

Allelic Frequency in Human SNPs Predicts the Rate of Non-Synonymous Nucleotide Substitutions between Human and Chimpanzee Genes

Hippokratris Kiaris, Athanasios G. Papavassiliou

Department of Biochemistry, University of Athens Medical School, Athens, Greece
Email: hkiaris@med.uoa.gr

Received December 5th, 2013; revised January 3rd, 2014; accepted February 2nd, 2014

Copyright © 2014 Hippokratris Kiaris, Athanasios G. Papavassiliou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Hippokratris Kiaris, Athanasios G. Papavassiliou. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

The combination of comparative genomics and population genetics may provide important clues regarding human evolution. We have hypothesized that the divergence between various human and chimpanzee orthologs will be reflected in the variability of single nucleotide polymorphisms (SNPs) that are localized in the vicinity of the corresponding loci in different human populations. Consistently with this notion, more diverged genes between humans and chimpanzees are more likely to be associated with human speciation and are anticipated to contain SNPs with reduced variability between different human populations. In order to test this hypothesis, we have compared the rate of non synonymous nucleotide substitutions (K_a) between 255 chimpanzee and human orthologs with the average deviation in the allelic frequencies of corresponding closely linked SNPs in two distinct human populations: The Yoruba people in Ibadan, Nigeria (YRI) and US residents with ancestry from Northern and Western Europe, collected in 1980 by the Centre d'Etude du Polymorphisme Humain (CEU). We found a significant ($p < 0.05$) negative association between the K_a and the degree of variation in the corresponding allelic frequencies between the human populations which implies that the most significant genes for human speciation are associated lower variability between the human populations examined. This observation is consistent with the strong selective advantage offered by these nucleotide substitutions during human evolution and predicts that a low polymorphic rate is consistent with the presence of genes with an essential role in human speciation.

Keywords: HapMap; Chimpanzee Genome; Human Speciation

Introduction

The genetic details of human speciation and the identification of the array of the genes with the most important role in the divergence of humans and chimpanzees from their common evolutionary ancestor, represents an intriguing problem in modern biology because of its obvious importance in the study of human evolution (Disotell, 2006; Raaum et al., 2005). In addition it possesses important implications in biomedicine since it ultimately aims to identify genes intrinsically related to human behavior, physiology and pathology (Fisher & Marcus, 2006; Premack, 2007). The recent completion of the human and the chimpanzee genome projects and the advancement of the emerging field of comparative genomics provided investigators with tremendous tools to study these processes and phenomena and offered unique opportunities to answer questions related to human speciation (The Chimpanzee Sequencing and Analysis Consortium, 2005). In example, a major question in this line of research is associated with the identification of the class of genes that played the most prominent role in the acquisition of traits that are unique and specific to humans. However, the simple comparison between genomes such as that of the human

and the chimpanzee provides only limited information, given the relatively low nucleotide divergence between these species. Such analyses usually take into consideration the frequency of non synonymous nucleotide substitutions that occur between different genomes. The later refer to changes in the primary sequences of genes that result in differences in the corresponding amino-acids and thus, in the resulting protein products.

Therefore, novel comparative genomic approaches have to be developed aiming to reduce the number genes potentially associated with "human-specific" traits and facilitate their analysis.

Methods

Here, we have tried to apply such a simple comparative genomic-population genetics approach that takes into consideration both the rate of non synonymous nucleotide substitutions between chimpanzees and humans in a given gene and the degree of variation of single nucleotide polymorphisms (SNPs) in two remotely related human populations. Such remotely related human populations are expected to have only minimal interaction with each other over the course of human history and therefore, to have only restricted genetic exchanges. Our ap-

proach is based on the following consideration: Depending on the experimental approach and modeling used the divergence between human and chimpanzee is considered to have occurred between 4.7 million years (Horai et al., 1992) to around 7 million years ago (Disotell, 2006; Raam et al., 2005; Langergraber et al., 2012). Consistently with the relatively conservative estimation of 7 million years and assuming an average time span for each generation of about 20 years, implies a rough estimate of about 350,000 generations for the human species. Assuming that each gamete undergoes only one division per generation and that single nucleotide polymorphisms (SNPs) are due to *de novo* mutations in the human species, SNPs should develop at a rate of 1:350,000. This exceeds by a magnitude of more than 1000× the endogenous mutation rate that is around 5×10^{-9} mutations per nucleotide per cell division (Kumar & Subramanian, 2002). Thus, it is reasonable to conclude that SNPs must have been inherited from our evolutionary ancestors, and particularly the common ancestor between the human and the chimpanzee lineage and not developed after speciation has occurred.

Obviously, the aforementioned calculations reflect an oversimplification since age differences between spouses, different values between males and females or differences and alternative estimations on the mutation rate have not been taken into consideration. However, the selected values were utilized based on the assumption that they will indicate a degree of genetic heterogeneity “pushed” into the higher limits. To that end, it is conceivable that the incidence of non synonymous nucleotide substitutions (Ka) between human and chimpanzee orthologs will be associated with the variation of the allelic frequencies in closely located SNPs in human populations that are remotely related. Such human populations can be in example the Yoruba people in Ibadan, Nigeria (YRI) and US residents with ancestry from Northern and Western Europe, collected in 1980 by the Centre d’Etude du Polymorphisme Humain (CEU), for which SNP population data are publicly available from the HapMap project (www.hapmap.org, The International HapMap Consortium, 2007). Here, we have tested this hypothesis by analyzing 255 randomly selected genes (no specific criteria have been applied for their selection) in combination with the allelic frequencies of corresponding SNPs.

Results

Ka values from 255 randomly selected human-chimpanzee orthologs were obtained from previously reported studies (The Chimpanzee Sequencing and Analysis Consortium, 2005). Subsequently, through the ENSEMBL (<http://www.ensembl.org>) database the human orthologs were identified and one SNP for each gene was selected randomly. The SNPs utilized in our study fulfilled the following criteria in this specific order: 1) Availability of HapMap data for YRI and CEU populations. 2) Intronic localization. If intron-localized SNPs were not available, 5’-prime UTR or 3’-prime UTR located SNPs were selected. Subsequently, the SNPs were assessed through the HapMap (www.hapmap.org) site and the allelic frequencies for the YRI and CEU populations were obtained. The variability of the SNPs was assessed by calculating the standard deviation of the frequency of each SNP for the reference populations (allelic variation). Statistical analysis was performed by the chi-test.

In 255 human-chimpanzee orthologs, from chromosomes 1 - 14 for which Ka values are available, randomly selected SNPs were identified in the human populations under study and the

standard deviation in the allelic frequency was calculated for the ancestral allele (allelic variation). This value reflects the difference between the 2 populations with respect to this specific marker. Subsequently, each marker was classified with respect to whether this value, its allelic variation, was greater or smaller than the average value of the standard deviation of the frequency for the SNPs utilized in the present study. In example, for the marker rs11260587 allelic variation between YRI and CEU has been calculated to be 0.157684812. This value is greater than 0.118935 that corresponds to the average value of the allelic variations for all the SNPs used here. Alternatively, the allelic variation value for rs3813204 is 0.106066 that is smaller than the average value (Table 1).

A similar approach was followed to classify the corresponding orthologs as to whether Ka values were greater or smaller than the average Ka value. In example, the Ka value for the hypothetical protein FLJ20584 that is located closely to rs11260587, is 0.017, that is greater than 0.003945 that reflects the average of the corresponding Ka values used. The Ka value for hypothetical protein FLJ36119 is 0.0148 which is also greater than the average value (Table 1). Cumulative results are shown in Table 1 and whether Ka and allelic variation values for the same locus were both greater or smaller has been evaluated. As shown, among the 255 orthologs tested only 108 had both Ka and allelic variation values smaller or greater than average, while in the remaining 147 SNPs larger than average Ka values were associated with smaller allelic variance values and vice versa ($p < 0.000656$). Thus, it is apparent that less conserved genes between human and chimpanzees, and thus presumable genes of which some will be associated with human-specific traits, are characterized by a lower polymorphic rate in genetically linked SNPs. This is consistent with a bottleneck effect during human speciation according to which variation was restricted in human genes that are more important for the acquisition of the significant human-specific traits. Consistently with this notion, the later must have limited the polymorphic rate of the genes with an essential role in this process, and is of particular interest in the light of coalescent theory predictions suggesting that the age of most SNPs is smaller than about 100,000 generations (Hodgkinson et al., 2009).

Conclusion

Collectively, this study represents an initial attempt to view variation in SNPs between different human populations in association with comparative genomic data from chimpanzees and humans, with implications in human physiology, behavior and pathology. The association found between the degree of allelic variation in humans and the difference between adjacently located human and chimpanzee orthologs suggests that such analysis is potentially informative. Taking into consideration the variation between different human populations may

Table 1.

Cumulative results of the comparative analysis between allelic variation of SNPs in YRI and CEU population and the corresponding Ka values of adjacently located human and chimpanzee orthologs.

	Ka > average	Ka < average
Allelic deviation > average	27	73
Allelic deviation < average	74	81

bear significant value in understanding the genetic events that concluded with human speciation.

Acknowledgements

This study was supported by a grant from the Empirikion Foundation.

REFERENCES

- Disotell, T. R. (2006). "Chumanzee" evolution: The urge to diverge and merge. *Genome Biology*, 7, 240. <http://dx.doi.org/10.1186/gb-2006-7-11-240>
- Fisher, S. E., & Marcus, G. F. (2006). The eloquent ape: Genes, brains and the evolution of language. *Nature Reviews Genetics*, 7, 9-20. <http://dx.doi.org/10.1038/nrg1747>
- Hodgkinson, A., Ladoukakis, E., & Eyre-Walker, A. (2009). Cryptic variation in the human mutation rate. *PLoS Biology*, 7, e1000027. <http://dx.doi.org/10.1371/journal.pbio.1000027>
- Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S., & Takahaka, N. (1992). Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *Journal of Molecular Evolution*, 35, 32-43. <http://dx.doi.org/10.1007/BF00160258>
- Kumar, S., & Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 803-808. <http://dx.doi.org/10.1073/pnas.022629899>
- Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J. C., Muller, M. N., Robbins, M. M., Schubert, G., Stoinski, T. S., Viola, B., Watts, D., Wittig, R. M., Wrangham, R. W., Zuberbühler, K., Pääbo, S., & Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 15716-15721. <http://dx.doi.org/10.1073/pnas.1211740109>
- Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 13861-13867. <http://dx.doi.org/10.1073/pnas.0706147104>
- Raam, R. L. et al. (2005). Catarrhine primate divergence dates estimated from complete mitochondrial genomes: Concordance with fossil and nuclear DNA evidence. *Journal of Human Evolution*, 48, 237-257. <http://dx.doi.org/10.1016/j.jhevol.2004.11.007>
- The Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437, 69-87. <http://dx.doi.org/10.1038/nature04072>
- The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449, 851-861. <http://dx.doi.org/10.1038/nature06258>