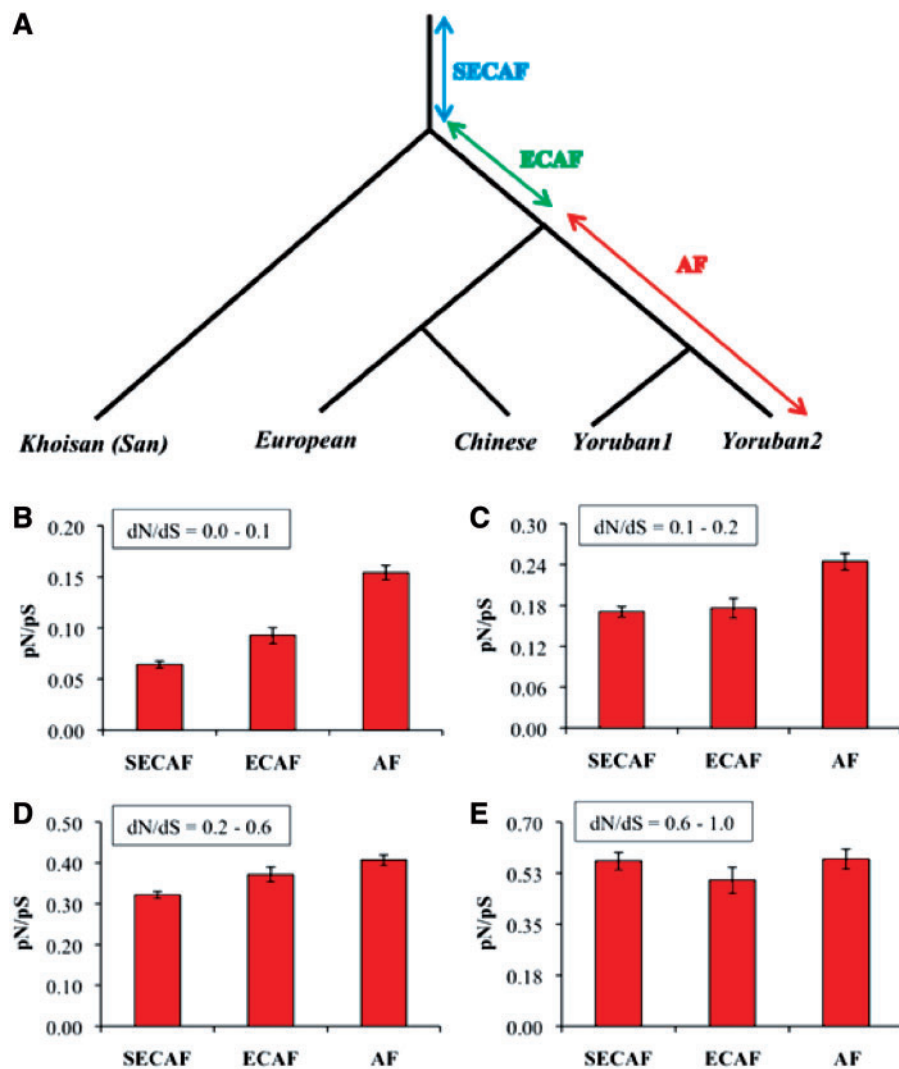
and Khoisan populations. Hence, this bottleneck effect might have drifted some deleterious SNPs to higher frequencies, and this might have some influence on the pattern observed for constrained genes (fig. 2A). To address this issue, we examined the pattern of nSNPs in the African lineage, which is not bottlenecked in the past (Gronau et al. 2011; Li and Durbin 2011). For this purpose, we included another Yoruban genome and examined  $pN/pS$  ratios for the branches of the human population tree constructed using the genomes from two Yorubans, a European, a Chinese, and a Khoisan (fig. 3). For constrained human genes ( $dN/dS < 0.1$ ), we found a significantly higher  $pN/pS$  ratio for Yoruban-specific SNPs (0.15) than that estimated using the SNPs shared between Yorubans and Eurasians (0.09), and this is higher than that computed for those shared between the ancient Khoisan and other genomes (0.06) (fig. 3B). In contrast, for genes under weak purifying selection ( $dN/dS > 0.6$ ),  $pN/pS$  ratios of old and young SNPs were similar ( $P = 0.44$ ) (fig. 3E). Figure 3B clearly supports a time-dependent pattern of deleterious SNPs rather than any bottleneck effect in the African lineage. Furthermore, a previous study suggested that the common ancestral population of all humans (SECAF) was smaller than the population ancestral to Yorubans (or Africans) and Eurasians (ECAAF), which was in turn smaller than the population of Yorubans (AF) (Gronau et al. 2011). Hence, it seems likely that population expansion occurred throughout the African lineage. If population size is assumed to modulate the  $pN/pS$  ratios, then a pattern opposite to that shown in figure 3 is expected: high  $pN/pS$  for SECAF and low for AF.

Therefore, population size effect might not explain our results shown in figures 2 and 3.

This study demonstrated how purifying selection on genes influences the pattern of time dependency of molecular rates. Previous studies on human mitochondrial genes also suggested enrichment of nonsynonymous polymorphisms in the tips of the human population tree, compared with the internal nodes (Kivisild et al. 2006; Subramanian 2009; Pereira et al. 2011). Furthermore, Pereira et al. (2011) showed a predominance of pathogenic mutations in the younger branches of the human tree. Previous studies based on pedigree analysis of D-loop regions reported much higher rates compared with phylogenetic studies, which suggests the presence of deleterious mutations (Parsons et al. 1997; Howell et al. 2003; Millar et al. 2008). In contrast, pedigree analyses based on complete nuclear genomes (Conrad et al. 2010; Roach et al. 2010; Kong et al. 2012; Sun et al. 2012) of humans showed that the rates obtained from these analyses were not significantly different from the interspecies rate estimated using pseudogene data from human and chimpanzee comparison (Nachman and Crowell 2000). Because the majority ( $>95\%$ ) of the sites of human genome are under neutral evolution, the similarity of the pedigree-based rate with the interspecific rate appears to be due to the predominance of neutral mutations.

In this study, we showed the variations in the rate of evolution observed between different timescales could largely be attributed to purifying selection. However, some of the variations could be due to other biological factors such as changes in population sizes. For instance, if the population size ( $N_e$ ) of





**Fig. 3.**—(A) Phylogeny of human genomes belonging to different populations. Abbreviations are SNPs specific to Yorubans (AF), those shared between Yorubans and Eurasians (ECAF), and Khoisan and other genomes (SECAF).  $pN/pS$  ratios were estimated for three branches at different depths of the tree. Human genes were grouped into four categories mentioned in figure 2. Error bars denote the standard error of the mean. In (E), the differences in  $pN/pS$  ratios between any two categories were not statistically significant ( $Z$  test,  $P > 0.20$ ).

a species has been declining over time, this might mimic the pattern of time dependency as  $pN/pS$  ratios could be modulated by  $N_e$ . However, time dependency has been observed in a large number of species from virus to vertebrates, which suggests that this could be a universal phenomenon. This seems a reasonable view because we do not expect a population decline in the large number of species examined to date. Although multiple substitutions, homoplasy, or back mutations are of biological origin, they might be missed particularly in mutational hotspots because of the lack of methods to identify them. Therefore, long-term divergences are subjected to a higher rate of underestimation than short-term distances, and this might also produce a time-dependent effect. However, we used nuclear genomes of human for which

the evolutionary rate is very low ( $<1$  per site per billion years [Conrad et al. 2010; Roach et al. 2010; Kong et al. 2012; Sun et al. 2012]), and therefore multiple, back, or parallel substitutions are unlikely to influence our results as our comparisons involve only closely related species and populations.

Our results suggest that although the observed rates of molecular evolution vary for different timescales, this variation is limited to genes or genomic regions under selection. Importantly, we showed that the extent of variation is determined by the magnitude of selection on these regions. Therefore, to estimate molecular evolutionary rates or divergence times between species/populations, it is advisable to use the genes and genomic regions that are under minimal selective constraints.

## Supplementary Material

Supplementary figure S1 is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgment

This work was supported by the Australian Research Council and Griffith University.

## Literature Cited

- Bandelt HJ. 2008. Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. *Heredity* 100: 1–2.
- Bentley DR, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- Burridge CP, Craw D, Fletcher D, Waters JM. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. *Mol Biol Evol.* 25:624–633.
- Conrad DF, et al. 2010. Variation in genome-wide mutation rates within and between human families. *Nat Genet.* 43:712–714.
- Cooper GM, et al. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15:901–913.
- Debruyne R, Poinar HN. 2009. Time dependency of molecular rates in ancient DNA data sets, a sampling artifact? *Syst Biol.* 58:348–360.
- Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679–682.
- Emerson BC. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst Biol.* 56:337–345.
- Endicott P, Ho SY. 2008. A Bayesian evaluation of human mitochondrial substitution rates. *Am J Hum Genet.* 82:895–902.
- García-Moreno J. 2004. Is there a universal mtDNA clock for birds? *J Avian Biol.* 35:465–468.
- Gronau I, et al. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet.* 43:1031–1034.
- Haag-Liautard C, et al. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445:82–85.
- Hay JM, et al. 2008. Rapid molecular evolution in a living fossil. *Trends Genet.* 24:106–109.
- Henn BM, Gignoux CR, Feldman MW, Mountain JL. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol.* 26:217–230.
- Ho SY, Kolokotronis SO, Allaby RG. 2007. Elevated substitution rates estimated from ancient DNA sequences. *Biol Lett.* 3:702–705.
- Ho SY, et al. 2011. Time-dependent rates of molecular evolution. *Mol Ecol.* 20:3087–3101.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol.* 22:1561–1568.
- Holmes EC. 2003. Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J Virol.* 77: 11296–11298.
- Howell N, et al. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet.* 72:659–670.
- Kim JI, et al. 2009. A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011–1015.
- Kimura M. 1983. *The neutral theory of molecular evolution.* Cambridge (UK): Cambridge University Press.
- Kivisild T, et al. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
- Kong A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488:471–475.
- Lambert DM, et al. 2002. Rates of evolution in ancient DNA from Adeline penguins. *Science* 295:2270–2273.
- Levy S, et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5:e254.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
- Li Y, et al. 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet.* 42: 969–972.
- Marth GT, Czaparka E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 166:351–372.
- Millar CD, et al. 2008. Mutation and evolutionary rates in Adeline penguins from the Antarctic. *PLoS Genet.* 4:e1000209.
- Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.
- Nelson MR, et al. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337: 100–104.
- Parsons TJ, et al. 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet.* 15:363–368.
- Pereira L, et al. 2011. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. *Am J Hum Genet.* 88: 433–439.
- Roach JC, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–639.
- Rocha EPC, et al. 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol.* 239:226–235.
- Schuster SC, et al. 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463:943–947.
- Subramanian S. 2009. Temporal trails of natural selection in human mitochondrial genomes. *Mol Biol Evol.* 26:715–717.
- Subramanian S. 2012. The abundance of deleterious polymorphisms in humans. *Genetics* 190:1579–1583.
- Subramanian S, et al. 2009. High mitogenomic evolutionary rates and time dependency. *Trends Genet.* 25:482–486.
- Subramanian S, Lambert DM. 2011. Time dependency of molecular evolutionary rates? Yes and no. *Genome Biol Evol.* 3:1324–1328.
- Sun JX, et al. 2012. A direct characterization of human mutation based on microsatellites. *Nat Genet.* 44:1161–1165.
- Tennesen JA, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337: 64–69.
- Tishkoff SA, et al. 2009. The genetic structure and history of Africans and African Americans. *Science* 324:1035–1044.
- Wang J, et al. 2008. The diploid genome sequence of an Asian individual. *Nature* 456:60–65.
- Xue Y, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol.* 19:1453–1457.
- Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Zhivotovsky LA, et al. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 74:50–61.
- Zhivotovsky LA, Underhill PA, Feldman MW. 2006. Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Mol Biol Evol.* 23:2268–2270.

Associate editor: George Zhang