An Upper Limit on the Functional Fraction of the Human Genome

Dan Graur*

Department of Biology and Biochemistry, University of Houston, TX

*Corresponding author: E-mail: dgraur@uh.edu.

Accepted: July 6, 2017

Abstract

For the human population to maintain a constant size from generation to generation, an increase in fertility must compensate for the reduction in the mean fitness of the population caused, among others, by deleterious mutations. The required increase in fertility due to this mutational load depends on the number of sites in the genome that are functional, the mutation rate, and the fraction of deleterious mutations among all mutations in functional regions. These dependencies and the fact that there exists a maximum tolerable replacement level fertility can be used to put an upper limit on the fraction of the human genome that can be functional. Mutational load considerations lead to the conclusion that the functional fraction within the human genome cannot exceed 15%.

Key words: human genome, mutational load, junk DNA.

Introduction

Many evolutionary processes can cause a population to have a mean fitness lower than its theoretical maximum. For example, deleterious mutations may occur faster than selection can get rid of them; recombination may break apart favorable combinations of alleles, thus creating less fit combinations; and genetic drift may cause allele frequencies to change in a manner that is antagonistic to the effects of natural selection. Genetic load (L) is defined as the reduction in the mean fitness of a population ($\bar{w}$) relative to the individual with the maximal fitness (w_{max}) in the population (Haldane 1937; Muller 1950).

$$L = \frac{w_{max} - \bar{w}}{w_{max}} \quad (1)$$

There are many kinds of genetic loads, such as the load caused by deleterious mutations, the segregation load, the substitutional load (also referred to as the "cost of natural selection"), the load due to recombination, and loads due to migration and inbreeding. In the following, we use the mutational load, that is, the reduction in mean population fitness due to deleterious mutations, as a proxy for the overall genetic load. This is a conservative approach, as the true genetic load can only be equal to or higher than our estimate.

The mutational load determines the mean fitness of a population, which in turn determines the mean fertility required to maintain a constant population size, that is, the replacement level fertility, as a function of the number of functional sites in the genome. Obviously, fertility values cannot be arbitrarily large, and that there exists a relatively modest upper limit for tolerable mean fertility values in human populations.

Here, we use empirical data on genome size, mutation rates, the fraction of deleterious mutations from among all mutations in functional regions, as well as data on fertility rates to estimate an upper limit on the functional fraction of the human genome.

Definitions

Throughout this paper, the term "function" is used to denote selected effect function, that is, a capacity that has been shaped by and is maintained by natural selection (Wright 1973; Graur et al. 2013, 2015; Brunet and Doolittle 2014). The selected effect function stands in contradistinction with the causal role function (or activity), which is ahistorical and nonevolutionary, and merely describes what an entity does (Cummins 1975; Amundson and Lauder 1994). A genomic segment is considered to possess a selected effect function if at least one out of all the possible mutations that can affect its sequence is deleterious (Graur 2016, pp. 492–496).

The mean fertility of a population is the mean number of offspring born per individual. Here, we are interested in the mean replacement level fertility ($\bar{F}$), that is, the fertility required to maintain a constant population size from generation to generation.

Model

The purpose of this model is to make a quantitative connection between the rate of deleterious mutation, the fraction of the genome that is functional, and replacement level fertility.

In the model, we assume that the probability of a mutation occurring in a certain region of the genome is independent of the functionality or lack of functionality of the region in which the mutation arises (Luria and Delbrück 1943; Lederberg and Lederberg 1952). We also assume that all mutations occurring in the nonfunctional fraction of the genome are neutral. Mutations occurring in the functional fraction of the genome, on the other hand, are assumed to be either deleterious or neutral. Advantageous mutations are known to be extremely rare (e.g., Eyre-Walker and Keightley 2009) and, hence, unlikely to affect the results.

By assuming that the fitness contributions of different loci are independent from one another, that is, that there is no epistasis, then the load of mutation can be approximated as

$$L \approx C\bar{\mu}_{\mathrm{del}} \tag{2}$$

where $\bar{\mu}_{\mathrm{del}}$ is the mean deleterious mutation rate and C is a constant between 1, for completely recessive mutations, and 2, for completely dominant mutations (Crow and Kimura 1970).

Thus, the mutational load does not depend on the strength of selection against any particular mutation. This surprising result comes from the fact that alleles under strong selection are relatively rare, but their effects on mean fitness are large, whereas the alleles under weak purifying selection are common, but their effects on mean fitness are small. As a result, the effects of these two types of mutation neatly cancel out. To understand the magnitude of the mutational load in a population, we need only determine the deleterious mutation rate, not the distribution of fitness effects.

The mean fitness of the population can be defined by two variables, the mean deleterious mutation rate per functional nucleotide site per generation ($\bar{\mu}_{\mathrm{del}}$) and the number of functional nucleotide sites (n) in the genome (Kimura 1961; Nei 2013).

$$\bar{w} = (1 - 2\bar{\mu}_{\mathrm{del}})^n \tag{3}$$

Note that the larger n is, the lower $\bar{w}$ will be.

Let us now consider the connection between mutational load and replacement level fertility ($\bar{F}$). If the mortality rate before reproduction age is 0 and mean fertility is 1, then the population will remain constant in size from generation to generation. In real populations, however, the mortality rate before reproduction is larger than 0 and, hence, mean fertility needs to be larger than 1 to maintain a constant population size. In the general case, for a population to maintain constant size, its replacement level fertility should be

$$\bar{F} = \frac{1}{\bar{w}} \tag{4}$$

Nei (2013).

Data

Genome Size

The maximal possible number of functional sites in the human genome equals the size of the diploid genome. The human diploid genome size has been estimated to be 6.114×10^9 nucleotides in length (Doležel and Greilhuber 2010).

Mutation Rates

Human germline mutation rates are known to vary among different regions of the genome (Harpak et al. 2016), to be different between males and females (Li et al. 2002), and to correlate with father's age (Kong et al. 2012). In humans, the mean germline point mutation rate at the DNA level has been inferred by many methods and by using a variety of data sets (Kondrashov and Crow 1993; Drake et al. 1998; Nachman and Crowell 2000, Kondrashov 2003; Xue et al. 2009; Roach et al. 2010; Campbell et al. 2012; Kong et al. 2012; Michaelson et al. 2012; Ségurel et al. 2014; Lipson et al. 2015). Notwithstanding the large number of estimates and estimation methodologies, the range of recent (i.e., 2010–2016) values for the germline mutation rate varies by merely a factor of 2.5, from 1.0×10^{-8} to 2.5×10^{-8} mutations per nucleotide site per generation (Scally 2016).

Fraction of Mutations in Functional Regions That Are Deleterious

What fraction of the mutations occurring in a functional region of the genome consists of deleterious mutation? Since at present we cannot answer this question as far as RNA-specifying and nontranscribed genes are concerned, we will use coding regions in protein-coding genes as models for functional genomic regions. Approximately 24% of all mutations occurring coding regions are synonymous and, hence, almost certainly not subject to purifying selection (Price and Graur 2016). If we assume that all missense and nonsense mutations are deleterious, then a maximalist estimate for the deleterious mutation rate in functional regions (μ_{del}) will be 76% of the total mutation rate (μ). Alternatively, we may assume that only nonsense

Table 1
Replacement Level Fertility Values in Humans As a Function of the Deleterious Mutation Rate (μ_{del}) and the Fraction of the Genome that is Functional[a]

μ_{del}	Functional fraction of the genome										
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.50	0.80	1.00
$4.00 \cdot 10^{-10}$	1.3	1.6	2.1	2.7	3.4	4.4	5.6	7.1	12	51	136
$5.00 \cdot 10^{-10}$	1.4	1.8	2.5	3.4	4.6	6.3	8.6	12	22	136	466
$6.00 \cdot 10^{-10}$	1.4	2.1	3.0	4.4	6.3	9.1	13	19	40	364	$1.6 \cdot 10^{3}$
$7.00 \cdot 10^{-10}$	1.5	2.4	3.6	5.6	8.6	13	20	31	74	974	$5.4 \cdot 10^{3}$
$8.00 \cdot 10^{-10}$	1.6	2.7	4.4	7.1	12	19	31	51	136	$2.6 \cdot 10^{3}$	$1.9 \cdot 10^{4}$
$9.00 \cdot 10^{-10}$	1.7	3.0	5.3	9.1	16	28	48	83	252	$7.0 \cdot 10^{3}$	$6.4 \cdot 10^{4}$
$1.00 \cdot 10^{-9}$	1.8	3.4	6.3	12	22	40	74	136	466	$1.9 \cdot 10^{4}$	$2.2 \cdot 10^{5}$
$2.00 \cdot 10^{-9}$	3.4	12	40	136	466	$1.6 \cdot 10^{3}$	$5.4 \cdot 10^{3}$	$1.9 \cdot 10^{4}$	$2.2 \cdot 10^{5}$	$3.5 \cdot 10^{8}$	$4.7 \cdot 10^{10}$
$3.00 \cdot 10^{-9}$	6.3	40	252	$1.6 \cdot 10^{3}$	$1.0 \cdot 10^{4}$	$6.4 \cdot 10^{4}$	$4.0 \cdot 10^{5}$	$2.5 \cdot 10^{6}$	$1.0 \cdot 10^{8}$	$6.4 \cdot 10^{12}$	$1.0 \cdot 10^{16}$
$4.00 \cdot 10^{-9}$	12	136	$1.6 \cdot 10^{3}$	$1.9 \cdot 10^{4}$	$2.2 \cdot 10^{5}$	$2.5 \cdot 10^{6}$	$3.0 \cdot 10^{7}$	$3.5 \cdot 10^{8}$	$4.7 \cdot 10^{10}$	$1.2 \cdot 10^{17}$	$2.2 \cdot 10^{21}$
$5.00 \cdot 10^{-9}$	22	466	$1.0 \cdot 10^{4}$	$2.2 \cdot 10^{5}$	$4.7 \cdot 10^{6}$	$1.0 \cdot 10^{8}$	$2.2 \cdot 10^{9}$	$4.7 \cdot 10^{10}$	$2.2 \cdot 10^{13}$	$2.2 \cdot 10^{21}$	$4.8 \cdot 10^{26}$
$6.00 \cdot 10^{-9}$	40	$1.6 \cdot 10^{3}$	$6.4 \cdot 10^{4}$	$2.5 \cdot 10^{6}$	$1.0 \cdot 10^{8}$	$4.0 \cdot 10^{9}$	$1.6 \cdot 10^{11}$	$6.4 \cdot 10^{12}$	$1.0 \cdot 10^{16}$	$4.1 \cdot 10^{25}$	$1.0 \cdot 10^{32}$
$7.00 \cdot 10^{-9}$	74	$5.4 \cdot 10^{3}$	$4.0 \cdot 10^{5}$	$3.0 \cdot 10^{7}$	$2.2 \cdot 10^{9}$	$1.6 \cdot 10^{11}$	$1.2 \cdot 10^{13}$	$8.8 \cdot 10^{14}$	$4.8 \cdot 10^{18}$	$7.7 \cdot 10^{29}$	$2.3 \cdot 10^{37}$
$8.00 \cdot 10^{-9}$	136	$1.9 \cdot 10^{4}$	$2.5 \cdot 10^{6}$	$3.5 \cdot 10^{8}$	$4.7 \cdot 10^{10}$	$6.4 \cdot 10^{12}$	$8.8 \cdot 10^{14}$	$1.2 \cdot 10^{17}$	$2.2 \cdot 10^{21}$	$1.4 \cdot 10^{34}$	$4.9 \cdot 10^{42}$
$9.00 \cdot 10^{-9}$	252	$6.4 \cdot 10^{4}$	$1.6 \cdot 10^{7}$	$4.0 \cdot 10^{9}$	$1.0 \cdot 10^{12}$	$2.6 \cdot 10^{14}$	$6.5 \cdot 10^{16}$	$1.6 \cdot 10^{19}$	$1.0 \cdot 10^{24}$	$2.7 \cdot 10^{38}$	$1.1 \cdot 10^{48}$
$1.00 \cdot 10^{-8}$	466	$2.2 \cdot 10^{5}$	$1.0 \cdot 10^{8}$	$4.7 \cdot 10^{10}$	$2.2 \cdot 10^{13}$	$1.0 \cdot 10^{16}$	$4.8 \cdot 10^{18}$	$2.2 \cdot 10^{21}$	$4.8 \cdot 10^{26}$	$4.9 \cdot 10^{42}$	$2.3 \cdot 10^{53}$

[a]Values above 1.8 are unrealistically high in humans.

mutations are deleterious, that is, that all amino acid replacements are neutral, in which case a minimalist estimate of the deleterious mutation rate will be 4% of the total mutation rate. Empirical data indicate that about half of all missense mutations in coding regions are deleterious (Soskine and Tawfik 2010). Adding the deleterious missense mutations to the deleterious nonsense mutations yields an empirical mean estimate for the deleterious mutation rate of ~40% of the total mutation rate.

Range of Deleterious Mutation Rates

By multiplying the lowest mutation rate estimate by the lowest possible fraction of deleterious mutations (4%), and by multiplying the highest mutation rate estimate by the highest possible fraction of deleterious mutations (76%), we infer that the rate of deleterious mutation ranges between 4×10^{-10} and 2×10^{-8} mutations per nucleotide site per generation. If we use the empirical estimate for the fraction of deleterious mutations out of all mutations (40%), then the range of deleterious mutation rates becomes 4×10^{-9} to 1×10^{-8} mutations per nucleotide site per generation.

Results

The required replacement level fertility was calculated for a range of deleterious mutation rate values from 4.0×10^{-10} to 10^{-8} mutations per nucleotide per generation as a function of the fraction of the human genome that is assumed to be functional. The results are shown table 1. We note that $\bar{F}$ scales positively and steeply with both the deleterious mutation rate and the number of functional sites in the genome.

Discussion

How high a replacement level fertility value can a human population tolerate? The answer is that $\bar{F}$ values cannot be arbitrarily large. One cannot imagine $\bar{F}=50$, that is, the situation in which each woman in a population gives birth to an average of 100 children of which on average 98 will die or fail to reproduce. Thus, there must exist a relatively modest upper limit for the tolerable mean replacement level fertility.

Although the oldest Homo sapiens fossil is ~315,000-years-old (Hublin et al. 2017), the common ancestor of all modern human populations is only 100,000–200,000-years-old (Green and Shapiro 2013). Throughout this period, mean replacement level fertility remained fairly constant (Davis 1986) and varied from <1.05 to nearly 1.75 per person, or from <2.1 to nearly 3.5 per couple (Espenshade et al. 2003). Given these numbers, we decided to use $\bar{F}=2$ as the maximum tolerable value.

From table 1, we see that even for low estimates of deleterious mutation rates, the fraction of the genome that can be functional cannot exceed 15%. These results agree with empirical estimates in the literature on the fraction of the human genome that is evolutionarily constrained (Rands et al. 2014).

Let us now see what happens if we assume that 80% of the diploid human genome is functional, as was claimed by The ENCODE Project Consortium (2012). By using the lower bound for the deleterious mutation rate (4×10^{-10} mutations

per nucleotide per generation), the mean individual fertility required to maintain a constant population size would be $\bar{F} = 51$. For 80% of the human genome to be functional, each couple in the world would have to beget on average 102 children and all but two would have to die or fail to reproduce. If we use the upper bound for the deleterious mutation rate (10^{-8} mutations per nucleotide per generation), then $\bar{F}$ becomes $\sim 5 \times 10^{42}$, that is, the number of children that each couple would have to have to maintain a constant population size would exceed the number of stars in the visible universe by ten orders of magnitude. The absurdity of such numbers was realized by Muller (1950, 1967) who suggested that genetic load values cannot exceed $L = 1$. Indeed, a recent estimate of the mutational load suggests that humans have a mutational genetic load of ~0.99 (Eory et al. 2010).

The situation becomes much more absurd and untenable if we assume that the entire genome is functional, as proclaimed by creationists such as Francis Collins, director of the National Institutes of Health (quoted in Zimmer 2015). Under the assumption of 100% functionality and the range of deleterious mutation rates used in this paper, maintaining a constant population size would necessitate that each couple on average produce a minimum of 272 and a maximum of 5×10^{53} children. For the genome to be entirely functional, the deleterious mutation rate should not exceed 10^{-11} mutations per nucleotide per generation, which is at least two or three orders of magnitude lower than estimates in the literature (see Reed and Aquadro 2006).

Above, we assumed that the mating pattern within human population is random and that deleterious mutations have independent effects on fitness. Deviations from either of these assumptions can affect the mutational load and consequently our estimate of the mean fertility required to maintain a constant population size. For example, both inbreeding and negative fitness epistasis (also referred to as synergistic epistasis on deleterious mutations) will reduce the mutational load by increasing the number of deleterious mutations removed from the population (Kimura and Maruyama 1966; Barrett and Charlesworth 1991). On the other hand, any factor that decreases the efficacy of selection, such as positive fitness epistasis (also referred to as antagonistic epistasis on deleterious mutations) or reduction in effective population size, will increase the mutational genetic load (Kimura et al. 1963).

Let us first deal with inbreeding. Empirical data pertaining to human populations show that with the exception of some isolates in Oceania and the Americas, genomic inbreeding coefficients in human populations are quite small and, in the context of mutational genetic load, negligible (Pemberton and Rosenberg 2014).

Dealing with fitness epistasis in humans and other nonmodel organisms is somewhat more complicated. In the largest study to date, Wang et al. (2017) took advantage of the fact that most African Americans inherited their genome from both African and European ancestors. In such a population, it is possible to discover fitness epistasis between two loci by detecting combinations of an African allele at one locus and a European allele at another locus that exist in the population at greater or lesser proportions than expected by chance. In Wang et al.'s study, more than 24 million pairwise-locus tests from about 16,000 individuals were performed. A single case of suspected epistasis was found, indicating that epistasis is exceedingly rare. This finding is in agreement with previous studies (e.g., Kouyos et al. 2007; Halligan and Keightley 2009), which showed that "there is little empirical evidence that net synergistic epistasis for fitness is common" (Keightley 2012).

Recently, Sohail et al. (2017) claimed to have found evidence for negative epistasis among deleterious alleles in humans. In their study, they divided mutations into synonymous, nonsynonymous (or missense), stop-codon gain, stop codon-loss, and mutations affecting splicing. The last three categories of mutations were grouped together into a category called loss-of-function (LoF). As a proxy for epistasis they used linkage disequilibrium as follows: In the absence of epistasis, alleles should contribute to the mutation burden independently, such that the variance of the mutation burden is equal to the sum of the variances at all loci, that is, to the additive variance (V_A). For rare mutant alleles, the mutation burden should follow a Poisson distribution with a variance (σ^2) equal to its mean (μ). Hence under no epistasis, $V_A = \sigma^2$ or $\sigma^2/V_A = 1$. If negative or synergistic epistasis on deleterious alleles operates, negative linkage disequilibrium will be observed and, as a result, the variance of the mutation burden will be reduced, leading to $\sigma^2/V_A < 1$. In contrast, under positive or antagonistic epistasis on deleterious alleles positive linkage disequilibrium between deleterious alleles will be observed, leading to $\sigma^2/V_A > 1$.

Sohail et al. (2017) reasonably assumed that most LoF mutations are deleterious. In the LoF category, the value of σ^2/V_A was 0.930, a very slight decrease in comparison with the expectation under no epistasis. This led them to claim that synergistic epistasis is prevalent among deleterious mutations. The problem with this conclusion is that LoF mutations constitute only ~3% of the mutations in the GoNL sample. Approximately 65% of the mutations were nonsynonymous. We know that to a greater or lesser extent, many nonsynonymous mutations have deleterious effects on fitness (e.g., Eyre-Walker et al. 2006; Eyre-Walker and Keightley 2007), but for this category of mutations, $\sigma^2/V_A = 2.077$, more than twice the expectation under no epistasis. This result indicates that for the vast majority of deleterious mutations, positive epistasis prevails. The existence of positive epistasis indicates that the 15% estimate for the upper limit on the functional fraction of the human genome may be exaggerated.

Finally, we note that in addition to inferring an upper limit on the functional fraction of the human genome, we can also conclude that the fraction of deleterious mutations out of all

mutations in functional regions should be very small. If >20% of all mutations in functional regions are deleterious, then the upper limit on the functional fraction of the human genome would be <2%, which is clearly false.

Acknowledgments

I thank Dr J. Spencer Johnston for his help with identifying the best estimate of human genome size. Drs Ricardo Azevedo, Diane Wiernasz, and Becky Zufall have provided valuable comments, and Mr Fei Yuan ran the calculations. Drs Shamil Sunayev and Mashaal Sohail guided me through the epistasis data. Two reviewers, Drs Charles Aquadro and Alexey Kondrashov, provided critical assessments that greatly improved this article.

Literature Cited

Amundson R, Lauder GV. 1994. Function without purpose. Biol Philos. 9:443–469.
Barrett SC, Charlesworth D. 1991. Effects of a change in the level of inbreeding on the genetic load. Nature 352:522–524.
Brunet TDP, Doolittle WF. 2014. Getting "function" right. Proc Natl Acad Sci U S A. 111:E3365.
Campbell CD, et al. 2012. Estimating the human mutation rate using autozygosity in a founder population. Nat Genet. 44:1277–1281.
Crow JF, Kimura M. 1970. An introduction to population genetics theory. Caldwell (NJ): Blackburn Press.
Cummins R. 1975. Functional analysis. J Philos. 72:741–765.
Davis K. 1986. Low fertility in evolutionary perspective. Pop Dev Rev. 12:48–65.
Doležel J, Greilhuber J. 2010. Nuclear genome size: are we getting closer? Cytometry 77A:635–642.
Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.
Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8:610–618.
Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol. 26:2097–2108.
Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173:891–900.
Eory L, Halligan DL, Keightley PD. 2010. Distributions of selectively constrained sites and deleterious mutation rates in the hominid and murid genomes. Mol Biol Evol. 27:177–192.
Espenshade TJ, Guzman JC, Westoff CF. 2003. The surprising global variation in replacement fertility. Pop Res Policy Rev. 22:575–583.
Graur D. 2016. Molecular and genome evolution. Sunderland (MA): Sinauer Associates.
Graur D, et al. 2013. On the immortality of television sets: "Function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol. 5:578–590.
Graur D, Zheng Y, Azevedo RB. 2015. An evolutionary classification of genomic function. Genome Biol Evol. 7:642–645.
Green RE, Shapiro B. 2013. Human evolution: turning back the clock. Curr Biol. 23:R286–R288.
Haldane JBS. 1937. The effect of variation on fitness. Am Nat. 71:337–349.
Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. Annu Rev Ecol Evol Syst. 40:151–172.
Harpak A, Bhaskar A, Pritchard JK. 2016. Mutation rate variation is a primary determinant of the distribution of allele frequencies in humans. PLoS Genet. 12:e1006489.
Hublin JJ, et al. 2017. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. Nature 546:289–292.
Keightley PD. 2012. Rates and fitness consequences of new mutations in humans. Genetics 190:295–304.
Kimura M. 1961. Some calculations on the mutational load. Jpn J Genet. 36:S179–S190.
Kimura M, Maruyama T. 1966. The mutational load with epistatic gene interactions in fitness. Genetics 54:1337–1351.
Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. Genetics 48:1303–1312.
Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat. 21:12–27.
Kondrashov AS, Crow JF. 1993. A molecular approach to estimating the human deleterious mutation rate. Hum Mutat. 2:229–234.
Kong A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471–475.
Kouyos RD, Silander OK, Bonhoeffer S. 2007. Epistasis between deleterious mutations and the evolution of recombination. Trends Ecol Evol. 22:308–315.
Lederberg J, Lederberg EM. 1952. Replica plating and indirect selection of bacterial mutants. J Bacteriol. 63:399–406.
Li W-H, Yi S, Makova K. 2002. Male-driven evolution. Curr Opin Genet Dev. 12:650–656.
Lipson M, et al. 2015. Calibrating the human mutation rate via ancestral recombination density in diploid genomes. PLoS Genet. 11:e1005550.
Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511.
Michaelson JJ, et al. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151:1431–1442.
Muller HJ. 1950. Our load of mutations. Am J Hum Genet. 2:111–176.
Muller HJ. 1967. The gene material as the initiator and the organizing basis of life. In: Brink RA, editor. Heritage from Mendel. Madison (WI): University of Wisconsin Press. pp. 417–447.
Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304.
Nei M. 2013. Mutation-driven evolution. Oxford: Oxford University Press.
Pemberton TJ, Rosenberg NA. 2014. Population-genetic influences on genomic estimates of the inbreeding coefficient: a global perspective. Hum Hered. 77:37–48.
Price N, Graur D. 2016. Are synonymous sites in primates and rodents functionally constrained? J Mol Evol. 82:51–64.
Rands CM, Meader S, Ponting CP, Lunter G. 2014. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. PLoS Genet. 10:e1004525.
Reed FA, Aquadro CF. 2006. Mutation, selection and the future of human evolution. Trends Genet. 22:479–484.
Roach JC, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639.
Scally A. 2016. The mutation rate in human evolution and demographic inference. Curr Opin Genet Dev. 41:36–43.
Ségurel L, Wyman MJ, Przeworski M. 2014. Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet. 15:47–70.
Sohail M, et al. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. Science 356:539–542.

Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 11:572–582.
The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.
Wang H, et al. 2017. Genome-wide survey in African Americans demonstrates potential epistasis of fitness in the human genome. Genet Epidemiol. 41:122–135.
Wright L. 1973. Functions. Philos Rev. 82:139–168.
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.
Zimmer C. 2015. Is most of our DNA garbage? New York Times. Available from: https://www.nytimes.com/2015/03/08/magazine/is-most-of-our-dna-garbage.html?_r=0

Associate editor: Bill Martin