6 My-old hybridization among human ancestors separated by ~4.5 My of evolution

as inferred from analysis of mtDNA pseudogenes (NUMTs).

Konstantin Popadin[1], Konstantin Gunbin[2], Leon Peshkin[3], Sofia Annis[4], Zoe Fleischmann[4], Natasha Markuzon[5], Yevgenya Kraytsberg[3], Rebecca Ackermann[6], and **Konstantin Khrapko[4] (kkhrapko@gmail.com)**

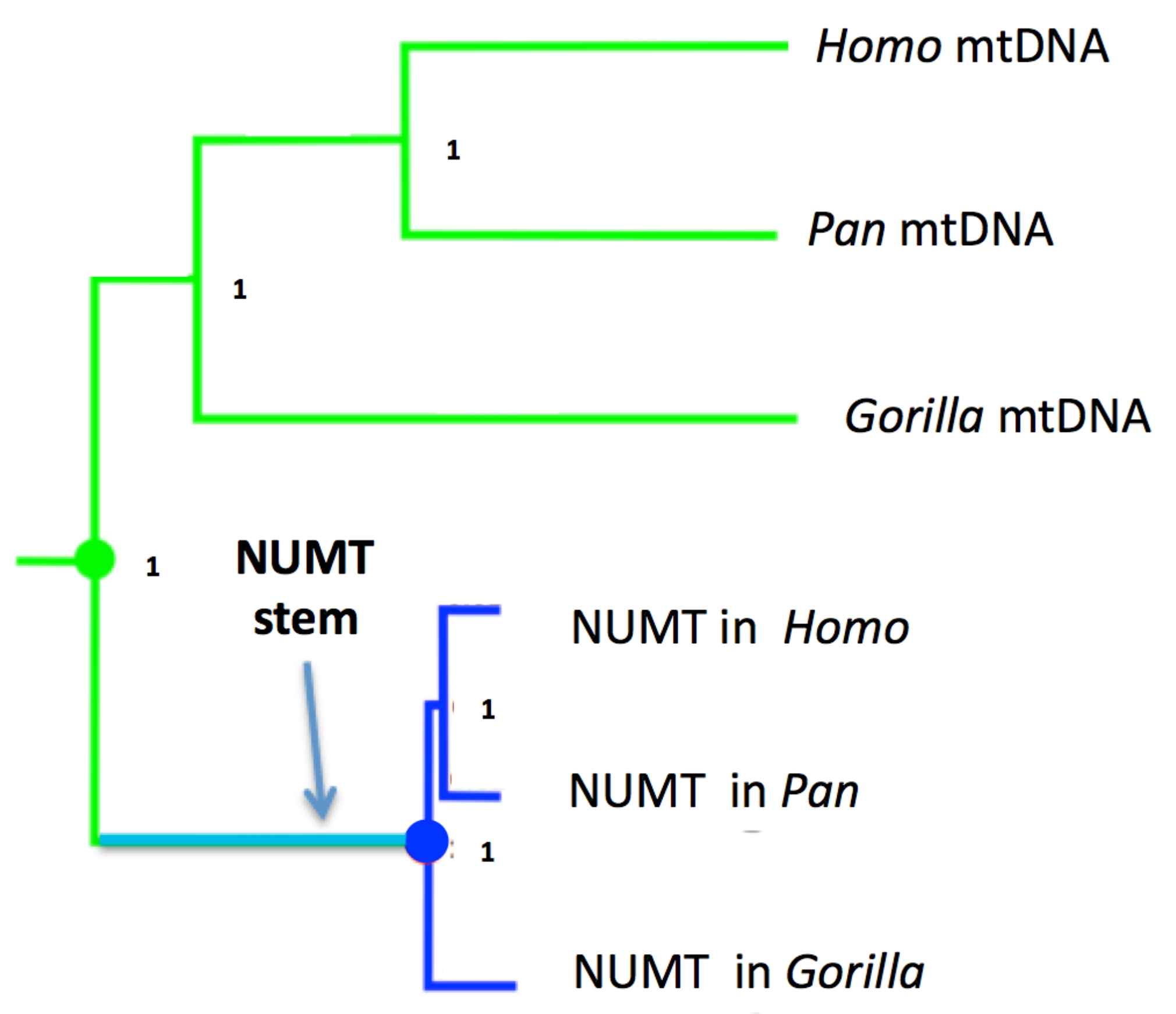
[1] University of Lausanne, Switzerland, [2] Novosibirsk State University, Russia, [3] Harvard Medical School, Boston USA,
[4] Northeastern University, Boston USA, [5] Draper Laboratory, Boston, USA, [6] University of Cape Town, South Africa

Abstract:

Increasingly, human evolution is being tied to genetic exchange between divergent lineages (e.g., Neanderthals, Denisovans). However, less is known about genetic exchange during earlier human evolution (pre-1Ma) and between more divergent lineages (a few My apart).

We use the analysis of nuclear pseudogenes of mtDNA ("NUMTs") to look for earlier, more distant hybridization events. **NUMTs** are fragments of mtDNA that were inserted into nuclear genome (a few hundreds of them have been permanently inserted into our nuclear genome over the last 60 My). NUMTs are considered "mtDNA fossils", as they preserve sequences of ancient mtDNA and thus carry unique information about ancestral populations. Our analysis of a NUMT sequence shared by humans, chimpanzees, and gorillas implies that, around the time of divergence between humans and chimpanzees, our ancestors interbred with hominines whose mtDNA at that time had diverged as much as ~4.5My from human-to-be mtDNA, strongly implying distant hybridization. Analysis of other NUMTs suggests that such events might have occurred more than once. Reassuringly, interspecies hybridization across such large genetic distances have been observed in other primates, e.g. baboons and colobines. We hypothesize that hybridization might have provided the diversity that promoted evolution of our lineage, and could explain the apparent saic morphology of some fossils of that epoch.

Source publication: Gunbin, K. Popadin, K. Peshkin, L., Annis, S*, Markuzon, N., Kraytsberg, Y., Ackermann, R., **Khrapko, K.** (2017). Mitochondrial pseudogenes suggest repeated inter-species hybridization in hominid evolution. ***BioArxiv,*** doi: https://doi.org/10.1101/134502



🟩 mtDNA mutations (mostly synonymous)
🟦 Pseudogene nDNA mutations (highly non-synonymous)
🟦 NUMT stem mutations (~75:25 mtDNA:nDNA synonymity)

Fig. 1. Main observation:

The primary subject of this study is a NUMTs on human chromosome 5, which is shared with chimpanzee and gorilla. A joint phylogenetic tree of the three NUMT variants and the ape mtDNA sequences reveals a very long **NUMT stem** and short branches of the NUMT subtree, in drastic contrast to the proportions of the mtDNA sub-tree. As expected for a dysfunctional sequence, mutations in the outer pseudogene branches (colored blue) contain a significantly higher proportion of non-synonymous changes ($p<0.00005$) than mutations of mtDNA branches (green). lintriguingly, mutations of the **NUMT stem** contain a very high proportion of synonymous changes, as if mutations of mtDNA branches were mixed with mutations of NUMT branches at 75:25 ratio (therefore colored teal).

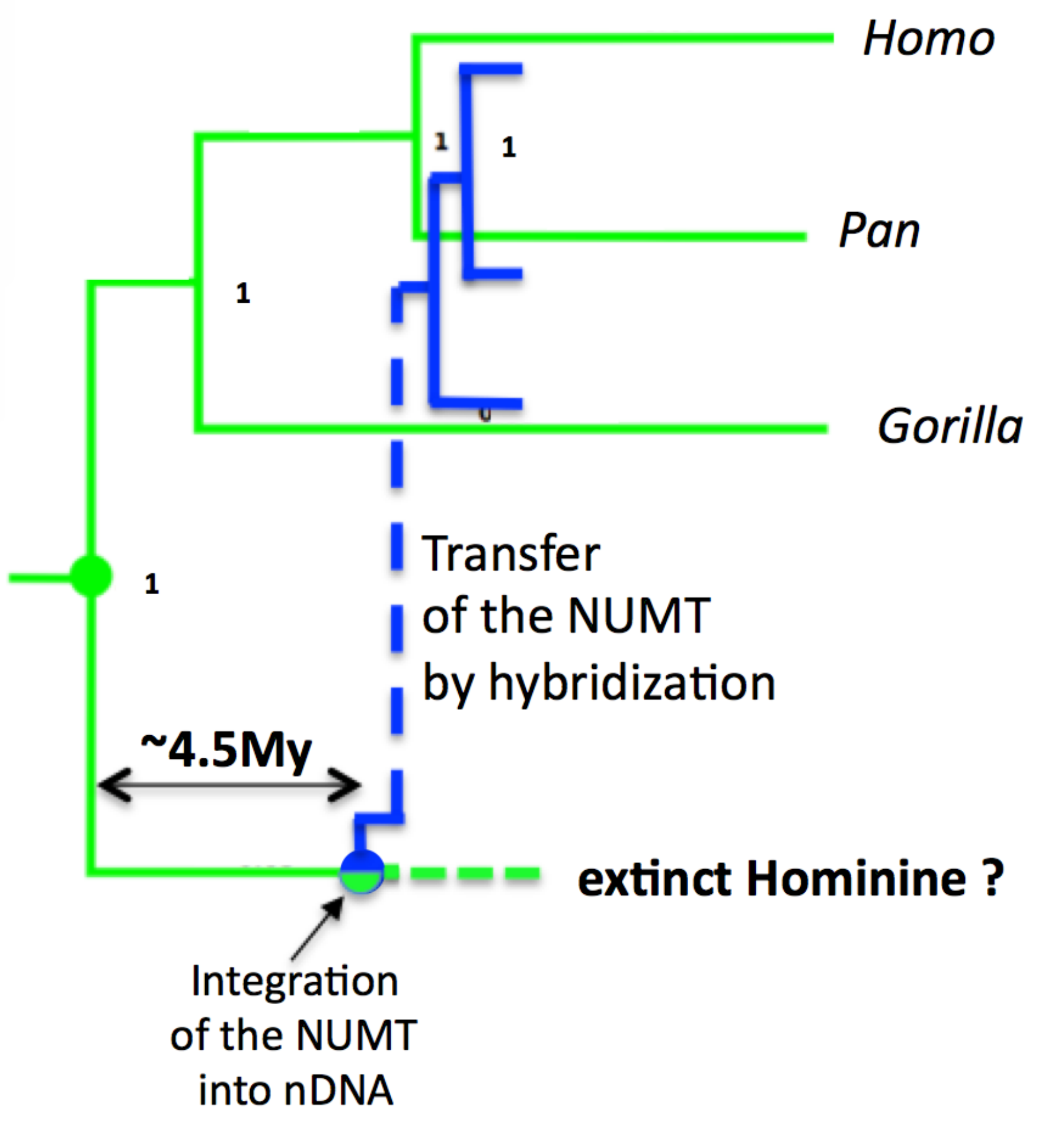


Fig. 2. Interpretation:

We hypothesize that mixed synonymity of mutations of the **NUMT stem** implies that NUMT sequence has been evolving as a functional mitochondrial genome, then turned into a NUMT, which then was inherited by the Homo, Pan, and Gorilla. The number of mtDNA mutations in the stem is equivalent to ~4.5 My of divergence. Thus, mtDNA lineage leading to the NUMT should have been diverging for 4.5My from the human-to-be mtDNA lineage. This implies that this pseudogene should have been created in a hominine (now extinct) that at that time had been diverged by about 4.5My of evolution from the hominine that at that time carried our mtDNA lineage. In order for this pseudogene and our mtDNA to end up in the same body, these two hominids should have mated. The large divergence implies that such mating was a distant interspecies (or even inter-generic) hybridization.

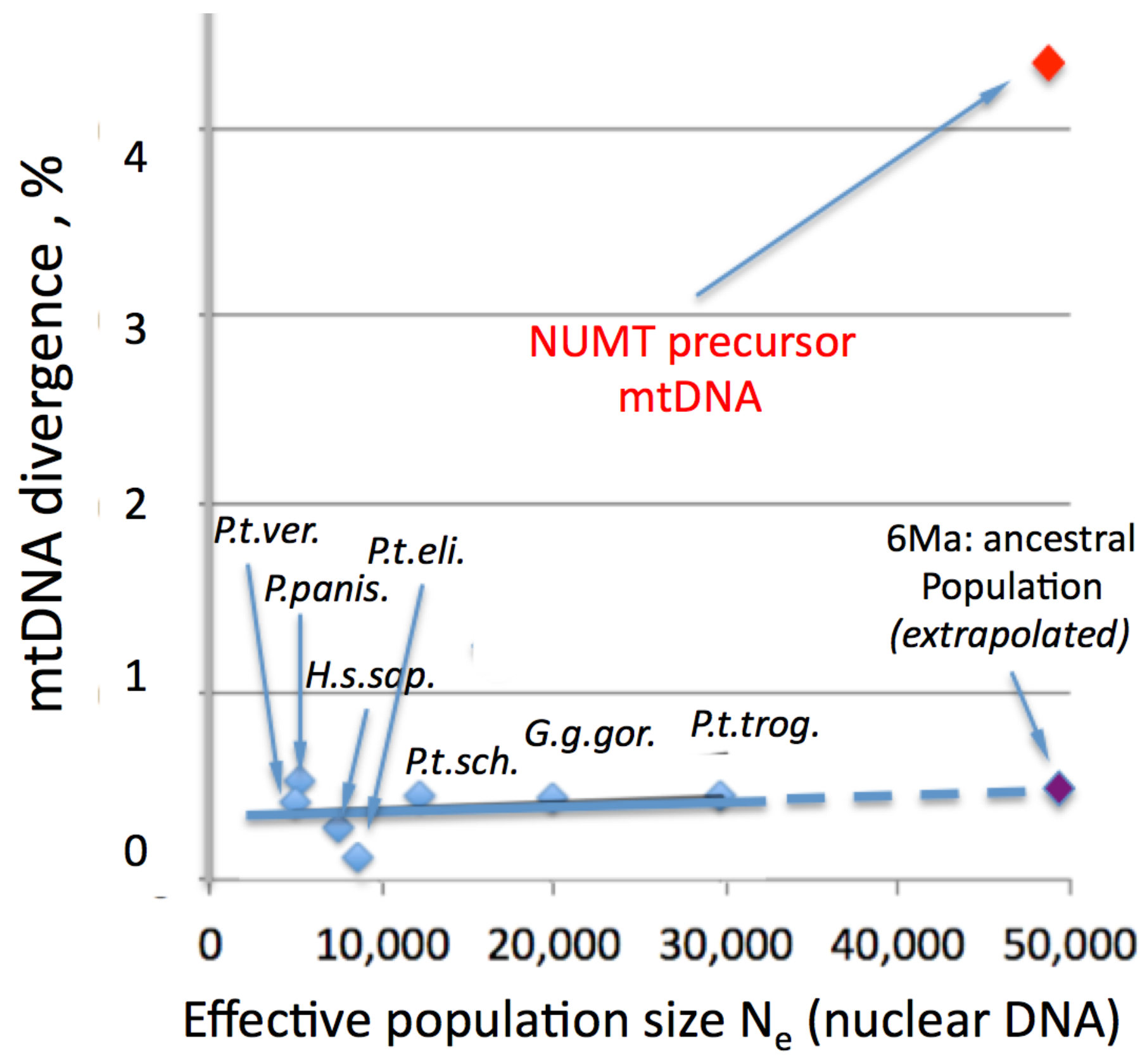


Fig. 3. Large ancestral N_e is not an explanation:

A potential alternative explanation of the high divergence of the mtDNA precursor of the NUMT from human-to-be mtDNA could be the high effective population size $\mathbf{N_e}$ of the ancestral population. In this scenario, genetic heterogeneity in population is so large that a divergent individual could be a regular member of the population, rather than an intruder from a distant species. Indeed, the ancestral $\mathbf{N_e}$ is believed to have been larger than $\mathbf{N_e}$ of modern ape populations. To explore how much mtDNA divergence depends on $\mathbf{N_e}$, we plotted (**Fig 3**) the observed maximal mtDNA divergence in present ape populations vs. $\mathbf{N_e}$ of the same populations (from Prado-Martinez et al., 2013). Strikingly, there is no correlation between $\mathbf{N_e}$ and mtDNA divergence. Moreover, even in the largest populations (e.g. common chimp, *P.t.trogl.*, $\mathbf{N_e}$=30,000, which is close to anticipated ancestral $\mathbf{N_e}$=50,000), mtDNA divergence is drastically lower than divergence of the NUMT precursor (red diamond), which discounts large $\mathbf{N_e}$ as an alternative explanation.

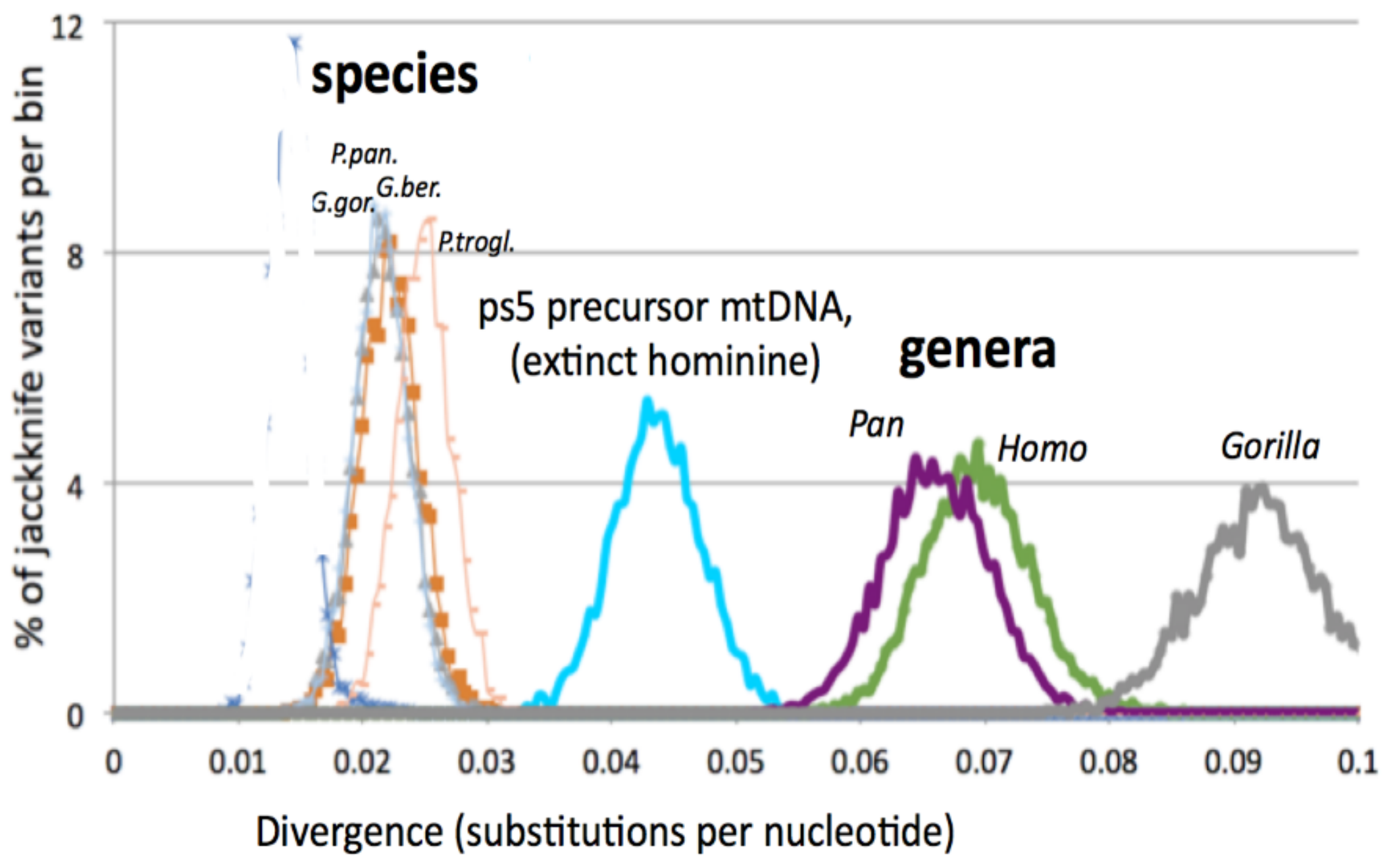


Fig. 4. Quantitation:

The estimated divergence of the mitochondrial portion of NUMT stem (representing the divergence of the hypothetical extinct hominine involved in the hybridization) lands in between the divergence between closely related species (Such as bonobo vs. common chimp) and divergence between genera (e.g. *Homo* and *Pan*).

Divergences were inferred using jackknifed ML phylogenic analysis (peak width represents the variance of the 50% jackknife). Phylogenies were constructed on the three NUMTs (human, chimp, gorilla) and 83 sequences of human, chimpanzee, gorilla and orangutan mtDNA from (Prado et al., 2013) combined with mitochondrial genomes of gibbon and baboon (outgroup), and two mitochondrial genomes of a Neanderthal and a Denisovan.

References:

Ackermann RR, Rogers J, Cheverud JM. (2006) Identifying the morphological signatures of hybridization in primate and human evolution. J Hum Evol; 51(6):632-645. doi:10.1016/j.jhevol.2006.07.009.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O 'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.

Wang, B., Zhou, X., Shi, F., Liu, Z., Roos, C., Garber, P.A., Li, M., and Pan, H. (2015). Full-length Numt analysis provides evidence for hybridization between the Asian colobine genera Trachypithecus and Semnopithecus. Am. J. Primatol. *77*, 901–910.

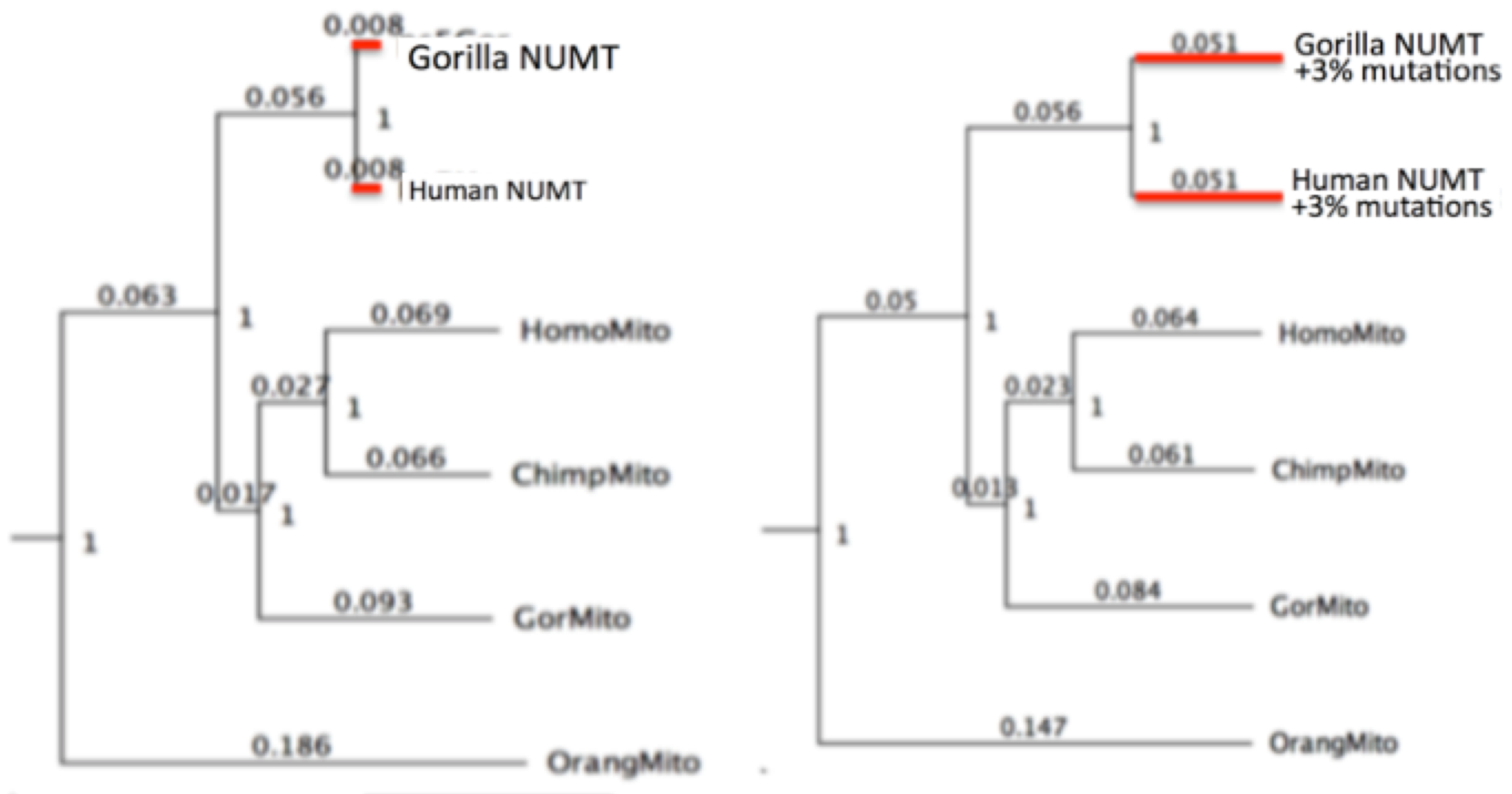


Fig. 5. Test for branch length artifacts:

short NUMT branches do not distort our estimates

We were concerned that unusually short outer branches of the pseudogene subtree might have artificially forced the ML algorithm to extend the NUMT stem in compensation. We therefore added extra mutations (transitions) into the pseudogene braches which made them essentially undistinguishable from the normal mitochondria branches by any parameters. Reassuringly, extension of branch lengths essentially did not affect the topology and topography of the remaining tree.

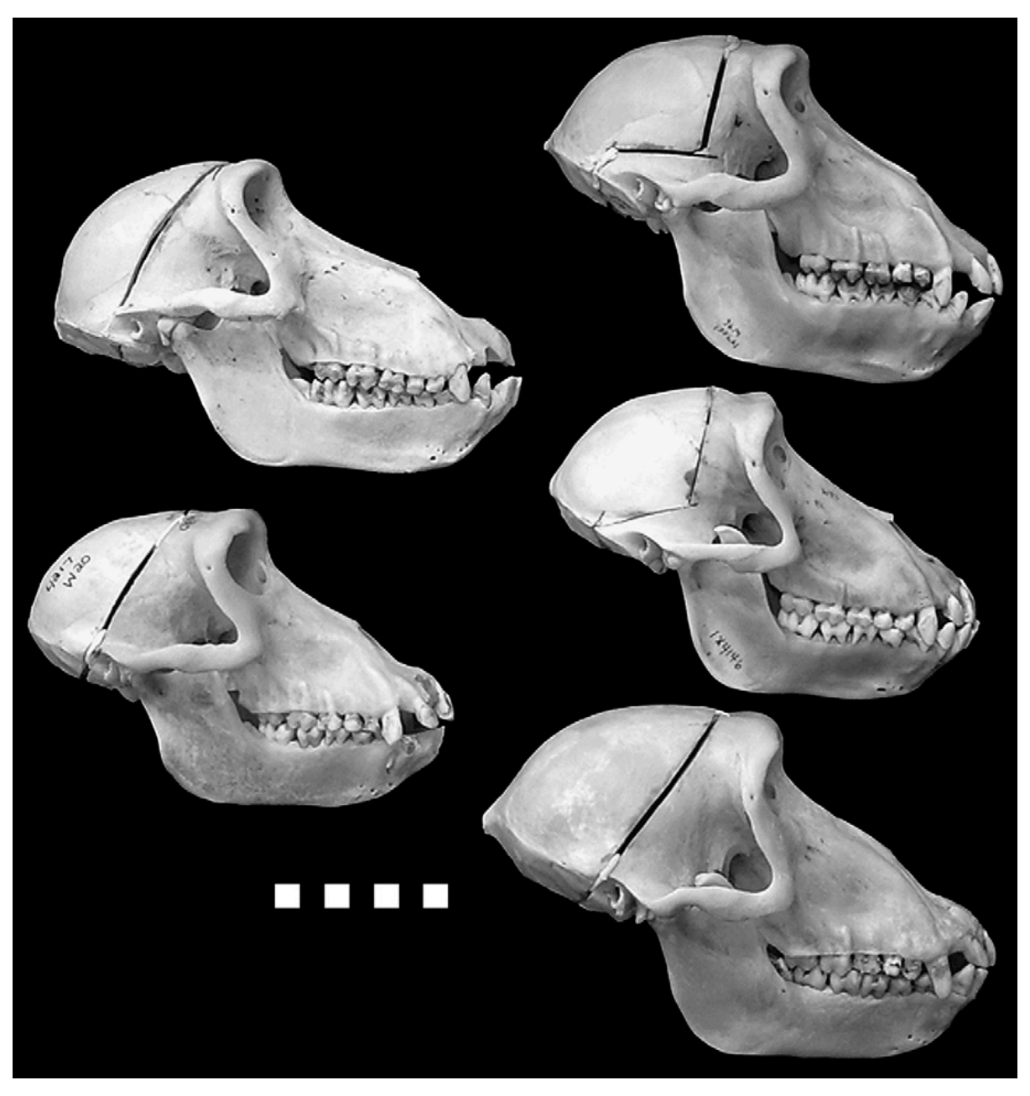


Fig. 6. Speculation:

Hybridization generated diversity which sparked hominin speciation

Of note, even a much less distant recent hybridization (~1My) between two baboon species produced a remarkable diversity of cranium sizes and shapes as seen in **Fig. 6** (Ackerman 2006). It is thus tempting to speculate that the more dramatic hybridization across 4.5 My of divergence around the time of human/chimpanzee separation generated the extra diversity that augmented the separation of the hominin lineage from the apes.

Discussion:

Hybridization across genetic difference of 4.5 My sounds too dramatic. However, such hybridizations are known in primates. E.g., intergenera hybridization in Colobine monkeys separated by ~5.5Ma (Wang 2015) involves a NUMT scenario similar to ours.

Interestingly, NUMT insertion/transfer happened around the time of the Homo/Pan split. Intriguingly, certain hominin fossils of that epoch have been interpreted alternately as more human-like or more ape-like. Such morphological **mosaicisity** could potentially be explained by hybridization.

Fixing of NUMTs in population should have been rather efficient, since these pseudogenes appear to have been fixed in more than one population (human, chimp, Gorilla). Was the spread of the NUMT driven by selection? Indeed, NUMTs on chr5 is located in 3' region of a functional genes. Most intriguingly, another NUMT with a similar phylogeny is located 3' to a gene essential for spermatogenesis. That NUMT might have served as an expression modifier, resulting in a reproductive advantage. Indeed, RNF141 demonstrates selectively driven expression shift in testis of the ancestor of hominines.