

Dependence of Heterozygosity on Body Weight in Mammals

V. G. Gorshkov and A. M. Makar'eva

Presented by Academician V.E. Sokolov June 13, 1996

Received June 14, 1996

Diploidy of Metazoa increases the population diversity and protects developing organisms against negative somatic mutations. Both these benefits occur at the expense of suppression of the phenotypic manifestation of recessive mutations in the heterozygous state, i.e., in only one of two copies of a diploid genome. Deleterious somatic mutations are manifested only when they affect the dominant genes of paired chromosomes (autosomes) in the heterozygous state or of unpaired (hemizygous) X and Y sex chromosomes. In mammals, unpaired X and Y chromosomes are present in males. During development of a multicellular organism, the numbers of mutations directly depends on the genome size and the number of cell divisions, i.e., on the logarithm of body weight. Deleterious somatic mutations may emerge with equal probability in any genome region and impair the functioning of entire organs. For a multicellular organism to develop, it is necessary that the majority of its cells contain no phenotypically manifested somatic mutations. This imposes a limitation on the relative content of alleles of genes in the heterozygous state (heterozygosity) with respect to the genome size and the logarithm of the body weight. This study shows that the maximum heterozygosity within each unit of the logarithm of body weight range depends on the reciprocal value of this logarithm, so that the mean number of mutations accumulated in mammalian cells appears to be constant and much smaller than one.

The rate (μ) of accumulation of phenotypically manifested somatic mutations in the genome upon the development of a multicellular organism may be presented as follows:

$$\mu = \frac{G}{2}va, \quad a = kH_{\text{tot}}, \quad H_{\text{tot}} = H_a + H_s, \quad (1)$$

where G is the size of the diploid genome (bp); v is the probability of a point mutation per nucleotide per cell division; k is the number of dichotomous divisions of the zygote in a somatic line that are required for the organism to develop, H_{tot} is the total heterozygosity that equals the sum of autosomal heterozygosity H_a (the

ratio between the number of heterozygous genes and the total number of genes under study) and the hemizygous part of the genome H_s (the ratio between the total length of the X and Y chromosomes and the length of the haploid genome). $(G/2)H_{\text{tot}}$ is the functionally haploid part of the genome, which is capable of accumulating mutations with phenotypic manifestations.

The absolute rate of mutation accumulation and the cell division rate are proportional to the metabolism rate. Therefore, the mutation probability per cell division v does not depend on the metabolism rate, the body size, or information contained in the genome. In mammals, the probability v should be a constant value of about 10^{-10} mutations per bp per division [1, 2].

The size of the diploid genome G varies slightly in all mammals and, on the average, is 7.8×10^9 bp [4, 7]. The total length G_g of protein-encoding genes (0.4×10^8 bp) is 1% of the entire genome length G [1, 3, 4]. The number of mutations in this part of the genome is expressed by a version of formula (1) in which G is replaced by G_g . The number of dichotomous divisions can be calculated based on the ratio between the body weight m and the cell weight m_c ($m/m_c = 2^k$), which is equivalent to the equation

$$k = \left(\log \frac{m}{m_c} \right) / \log 2 = 3.3 \log \frac{m}{m_c}. \quad (2)$$

As a result, formula (1) can be presented as the following linear function:

$$H_a = -H_s + ak^{-1}. \quad (3)$$

Taking into account that $(G/2)v \approx 0.4$ and $(G_g/2)v \approx 0.4 \times 10^{-2}$, we obtain the following equations:

$$\mu = \frac{G}{2}va \approx 0.4a, \quad \mu_g = \frac{G_g}{2}va \approx 0.4 \times 10^{-2}a. \quad (4)$$

The mean weight of a mammalian cell is accepted to be 6×10^{-8} g (the cell size is 40 μm) [5].

We used the earlier published data on the heterozygosity of protein loci estimated from gel electrophoresis of homologous proteins [6] from 274 mammalian species. The total weight range, from shrews ($m \approx 2$ g) to whales ($m \approx 30$ t), was divided into six logarithmic intervals. Within each interval, the maximum values

(10%) of protein heterozygosity H_a in natural species were averaged. The mean values were considered to be the extreme permissible heterozygosity levels for natural mammalian species, $H_a = H_{ac}$, within a weight interval (Fig. 1). Processing of the linear regression by the method of least squares revealed the following inclination ($a = a_c$) and intersection ($-H_s$) values in the formula (3) and the correlation coefficient (r):

$$a = a_c = 4.8 \pm 1.4; H_s = 0.05 \pm 0.01; \quad (5)$$

$$r = 0.93.$$

Note that the relative contribution of X and Y chromosomes in the genome (approximately 5% for mammals [7]) is in a good agreement with the independently obtained intersection value H_s (5). The inclination $a = a_c$ (5) determines the mean permissible number of mutations $\mu = \mu_c$ (4) in mammalian somatic cells. In accordance with (4), this value is about one ($\mu \sim \mu_c \sim 1$) for the entire genome G or 10^{-2} ($\mu_g \sim \mu_{gc} \sim 2 \times 10^{-2}$) for protein-encoding genes G_g in all mammals. According to the Poisson distribution, the percentage of cells lacking somatic mutations in protein-encoding genes is $e^{-\mu_{sc}} \approx 0.98$. For natural mammalian species to exist, it can be considered that this percentage must be equal to or greater than this value.

Under unnatural conditions when the competition between specimens in a population is decreased or eliminated (e.g., upon artificial selection), accumulation of deleterious alleles and the heterozygosity increase to a certain lethal level $H_{tot} = H_{L_{tot}}$. Species which have been removed from their natural ecological niches for a long time (e.g., horses, humans, and domestic cats) are characterized by a heterozygosity that is several times higher than that in natural species of the same body weight (Figs. 1, 2).

Figure 2 presents the distribution of mammalian species with respect to the rate of accumulation of phenotypically manifested somatic mutations, $a = kH_{tot}$ (1)–(5). It is characterized by a sharp decline with density of the species number in the region above the permissible number of somatic mutations (5). The mutagenesis rate corresponding to this distribution determines the resolution capacity of the competition between individuals in natural populations $a = a_c \approx 5$. Eleven species are located to the right of the value $a = 5.5$; five of them are domestic animals, and the other six species living under altered environmental conditions. The value $a = a_L$ determines the maximum observed rate of somatic mutation accumulation (horse) and may be considered to be the lethal limit for all mammals.

If no lethal limit were present and at the heterozygosity level equal to one, mammals could exist in the haplophase. The absence of haploid mammals and the fact that the observed heterozygosity does not exceed $H_{amax} = 30\%$ [6] unambiguously indicate that the het-

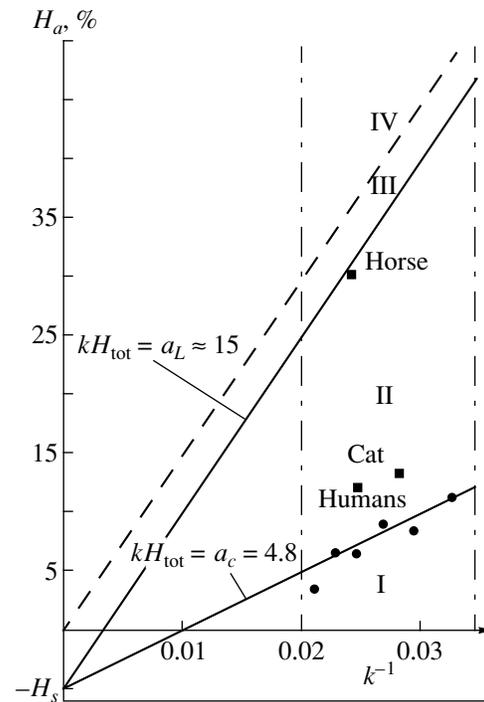


Fig. 1. The dependence of the autosomal heterozygosity on the logarithm of body weight in mammals and the applicability range for the Haldane rule. Solid lines indicate heterozygosities (H_a) with equal intersections and different inclinations: the lower line is the extreme permissible values for natural species under normal environmental conditions, and the upper line is the lethal threshold. The dashed line is parallel to the upper solid line with the added sexual heterozygosity value (H_s) and crosses the ordinates at zero. Between the upper solid line and the dotted line, only the heterogametic sex and parthenogenetic (unisexual) species consisting of homogametic females occur; the heterogametic sex is beyond the limits of the lethal threshold, and the Haldane rule for hybrids is fulfilled. Region I, natural species under normal environmental conditions; region II, domestic strains and species under altered environmental conditions; region III, the Haldane rule for hybrids is fulfilled; and region IV, the absence of viable mammals.

erozygosity lethal limit H_L exists in mammals is close to $H_{amax} \sim 30\%$.

The maximum rate of accumulation of somatic mutations that impair an organism's viability (μ_L) can be considered to be a universal constant for mammals that satisfies equations of the type (1):

$$\mu_L = a_L \frac{G}{2} v; \mu_{gL} = a_L \frac{G_g}{2} v; a_L = kH_{L_{tot}}. \quad (6)$$

Assuming $H_{amax} = 0.3$ and $H_{tot} = H_{amax} + H_s = H_{L_{tot}} \sim 0.35$ for the maximum body weights, we obtain from (6) that $\mu_g \sim \mu_{gL} \sim 0.1$. In this case, the number of cells lacking newly emerged, phenotypically manifested somatic mutations in protein-encoding genes decreases to $e^{-\mu_{gL}} \sim 0.90$; i.e., one newly emerged somatic mutation in every ten cells of an adult organism is lethal.

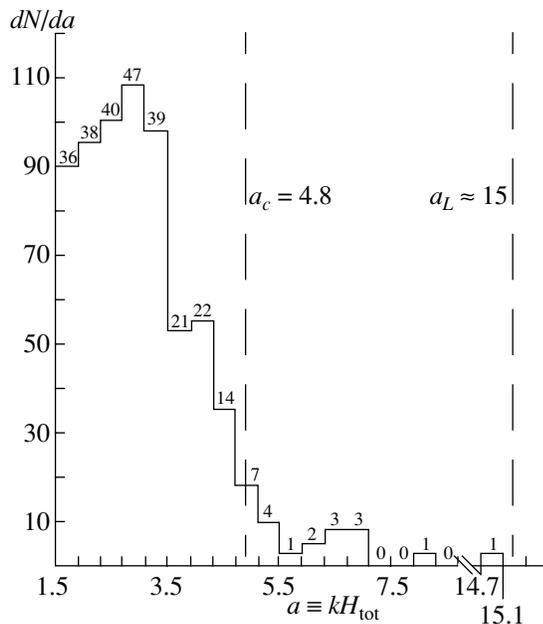


Fig. 2. Distribution of mammalian species with respect to the mean number of phenotypically manifested somatic mutations. Abscissa, $a = kH_{\text{tot}}$, the percentage of somatic mutations, see (1) and (4); (H_{tot}), total protein heterozygosity; (k), the number of dichotomous divisions of a zygote in a somatic line. Ordinate, the density of mammalian species (dN/da) falling into a unit interval of a . Over each interval, the number of species within this interval is indicated. To the right of the break in the scale, only one species (horse) occurs. Vertical dashed lines, a_c , the extreme permissible level of the genetic diversity that is maintained by the competition between natural species under normal environmental conditions; and (a_L), the lethal level of the genetic diversity.

Considering μ_L to be a universal value, the dependence of the lethal protein heterozygosity on the body weight can be plotted according to the formula (6), taking $\mu_g = \mu_{gL} \sim 0.1$ (Fig. 1). No viable mammals of either sex fall above this line, in region IV. In region III, between the upper solid line and the dashed line that lies at a distance of the sexual heterozygosity value H_s from the upper line, only viable homogametic females occur, not heterogametic males. Finally, in region II, between the

dashed and lower solid lines, viable mammals of both sexes occur under altered conditions. In region I, below the solid line determined by the parameters of (5), mammals inhabiting natural ecological niches occur.

The existence of region III gives a simple explanation for the well-known Haldane rule that if either sex of a hybrid is nonexistent, sterile, or inviable, this is the heterogametic sex [8]. In fact, in a hybrid, the total heterozygosity $H_{\text{tot}} = H_a + H_s$ of (1) is augmented by the hybrid heterozygosity H_h associated with certain genetic differences between the two interbred species. Thus, in hybrids, the values of H_a and H_s are equal to half the sum of the corresponding values in each of the species [9]. When addition of the hybrid heterozygosity results in the hybrid falling into region III, the heterogametic sex of this hybrid follows the Haldane rule.

ACKNOWLEDGMENTS

We are grateful to L.Z. Kaidanov, V.G. Kuznetsova, and M.B. Filatov for helpful discussion.

REFERENCES

1. Gershenzon, S.M., *Osnovy sovremennoi genetiki* (Fundamentals of Modern Genetics), Kiev: Naukova Dumka, 1983.
2. Drake, J.W., *Proc. Natl. Acad. Sci. USA*, 1993, vol. 90, pp. 4171–4175.
3. Ayala, F.J. and Kiger, J.A., *Modern Genetics*, Menlo Park: Benjamin/Cummings, 1984. Translated under the title *Sovremennaya genetika*, Moscow: Mir, 1987.
4. Lewin, B., *Genes*, New York: Wiley, 1983. Translated under the title *Geny*, Moscow: Mir, 1987.
5. Olmo, E., *Basic Appl. Histochem.*, 1983, vol. 27, pp. 227–256.
6. Nevo, E., Beiles, A., and Ben-Shlomo, R., in *Lecture Notes in Biomathematics*, New York: Springer.
7. Orlov, V.N. and Bulatova, N.Sh., *Sravnitel'naya tsitogenetika i kariosistematika mlekopitayushchikh* (Comparative Cytogenetics and Karyosystematics of Mammals), Moscow: Nauka, 1983.
8. Haldane, J.B.S., *J. Genet.*, 1922, vol. 12, pp. 101–109.
9. Nei, M., *Molecular Population Genetics and Evolution*, Amsterdam: Elsevier, 1975.