Mathematical Analysis of Genomic Evolution

Cedric Green
University of South Florida

Mathematical Analysis of Genomic Evolution

Abstract

Changes in nucleotide sequences, or mutations, accumulate from generation to generation in the genomes of all living organisms. The mutations can be advantageous, deleterious, or neutral. The goal of this project is to determine the amount of advantageous mutations it takes to get human (Homo sapiens) DNA from the DNA of genetically distinct organisms. We do this by collecting the genomic data of such organisms, and estimating the amount of mutations it takes to transform yeast (Saccharomyces cerevisiae) DNA to the DNA of a human. We calculate the typical number of mutations occurring annually through the organism's average life span and the average mutation rate. This allows us to determine the total number of mutations as well as the probability of advantageous mutations. Not surprisingly, this probability proves to be fairly small. A more precise estimate can be determined by accounting for the differences in the chromosomal structure and phenomena like horizontal gene transfer.

Keywords

Genomic Evolution, Beneficial Mutations, DNA

Creative Commons License

TABLE OF CONTENTS

PROBLEM STATEMENT

The goal of this project is to estimate the percentage of beneficial mutations necessary to transform genome of an early eukaryote to the human genome.

MOTIVATION

What is the likelihood of humans (*Homo sapiens*) coming about by the route of evolution? Embryology, studies of the DNA sequence homology, and radiometric dating seem to confirm the theory of evolution and prove that this likelihood is very high. We can, without a doubt, see populations changing as they adapt to changes in their environment. Harmful bacteria developing immunity to certain antibiotics is one example.

Instead of discussing microevolution, however, we concentrate on evolutionary changes happening over much larger periods of time. One seemingly common critique of long-term evolution used by many creationists is the lack of time needed for the evolution to lead from bacteria to the modern human being. Therefore the objective of this paper is to use the knowledge about mutations that occur from generation to generation to calculate the probability of finding enough useful genes in each generation for such a transformation to take place. An egregiously high percentage of necessary advantageous mutations would serve to discredit the idea of slow adaption over time from generation to generation. A low percentage would be expected for pragmatic adaption over time.

MATHEMATICAL DESCRIPTION AND SOLUTION APPROACH

We have collected the genomic data on various species (see Table 1 in Appendix). The list includes a relatively simple, single-celled eukaryote (yeast), as well as more complex organisms whose biological organization closer resembles that of human. We also include the estimate of similarity between these species' genomes and the human genome (Table 1).

Average lifespan. Our first goal is to use the data in Table 1 to find the average life span. The first billion years on the earth out of the approximately 4.5 billion were completely lifeless. During the subsequent 3 billion years, life consisted of single celled organisms and simple multicellular organisms. More complex multicellular organisms existed for about the last half a billion years. In our calculations we consider only the time elapsed after endosymbiosis led to emergence of the first eukaryotes (approximately the last 2.5 billion years). The following calculations estimate the average life span of organisms on the planet since the appearance of the eukaryotes.

We use yeast as a representative eukaryote of Proterozoic Eon, which gives the life span estimate of 90 min or 1.711×10^{-4} years. For Phanerozoic Eon, the rest of species listed in Table 1 give 4.725 years. The overall average is calculated by the following expression:

$$\frac{2}{2.5}(1.711 \cdot 10^{-4}) + \frac{0.5}{2.5}(4.725) = 0.945$$

years per generation.

Mutations. Note that: $0.5/2.5 \cdot 175$ (mutations per genome in *Homo sapiens* (Estimate of Human Mutation Rate, 2010)) $+ 2.0/2.5 \cdot 0.000063$ (mutations per genome in *Saccharomyces cerevisiae* (Mutations in Saccharomyces Cerevisiae., 2012)) $= 35$ mutations per genome per

generation. This estimate represents the average number of mutations per genome (or generation) for all of the considered species. Compared to the other species we selected, *Saccharomyces cerevisiae* is most representative of early eukaryotes that existed during the first 2 billion years. We take this fact into account in the above equation of the weighted average.

Dividing the average amount of mutations by the average life expectancy above gives a rough estimate of the average amount of mutations per year, i.e., $35/0.945 = 37.037$ mutations per year.

Percent Advantageous Mutations. The *Saccharomyces cerevisiae* will be used in order to find the number of genes that are different and give a rough estimate of how many mutations had to be advantageous and stay throughout time.

There are approximately $1.2 \cdot 10^7 \cdot 0.31 = 3.72 \cdot 10^6$ base pairs in common between the yeast genome and the human genome. Thus, there must be $3.2 \cdot 10^9 - 3.72 \cdot 10^6 = 3.19628 \cdot 10^9$ advantageous mutations.

The mutations per year (assuming each point mutation is a change of an already existing nucleotide) will now be used to calculate the amount of mutations that could have occurred since the rise of eukaryotes through a single line of organisms slowly changing from generation to generation. This result, in turn, will be used to calculate the probability of getting a beneficial genome. Our calculation accounts for only the last 2.5 billion years because this is the time elapsed since the appearance of the first eukaryotes. Thus we get 37.037 (mutations per year) $\cdot 2.5 \cdot 10^9$ (years) $= 9.25925 \cdot 10^9$ mutations total since eukaryotes originated, and $3.19628 \cdot 10^9/9.25925 \cdot 10^{10} = 0.0345$ or 3.45% gives the probability of acquiring the needed beneficial mutations.

DISCUSSION

This means that approximately 3.45% of all mutations that occur in each generation would have to be beneficial in order for the transition from single-celled eukaryotes to be pragmatic. This, however, does not account for other phenomena such as horizontal gene transfer. The presented model also assumes that natural selection is working at a constant rate that never allows beneficial genes to become extinct. The expected rate of acquiring beneficial genes from mutations would actually be even lower when considering these and other phenomena that cause modification of genomes over time.

CONCLUSION AND RECOMMENDATIONS

The resulting percentage of maximum beneficial mutations is relatively small. It shows that the number of beneficial mutations found in nature does not need to be high in order for the species to undergo cladogenesis. Researchers investigating this problem in the future should also account for the horizontal gene transfer and use a wider variety of species.

REFERENCES

Zebrafish | Developmental Biology Interactive. (2012, Jan 20). Retrieved from Developmental Biology Interactive | A Learning Resource by Students for Students: http://www.devbio.biology.gatech.edu/?page_id=204

Estimate of the Mutation Rate per Nucleotide in Humans. (2010, May 11). Retrieved from Genetics: http://www.genetics.org/content/156/1/297.full

Dye, L. (2011, May). *A Fish That's Almost Human?* Retrieved from ABC News: http://abcnews.go.com/Technology/story?id=3432823&page=1

Pavel, T. (n.d.). *Current Trends in Chicken Genome Manipulation: Transgenesis and Biotechnology of.* Retrieved from http://www.cabi.org/animalscience/Uploads/File/AnimalScience/additionalFiles/WPSA2_files/Trefil.pdf

Primate Factsheets: Chimpanzee (Pan Troglodytes) Behavior. (2012, Jan 2012). Retrieved from http://pin.primate.wisc.edu/factsheets/entry/chimpanzee/behav

Species: Comparing Their Genome (ActionBioscience). (2010, May 11). Retrieved from ActionBioscience - Promoting Bioscience Literacy.: http://actionbioscience.org/genomic/hhmi.html

Spontaneous Mutations in Diploid Saccharomyces Cerevisiae. (20, Jan 2012Jan). Retrieved from Genetics: http://www.genetics.org/content/168/4/1817.long

The Genome of the Western Clawed Frog XenopusTropicalis. (23, Jan 2012). Retrieved from Office of Scientific and Technical Information, OSTI, U.S. Department of Energy: http://www.osti.gov/bridge/servlets/purl/1010537-cthNU1/1010537.pdf

APPENDIX-TABLES

Organism	Species	# genes	# base pairs (millions)	Reproduction Rate (new generation)	Correlation w/ *Homo sapiens*
Yeast	*Saccharomyces cerevisiae*	6,000	12	90 minutes	31%
Worm	*Caenorhabditis elegans*	19,099	99	3 days	40%
Fly	*Drosophila melanogaster*	13,600	165	10 days	50%
Zebra Fish	*Daniorenia*	No data found	385	3.5 months	80%
Clawed frog	*Xenopus tropicalis*	26,000	1,760	1 year	75%
Chicken	*Gallus gallus*	22,000	1,000	5 months	60%
Mouse	*Mus musculus*	40,000	3,000	9 weeks	92%
Chimpanzee	*Pan troglodytes*	40,000	3,200	16 years	98%
Human	*Homo sapiens*	40,000	3,200	20 years	100%

Table 1: Similarities between *Homo sapiens* and Other Species (Sources (Species: Comparing Their Genome (ActionBioscience), 2010) (The Genome of the Western Clawed Frog XenopusTropicalis., 23) (Pavel) (Primate Factsheets: Chimpanzee (Pan Troglodytes) Behavior, 2012) (Zebrafish | Developmental Biology Interactive, 2012)).