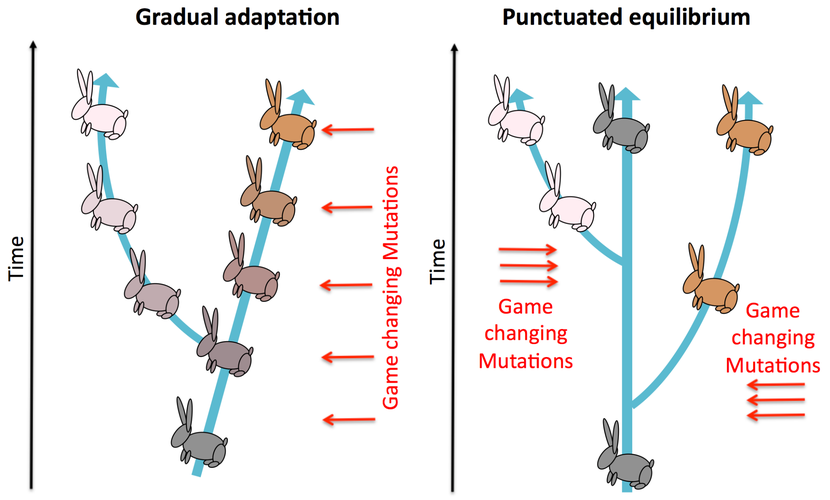
# Human evolution: Gradual or Punctuated?
## Accelerated integration of mtDNA pseudogenes coincides with emergence of the human genus. A Hypothesis.

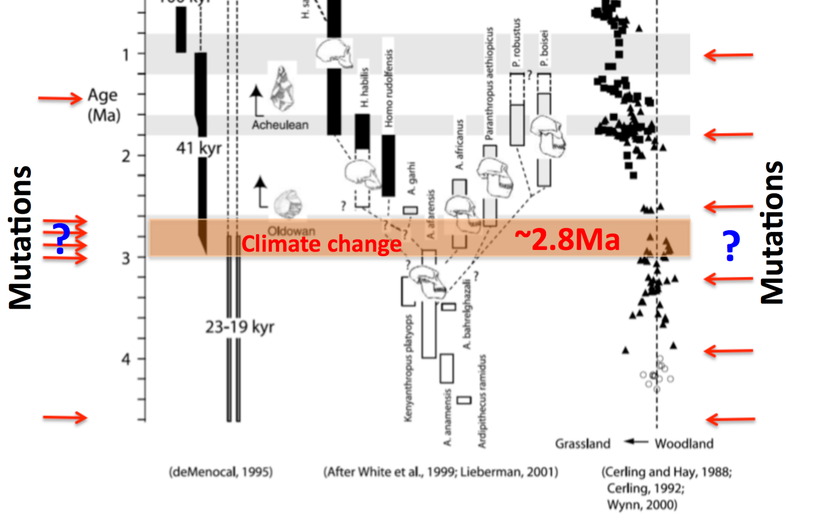
Konstantin Gunbin[1], Konstantin Popadin[2], Leon Peshkin[3], Sofia Annis[4], Rebecca Ackermann[5], and **Konstantin Khrapko[4]***

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

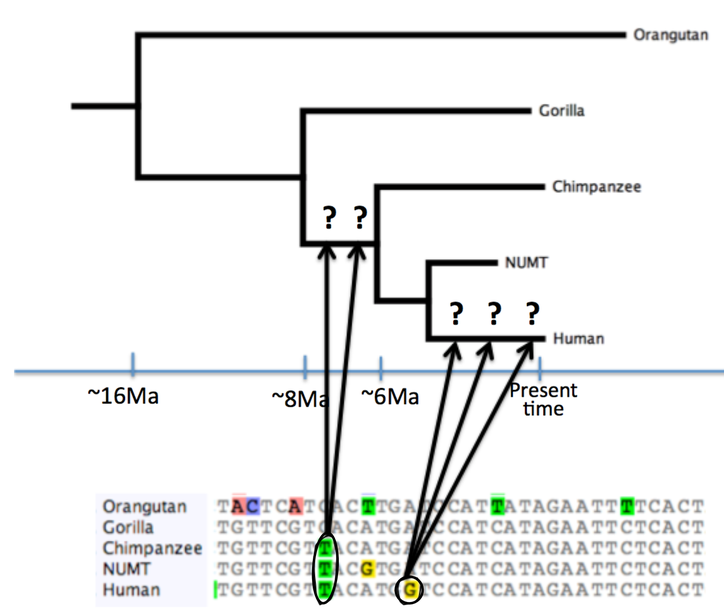
**Abstract**:
Fragments of mitochondrial DNA are known to get inserted into nuclear DNA to form NUMTs, i.e. nuclear pseudogenes of the mtDNA. The insertion of a NUMT is a rare event. Hundreds of pseudogenes have been cataloged in the human genome. NUMTs are, in essence, a special type of mutation, but, unlike conventional point mutations, NUMTs are mutations with their own internal timer, which is synchronized with an established molecular clock, the mtDNA. Thus insertion of NUMTs can be timed with respect to evolution milestones such as the emergence of new species. We asked whether NUMTs were inserted uniformly over time or preferentially during certain periods of evolution, as implied by the "punctuated evolution" model. Interestingly, the NUMT insertion times do appear nonrandom with at least one cluster positioned at around 2.8 million years ago (Ma). Interestingly, 2.8 Ma closely corresponds to the time of emergence of the genus Homo, and to a well-documented period of major climate change ca. 2.9–2.5 Ma. It is tempting to hypothesize that the insertion of NUMTs is related to the speciation process. NUMTs could be either "riders", i.e., their insertion could be facilitated by the overall higher genome flexibility during speciation, or "drivers", i.e. they may more readily get fixed in the population due to positive selection associated with speciation. If correct, the hypothesis would support the idea that evolution of our genus may have happened in a rapid, punctuated manner. Ref: (Gunbin et al. 2017).



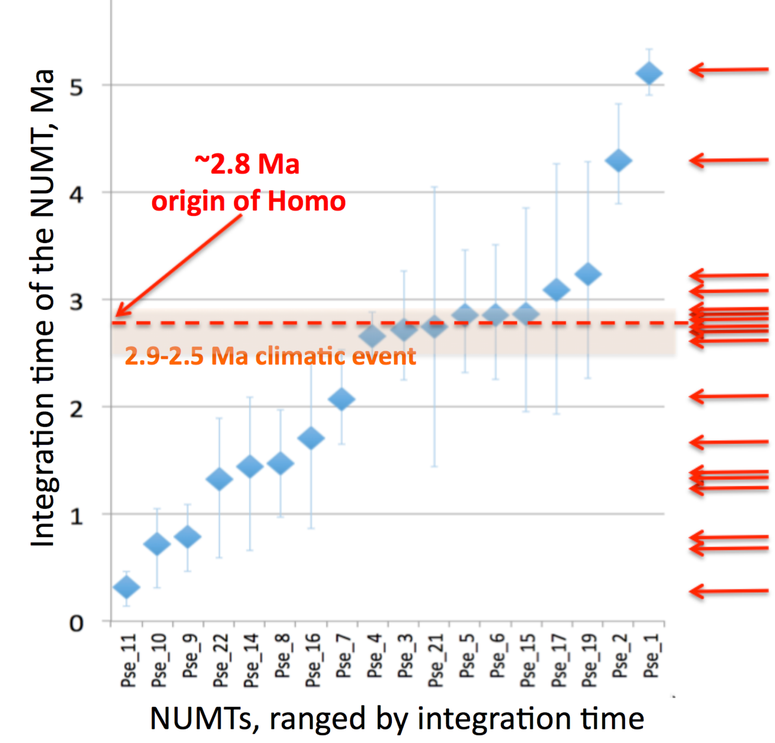

### Fig. 1. The General Question:
**Is evolution a process of gradual adaptation and continuous change of form/function OR a succession of rapid discontinuous shifts? And can this continuity/discontinuity be observed at the level of the rate of genetic change, i.e. is the rate of "critical" mutations constant or punctuated?**



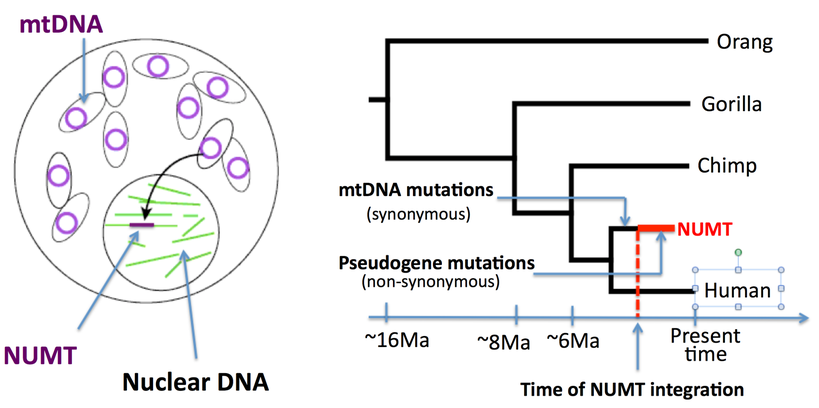

### Fig. 2. The Specific Question:
**Is *human* evolution a gradual process or a succession of rapid discontinuous changes?** This has been a focus of intense debate for decades. Of particular interest has been the mid-Pliocene climate change event ~2.9-2.5 Ma, which is thought to have precipitated the separation of the genus Homo as an independent lineage, i.e. separation of Humans from non-Humans (~2.8Ma). The debate mostly concerned continuity/punctuality of the fossil record, but of course the rate of the underlying genetic change is of ultimate interest/importance. Did hominid lineage experience an increased mutation rate when a large number of distinct hominins emerged and eventually gave rise to the split between *Australopitecus/Paranthropus* and *Homo*? Ref: (deMenocal 2004).




### Fig. 3. Technical Problem:
**Imprecise timing of conventional mutations**. The difficulty in answering the above questions lies in the way ancient mutations have to be timed. Conventional point mutations are assigned to specific branches of the DNA phylogenetic trees. The essence of the problem is that mutations can be located within branch segments from branching point to branching point, but the exact position within the segment is principally unknown. Because the hominid DNA-derived phylogenetic tree has very low branch density, the precision of mutation timing is low, e.g., human-specific mutations can be positioned within ~6 My from separation from chimpanzee till the branching of the Denisovans.




### Fig. 5. Results:
Hundreds of pseudogenes have been cataloged in the human genome that have been inserted over the last ~60 My of which we considered the last 6 My. Various quality filters resulted in the selection of 18 NUMTs most suitable for phylogenetic analysis. Insertion times of these 18 NUMTs appear non-randomly distributed with one cluster positioned around 2.8Ma. While timing of insertion of individual NUMTs is imprecise, the overall probability of forming such a cluster by chance is low, which makes this observation highly statistically significant.




### Fig. 4. Solution:
**NUMTs – mutations with an internal timer**. NUMTs are insertions of mtDNA sequences into the nuclear genome. Unlike point mutation, each NUMTs actually represents a branch on the mtDNA phylogenic tree and thus its time of insertion can be determined as precise as their branching point can be positioned on the tree. In a sense, NUMTs are "mutations with an internal clock", which is synchronized with the well-established mtDNA mutation evolution clock. By determining the NUMTs' insertion time points, one can ask whether NUMTs were inserted uniformly over time or preferentially during certain periods of evolution, as implied by the "punctuated evolution" model.

A technical peculiarity is that a NUMT does not necessarily insert into the nuclear genome at the time NUMT branching. The mtDNA ancestor of a NUMT might have been evolving as an independent mtDNA lineage prior to insertion, therefore NUMT branch may consist of the mitochondrial and the pseudogenic segments (black and red in the figure above). The proportion of the mitochondrial segment and thus the time of integration may be estimated based of the synonymity of the mutations of the branch, which ought to be a linear combination of high synonymity mitochondrial mutations, which are under selective pressure and low synonymity of the pseudogene mutations (which are not).

### References:
Gunbin, K.V., Popadin, K., Peshkin, L., Ackermann, R.R., and Khrapko, K. (2017). Integration of mtDNA pseudogenes into the nuclear genome coincides with critical periods in human evolution. A Hypothesis. *Mitochondrion* 34, pp. 20-23, doi: 10.1016/j.mito.2016.12.001 (2017).

deMenocal, P. B. (2004). African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, *220*(1-2), 3–24. Elsevier.

Assumed Human/Chimpanzee divergence time

| Event Interval width | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 0.2 | 0.453 | **0.021** | **0.022** | 1.000 | 1.000 | 1.000 | 0.460 | 1.000 | 1.000 |
| 0.4 | 0.340 | **0.005** | **0.001** | 0.713 | 1.000 | 0.711 | 0.711 | 1.000 | 0.712 |
| 0.6 | 0.549 | **0.000** | **0.007** | 0.266 | 0.850 | 0.851 | 0.850 | 0.552 | 0.849 |
| 0.8 | 0.083 | **0.001** | **0.006** | 0.080 | 0.712 | 0.924 | 0.924 | 0.713 | 0.923 |
| 1 | **0.006** | **0.005** | **0.020** | **0.021** | 0.594 | 0.964 | 0.829 | 0.828 | 0.598 |
| 1.2 | **0.016** | **0.017** | **0.016** | **0.050** | 0.502 | 0.902 | 0.903 | 0.903 | 0.498 |
| 1.4 | **0.040** | **0.040** | **0.012** | **0.041** | 0.107 | 0.828 | 0.946 | 0.635 | 0.632 |
| 1.6 | 0.079 | 0.079 | **0.030** | 0.079 | 0.079 | 0.548 | 0.748 | 0.747 | 0.549 |
| 1.8 | 0.061 | 0.058 | 0.060 | **0.020** | 0.140 | **0.020** | 0.665 | 0.667 | 0.669 |
| 2 | **0.043** | 0.107 | **0.043** | **0.043** | **0.045** | **0.045** | 0.225 | 0.772 | 0.768 |

### Fig. 6. Statistical significance:
Significance was estimated by simulation. We generated $10^6$ 18–point random data sets drawn from the uniform distribution between 0 and 6 Ma. Then we calculated the fraction of 18-point sets that were "as highly clustered or more extremely clustered than the distribution of NUMT insertion times" (Fig. 5), i.e., the number of time points that landed within 2.5-2.9 Ma interval was >= 6. The corresponding probability (p-value proxy) is **0.0006**. This impressively low probability, however, depends on the interval width and the assumed Human/Chimpanzee divergence time (because we use it to calibrate the mtDNA clock). We therefore repeated these estimates with different widths of intervals centered at 2.8 Ma and with different divergence times (5, 5.5, …, 9 Ma). The results of this analysis are shown in **Figure 6**: p-value stayed well below the critical 0.05 value within a wide range of interval widths and human/chimp divergence times. At the same time, these results put limitations on our conclusions. For example, if divergence time between humans and chimpanzees were as ancient as 8Ma, then our data are not sufficient to statistically support temporal association of accelerated NUMT insertion into nuclear DNA with the emergence of Homo and/or climatic change 2.8Ma.

### Discussion:
It is tempting to hypothesize that accelerated insertion of NUMTs is somehow linked to the *Homo* speciation process. NUMTs could be either **"riders"**, i.e., their insertion could be facilitated by the overall higher genome flexibility during the speciation period (if such extra flexibility indeed exists), or **"drivers"**, i.e. they may more readily get fixed in the population during the period of speciation due to increased selective pressures.
If correct, the hypothesis of accelerated pseudogenization would support the idea that evolution of our genus might have been a punctuated process.