Jukes-Cantor distance

In the [Jukes and Cantor (1969)](#) model, the rate of nucleotide substitution is the same for all pairs of the four nucleotides A, T, C, and G. As is shown below, the multiple hit correction equation for this model produces a maximum likelihood estimate of the number of nucleotide substitutions between two sequences. It assumes an equality of substitution rates among sites (see the related [gamma distance](#)), equal nucleotide frequencies, and it does not correct for higher rate of [transition](#)al substitutions as compared to [transversion](#)al substitutions.

The Jukes-Cantor model

	A	T	C	G
A	-	α	α	α
T	α	-	α	α
C	α	α	-	α
G	α	α	α	-

MEGA provides facilities for computing the following quantities:
d: **Transitions + Transversions** : Number of nucleotide substitutions per site.
L: **No of valid common sites**: Number of sites compared.

Formulas for computing these quantities are as follows:

Distance

$$d = -\tfrac{3}{4} \log_e \left(1 - \tfrac{4}{3} p\right)$$

where *p* is the proportion of sites with different nucleotides.

Variance

$$\mathrm{Var}(d) = p(1-p) \Big/ \left[\left(1 - \tfrac{4}{3} p\right)^2 L\right]$$

See also [Nei and Kumar (2000)](#), page 36.