Introgression from Gorilla caused the Human-Chimpanzee split

Johan Nygren
2018
Independent researcher

Corresponding author: Johan Nygren, johanngrn@gmail.com, Independent researcher

ABSTRACT: The Gorilla Genome Project (Scally, 2012) showed that 30% of the gorilla genome introgressed into the ancestor of humans and chimpanzees, and that the two species diverged through lineage sorting with 15% ending up in *Pan* and another 15% in *Homo*. That introgression is the *Pan-Homo* split, hybridization, which led to speciation as the new hybrid lineages became reproductively isolated from one another.

The NUMT on chromosome 5 ("ps5") (Popadin, 2017) fits perfectly with the introgression speciation model, it was formed from mtDNA that had diverged from the common ancestor of *Pan-Homo* for 1.8 Myr at the time of insertion into the nuclear genome, and originated in the *Gorilla* lineage. The ps5 pseudogene was transferred to *Pan* and *Homo* during the introgression event that led to the *Pan-Homo* split, 6 million years ago.

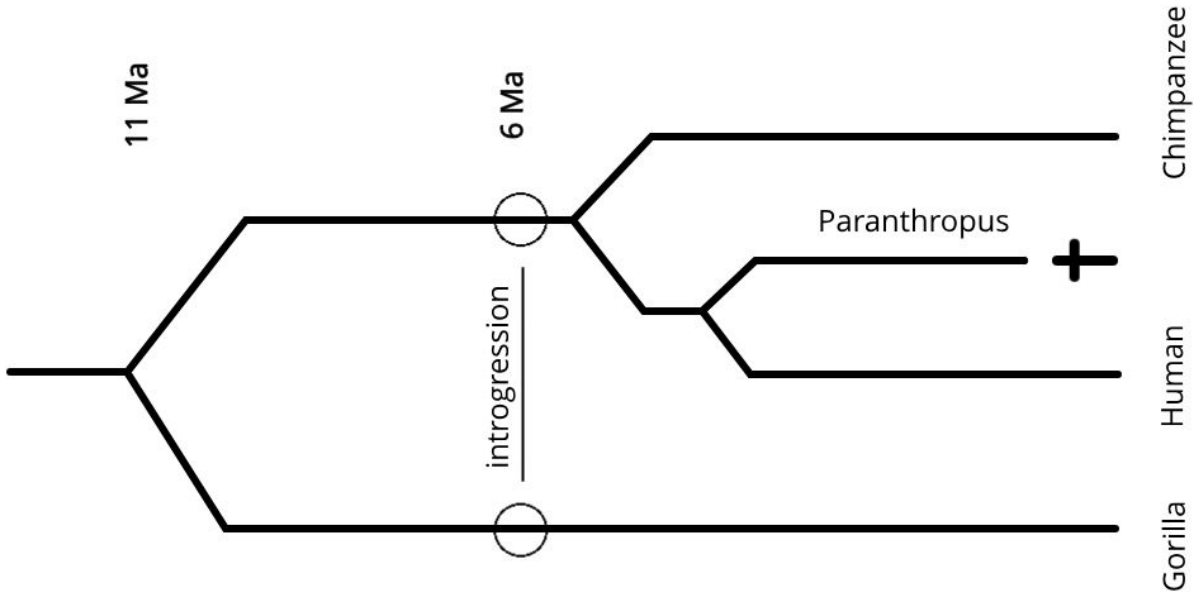

Fig. 1. Phylogenetic tree showing how introgression from *Gorilla* into the common ancestor of *Pan* and *Homo* led to hybridization, with multiple descendant lineages forming through lineage sorting.

Introduction

> *"During many years I collected notes on the origin or descent of man, without any intention of publishing on the subject, but rather with the determination not to publish, as I thought that I should thus only add to the prejudices against my views." - Charles Darwin, 1871*

Genome sequencing has been evolving along the law of accelerating returns (Kurzweil, 2004), the total amount of sequence data produced doubling approximately every seven months. (Stephens, 2015) With the genetic revolution, phylogenetic relationships are no longer limited to morphological characters, they can instead be read like an open book. This thesis will explore a new chapter, with roots in genetic data from the Gorilla Genome Project (Scally, 2012).

The Gorilla Genome Project was the first complete genome of *Gorilla*, from a female western lowland gorilla, and it revealed a closer relationship between humans and gorilla than what morphological analyses had shown: in 30% of the genome, gorilla is closer to human or chimpanzee than the latter are to each other. At the time interpreted as incomplete lineage sorting (Scally, 2012), genetic evidence of gene transfer between *Gorilla*, *Pan* and *Homo* around the time of the *Pan-Hom*o split (Popadin, 2017) shows that the lineage sorting is more parsimonious as a result of introgression.

Introgression

Introgression is the transfer of genetic information from one species into the gene pool of another by repeated backcrossing of an interspecific hybrid with one of its parent species.

Introgression may lead to speciation, in which the new hybrid lineages become reproductively isolated from parental populations (Baack, 2007), and since *Pan* and *Homo* have diverged through lineage sorting, with 15% of the introgressed genes ending up in *Pan* and another 15% in *Homo*, it is reasonable to conclude that the introgression caused the *Pan-Homo* split (Fig. 1), and therefore that it occurred at the time of the *Pan-Homo* split, around 6 million years ago.

Ps5

In early screening of mitochondrial pseudogenes within the human genome, a pseudogene sequence on chromosome 5 was discovered (Li-Sucholeiki et al., 1999), which later turned out to be a large (~9kb) NUMT, tentatively called "ps5" (Popadin, 2017). With advances in genome sequencing of *Gorilla* and *Pan*, the same ~9kb pseudogene sequence was discovered at homologous chromosomal positions in both those lineages, while it was absent in *Pongo*.

The pseudogene, when compared to mitochondrial branches of *Gorilla*, *Pan* and *Homo*, is shown to have diverged between the three lineages not at the *Gorilla/Pan-Homo* split, rather at the *Pan-Homo* split (Popadin, 2017), clear evidence that there was gene transfer between the three lineages at that time.

The ps5 pseudogene shares affinities with the gorilla lineage mtDNA (Popadin, 2017) which suggests that it originated in the gorilla lineage. With the probability of a NUMT insertion being unaffected by hybridization, it is clear that the insertion happened prior to the introgression event, and that the pseudogene had been evolving in the gorilla lineage for a period of time before introgressing into *Pan* and *Homo*. (Popadin, 2017)

With high availability of genetic data for both mitochondrial DNA and the pseudogene sequence, the exact history of ps5 can be read by comparing mutations within all three lineages.

The ratio of synonymous to non-synonymous mutations is a marker to distinguish between coding and non-coding gene sequences, because non-synonymous mutations are selected against until the gene is inactivated (Tomoko, 1995). For the "stem" of the ps5 pseudogene (the mutations that have accumulated prior to its divergence into three lineages), the fraction of coding ("mitochondrial") mutations to non-coding ("pseudogenic") mutations is 3/4 (Popadin, 2017).

The mutation rate in the mitochondrial genome is significantly higher than in the nuclear genome, which means that the 25% pseudogenic mutations have needed proportionally longer time to accumulate. With the estimate of 10x higher mutation rates in mtDNA (Brown, 1979), and 3x more "mitochondrial" mutations, it took 3.3x longer to accumulate the "pseudogenic" mutations, giving a rough estimate of the insertion happening at 1.8 Myr after the *Gorilla/Pan-Homo* split, 4.2 Myr before the introgression event that led to the *Pan-Homo* split.

Method

Phylogenetic relationships can be read from genome comparison. That there was gene transfer between *Gorilla*, *Pan* and *Homo* around the time of the *Pan-Homo* split can be read from a NUMT on chromosome 5 (ps5), which diverged between *Gorilla*, *Pan* and *Homo* at the time of the split. (Popadin, 2017) The ps5 NUMT as evidence of gene flow shows that introgression is a more parsimonious explanation for the lineage sorting from *Gorilla* than incomplete lineage sorting (ILS), and since *Pan* and *Homo* diverged through lineage sorting, it can be read that the introgression caused the *Pan-Homo* split. (Fig. 1)

Results

The lineage sorting of 30% of the gorilla genome that is seen in humans and chimpanzees (Scally, 2012) is a result of introgression, an event that caused the speciation of *Pan* and *Homo* (Fig. 1), and the two lineages diverged through lineage sorting with 15% of the introgressed genes ending up in *Pan* and another 15% in *Homo*.

Discussion

The indisputable evidence that an introgression event caused the speciation of *Pan* and *Homo* is made possible by the genome revolution, and it provides a map, a reference frame, that makes it

possible to read the world in ways that were previously out of sight, and can provide an important reference for continued research into hominin evolution. The fossil record shows that there were multiple lineages of hominin coexisting throughout the Pliocene and Pleistocene (Haile-Selassie, 2016), and the introgression speciation model can provide clues to how those lineages relate to one another. What remains to be understood is what environmental and ecological factors triggered the hybridization.

References

Baack, E. J., & Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development, 17(6), 513–518. https://doi.org/10.1016/j.gde.2007.09.001

Brown, Wesley & Jr George, M & Wilson, A.C.. (1979). Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America. 79. 3246-3250.

Haile-Selassie, Y., Melillo, S. M., & Su, D. F. (2016). The Pliocene hominin diversity conundrum: Do more fossils mean less clarity? Proceedings of the National Academy of Sciences, 113(23), 6364–6371.https://doi.org/10.1073/pnas.1521266113

Kurzweil, R. (2004). The Law of Accelerating Returns. In Alan Turing: Life and Legacy of a Great Thinker (pp. 381–416). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-05642-4_16

Li-Sucholeiki, X.-C., Khrapko, K., André, P. C., Marcelino, L. A., Karger, B. L., & Thilly, W. G. (1999). Applications of constant denaturant capillary electrophoresis / high-fidelity polymerase chain reaction to human genetic analysis. Electrophoresis, 20(6), 1224–1232. https://doi.org/10.1002/(sici)1522-2683(19990101)20:6<1224::aid-elps1224>3.0.co;2-o

Popadin, K., Gunbin, K., Peshkin, L., Annis, S., Fleischmann, Z., Kraytsberg, G., … Khrapko, K. (2017, May 9). Mitochondrial pseudogenes suggest repeated inter-species hybridization in hominid evolution. Cold Spring Harbor Laboratory. https://doi.org/10.1101/134502

Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., … Durbin, R. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature, 483(7388), 169–175. https://doi.org/10.1038/nature10842

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., … Robinson, G. E. (2015). Big Data: Astronomical or Genomical? PLOS Biology, 13(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Tomoko, O. (1995). Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. Journal of Molecular Evolution, 40(1), 56–63. https://doi.org/10.1007/bf00166595