Why Human and Chimp whole genomes are 99.99% close considering Codons Usage frequencies ?

Jean-claude Perez #
 # 7 av de terre-rouge F33127 Martignas France
Email : jeanclaudeperez2@free.fr

Abstract:

While analyzing in an exhaustive way the distribution of the respective frequencies of the 64 codons (« codons usage »), we show here that the whole genomes of human and chimpanzee are homologous with more than 99.99%. that is much higher than the commonly allowed homologies of 98.5% and even of 95%. We discover also a heterogeneity common to both genomes highlighting a partition of the chromosomes into 2 clusters demonstrating evidence of 2 common ancestors. Then, finally, this paper output must be a full reconsideration of evaluations for a date and genesis of an hypothetic "most recent common ancestor" (MRCA)...

Introduction:

One speculates since millenia in the resemblance between the man and the monkey, and more particularly the chimpanzee. The progressive availability of fragments of the two genomes resulted in affirming that they would be homologous in particular to 98.5% [1] for the proteins coding regions. In 2002, Roy Britten lowers this homology with only 95% by analyzing the hybridization of long sequences of the 2 genomes in which he includes the "indels" (nonhomologous areas) [2] . Because one has today the complete sequences of the human genome [3] and of the chimpanzee [4], it becomes from now on possible to compare overall these two genomes by using other operators like global distance inter-genomes operators.

The 99.99% Homology:

We analyzed then compared the integrality of these two genomes in their most recent releases [5 6]. Our method consists in computing the frequency of the populations of each of the 64 Universal Genetic Code codons ("codons usage") for the totality of the chromosomes (24 for the human genome and 25 for the chimp. genome). This task is repeated for each of the 3 codons reading frames related to the main DNA strand. Unspecified « N » bases are ignored. Each sequence is thus compressed in a pseudo-sequence comprising only bases TCAG. One thus obtains 2843411612 bases for the human genome and 2407863342 bases for that of the chimpanzee. The populations of codons are cumulated for the totality of the chromosomes of each of the two genomes. One thus obtains two vectors of 64 values each one corresponding to the populations of each one of 64 codons calculated for the whole genomes. The correllation of these two populations is computed using "Bravais Pearson"'s correllation method (please see WEB suplementary data). We obtain the following results:
If one cumulates the 3 codons reading frames, one obtains a total correllation of both genomes of **99.99022426%**. Results detailed by codons reading frames are of the same order (see Figure1).

Human Chimp. Codons Usage %	Chimp 1st codons reading frame	Chimp2nd codons reading frame	Chimp 3rd codons reading frame
Human 1st codons reading frame	99.9901562	99.9903325	99.9902947
Human 2ndcodons reading frame	99.9900776	99.9902470	99.9902058
Human 3rdcodons reading frame	99.9900617	99.9902208	99.9901864

Figure1- Human/Chimp. Whole genomes Codons usage homologies by codons reading frames.

Common Archaic Chromosomes Heterogeneity:

If one carries out this same type of correllations on the level of each chromosome, it appears a remarkable phenomenon: certain chromosomes of the man are closer to certain chromosomes of the chimpanzee than of the other human chromosomes, and vice versa. One highlights quickly, in each 2 genome, two clusters of chromosomes revealing a chromosomal HETEROGENEITY on the level of codons usage. Thus, we call MINOR-HUMAN the cluster human genome composed of the 5 chromosomes 16 17 19 20 22, and MAJOR-HUMAN, the cluster related to the remaining chromosomes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 21 X Y. We make in the same way for Chimpanzee: MINOR-CHIMP = chromosomes 18 19 20 21 23, and MAJOR-CHIMP = chromosomes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 22 X Y. Figure 2 below shows this heterogeneity common to the 2 genomes (% computed cumulating the 3 codons reading frames). Internal heterogeneity with each 2 genome is obvious: < 90% while the homogeneities between clusters in the same classes (Major/Major and Minor/Minor) are 99.994% (majors human/chimp) and of 99.997% (minors Human/chimp).

% chromosomes Heterogeneity	minor human	major human	minor chimp	major chimp
minor human	100	89.840	99.997	89.402
major human	0	100	89.872	99.994
minor chimp	0	0	100	89.437
major chimp	0	0	0	100

Figure2-Human/Chimp. Heteroheneity by chromosomes clusters evidence.

Two common ancestors:

What to say in connection with this heterogeneity common to the chromosomes of the man and the chimpanzee? On the one hand, in the plan of EVOLUTION, that would mean that both genomes Human and Chimpanzee would be resulting from a possible FUSION between two ancestral genomes roots. In addition, in the functional plan, a recent study of affinities of integration of retroviruses genomes like HIV within the various human chromosomes [7] shows a "permeability" 2 to 3 times higher than that of all the other chromosome for the human chromosomes 16 17 19 and 22 (see Figure 3 and WEB supplementary materials). However these 4 chromosomes all belong to the same cluster of the human genome (MINOR-HUMAN cluster). That would thus mean that the rétroviruses genomes of type HIV are integrated more easily within this chromosomal cluster.

Lastly, other results which we will publish separately reveals very high levels of numerical structures common to the genomes of the man and the chimpanzee. We will also show that it would seem that these high level structures are higher in the case of the human genome...

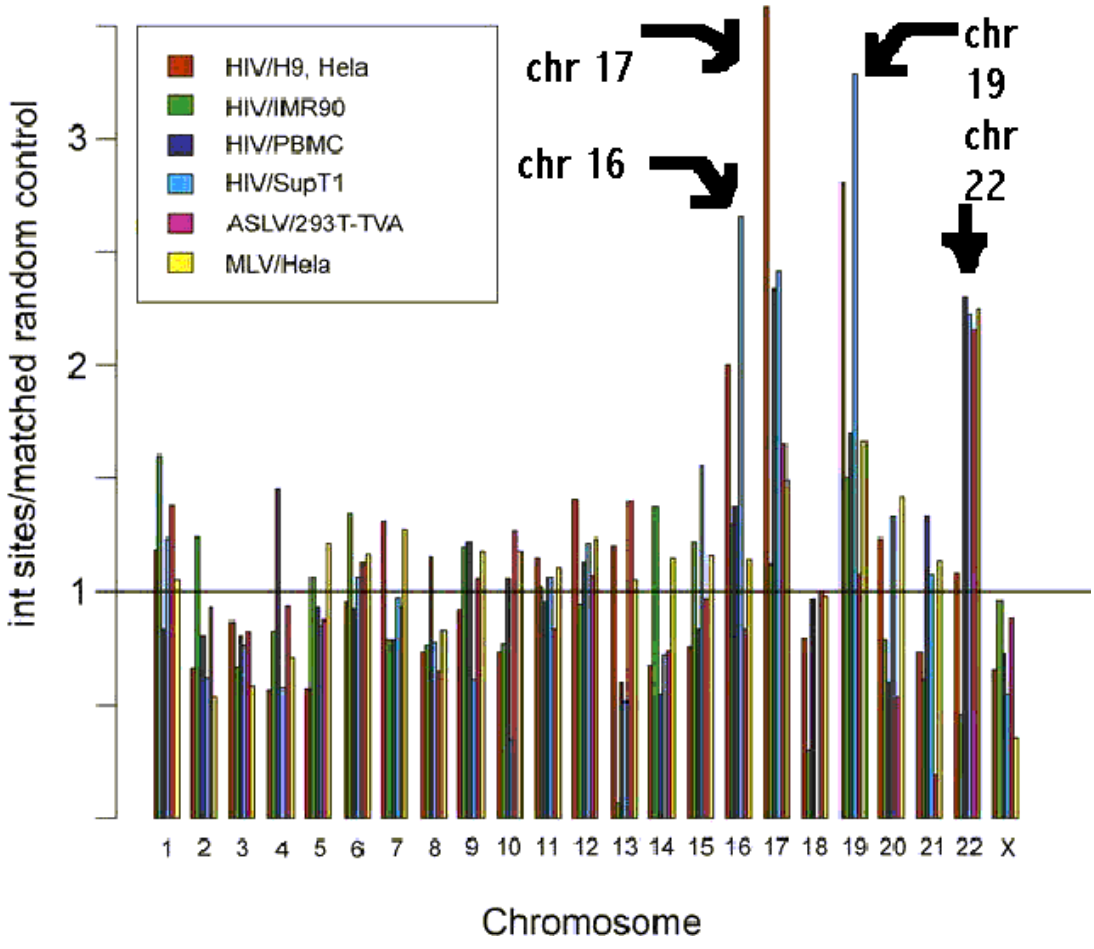

Figure 3 – Evidence of Human genome HIV strong permeability for Chr 16 17 19 22 (from reference 7, please visit WEB supplementary materials for details).

Towards perspectives in the EVOLUTION and ORIGINS of HUMAN LIFE fields :

It remain very difficult to compare whole human and chimp genomes for various reasons like different numbers of chromosomes or different numbers of nucleotides differenciating both genomes, meanwhile there are facts:

- Our codons based comparison operator is independant with the lenght of compared genomes.
- evaluations for a date of a "most recent common ancestor" (MRCA) by evolutionists could vary a bit between a recent single origin of about 100,000-200,000 years ago and values of approximately between 4 and 7 million years ago.
- If we take the Britten's 95% correlation, there are about 150 millions bases differenciating human and chimp genomes.
- If we take the commonly estimated value of 98.8% correlation, there are 40 millions base pairs differentiating both genomes.
- If wa take our 99.99% value, there are only about 300000 bases pairs separating human and chimp genomes...

- MEANWHILE:
- Human SNP differentiating each individual human with another is estimated of about 3 million bases pairs (99.9%)...
- General genomes sequencing rate error is about 1 / 10000 nucleotides then 99.99%.
- Strange overlapped and embedded percentages and facts involving both Human and Chimp genomes evolution and evaluation...

Then, if the nucleotides gap separating human and chimp is equivalent with sequencing accepted error or with SNP variability differenciating two individual single humans, what about the hypothetic date and genesis of our common hypothetical common ancestiors ("most recent common ancestor" MRCA) .
Could SNP be the central key of this major evolution?
In other unpublished research I demonstrate that SNP are strategic in human genome not specifically by their value but by their exact location within the genome.
What about an hypothetical theory of a possible global SNP massive mutation doing the breakthough from MRCA towards both chimp and humans?
It remain now possible to analyse this scenario comparing SNP within both haman and chimp genomes...
Why not?

---

[1] Ebersberger, I. , Metzler, D. , Schwarz, C. & Paabo, S. (2002) *Am. J. Hum. Genet.* **70**, 1490-1497.
[2] Britten, R. J., 2002. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proceedings of the National Academy of Science* 99(21):13633-13635.
[3] Baltimore D.*Nature* **409**, 814-816 (2001).
[4] NIH 2003 Chimp Genome Assembled by Sequencing Centers available doing http://www.genome.gov/11509418
[5] Human genome finalized BUILD34": http://hgdownload.cse.ucsc.edu/goldenPath/hg16/chromosomes/
[6] FTP UCSC complete Chimp genome: http://hgdownload.cse.ucsc.edu/goldenPath/panTro1/chromosomes/
[7] Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, et al. (2004) Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. PLoS Biol 2(8): e234