A nucleotide sequence that is not under selection (or only under very weak selection) experiences substitutions with a rate of 10^-8 substitutions per site and per year. (This rate applies for intron sequences, many synonymous substitutions, sequences between genes. This does not imply that these are not under selection – there is a preferred codon usage for example; but it says that the selection often is not very strong.) Without selection, two sequences that evolved from a common ancestor 3,500 million years ago (in total separated by 7 billion years), experienced rate times time = 10^-8 (substitutions per site and per year) \* 7 \*10^9 (years) = 70 substitutions per site.

To find out how long it takes until 50% of the sites have experienced a substitution, and ignoring multiple substitutions and back mutations, one could write:
rate \* unknown time =0.5
or with the time in years being X:
10^-8 \* X=0.5

X=0.5\*10^8 = 50 million years. A common ancestral sequence would have diverged to two extant sequences with that difference in 25 million years.

To find out how long it takes until 80% of the sites have experienced a substitution, and ignoring multiple substitutions and back mutations, one could write:
rate \* unknown time =0.8
or with the time in years being X:
10^-8 \* X=0.8

X=0.5\*10^8 = 80 million years. A common ancestral sequence would have diverged to two extant sequences with that difference in 40 million years.

Obviously, this reasoning is deeply flawed, because two random sequences with 4 letters are already 25% identical (provided the 4 nucleotides occur with equal frequency). A simple substitution model that takes back mutations and multiple substitutions into account is the Jukes Cantor model.

The Jukes Cantor estimate for divergence (see <http://en.wikipedia.org/wiki/Models_of_DNA_evolution> ) is

The Jukes-Cantor relationship between observed differences p between two sequences and the number of substitution events d is

d=-(3/4)\*ln(1-4/3p)

or

p = 3/4-3/4\*EXP((-(4/3)\*d))

With d = .5 (substitutions per site on average), p is .36 (average differences per between sites).

With d=1, p is about .55

With d=5 substitutions per site, p=0.749

An Excel spreadsheet with tables for both nucleotide and amino acid distances according to the Jukes Cantor model is at

http://gogarten.uconn.edu/mcb3421\_2014/ JukesCantorCorrection.xls