

The Estimated Pacemaker for Great Apes Supports the Hominoid Slowdown Hypothesis

Beatriz Mello  and Carlos G Schrago

Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Evolutionary Bioinformatics
Volume 15: 1–8
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934319855988



ABSTRACT: The recent surge of genomic data has prompted the investigation of substitution rate variation across the genome, as well as among lineages. Evolutionary trees inferred from distinct genomic regions may display branch lengths that differ between loci by simple proportionality constants, indicating that rate variation follows a pacemaker model, which may be attributed to lineage effects. Analyses of genes from diverse biological clades produced contrasting results, supporting either this model or alternative scenarios where multiple pacemakers exist. So far, an evaluation of the pacemaker hypothesis for all great apes has never been carried out. In this work, we tested whether the evolutionary rates of hominids conform to pacemakers, which were inferred accounting for gene tree/species tree discordance. For higher precision, substitution rates in branches were estimated with a calibration-free approach, the relative rate framework. A predominant evolutionary trend in great apes was evidenced by the recovery of a large pacemaker, encompassing most hominid genomic regions. In addition, the majority of genes followed a pace of evolution that was closely related to the strict molecular clock. However, slight rate decreases were recovered in the internal branches leading to humans, corroborating the hominoid slowdown hypothesis. Our findings suggest that in great apes, life history traits were the major drivers of substitution rate variation across the genome.

KEYWORDS: evolutionary rates, molecular dating, genome evolution

RECEIVED: February 26, 2019. **ACCEPTED:** May 17, 2019.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Beatriz Mello, Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, CCS, A2-092, Rua Prof. Rodolpho Paulo Rocco, SN, Cidade Universitária, Rio de Janeiro RJ 21941-617, Brazil. Email: biamello@gmail.com

Introduction

The mode of evolution of genes/genomic regions has been the subject of many studies since the molecular clock hypotheses was first proposed in the 1960s.^{1–6} The strict molecular clock assumption of homogeneity in substitution rates across lineages has been frequently demonstrated as a poor model to describe the observed evolutionary pattern for several portions of the genome^{7–11} (Figure 1A). Alternative models alleviated the assumption of rate constancy across distinct branches of a phylogenetic tree,^{3,4,12} accommodating variation of evolutionary rates that may be due to *lineage effects*. These effects are defined by any source of variation in evolutionary rates driven by life history traits, such as generation times.^{13,14} More recently, the availability of genomic data prompted the investigation of rate variation across the genome, bringing up new theoretical challenges.^{15–21} Presumably, selective regimes that are not shared among loci generate distinct patterns of evolutionary rate variation among genes. Moreover, differences in spontaneous mutation rates across the genome may also impact between-gene rate variation.²² As a consequence, evolutionary trees inferred from different genes (gene trees) may not be overlaid, which means that branch lengths are not proportional among gene trees. Therefore, evolutionary rates will vary between loci due to the so-called gene effects.^{14,22}

Previous analyses have suggested that selection is not the main force driving the evolution of genes, since a universal pacemaker (UPM) exists during the evolution of most genes in the genome^{18,19,23} (Figure 1B). This finding highlighted the importance of lineage effects in shaping the evolution of

substitution rates across the genome. Thus, the evolutionary rate may change across lineages, driven by fluctuations in effective population size, generations times, and other life history traits. These consist of top-down factors with homogeneous effects across the genome. Therefore, differences between gene trees, with branch lengths measured in substitutions per site, would be modulated by scaling factors that reflect the intensity of natural selection, determining how fast a genomic portion evolves compared with others. The UPM was suggested for diverse biological groups, such as bacteria, archaea, *Drosophila*, and yeast.^{18,19,23}

In mammals, a variation of the UPM was proposed, the multiple pacemaker (MPM)^{14,20} (Figure 1C). In this case, several pacemakers govern the evolution of genomic regions, as if distinct classes of pacemakers acted along the evolution of the genomes. The most extreme case of the MPM would be the degenerate multiple pacemaker (DPM), in which each genomic portion has its own evolutionary dynamics²⁰ (Figure 1D). Here, genomic evolution is mostly dictated by *residual effects*, which represents the interaction between gene and lineage effects.^{20,22} DPM has not yet been detected for any biological group, suggesting that life history traits have an effect on rate evolution to some extent.

Although several works have investigated the variation of the mode of rate evolution along genomes, a comprehensive analysis of pacemaker model for all great apes has never been implemented. For instance, studies have reported the slowdown on the substitution rates in the human lineage, but never compared this hypothesis with pacemaker models.^{24–27} Moreover, the



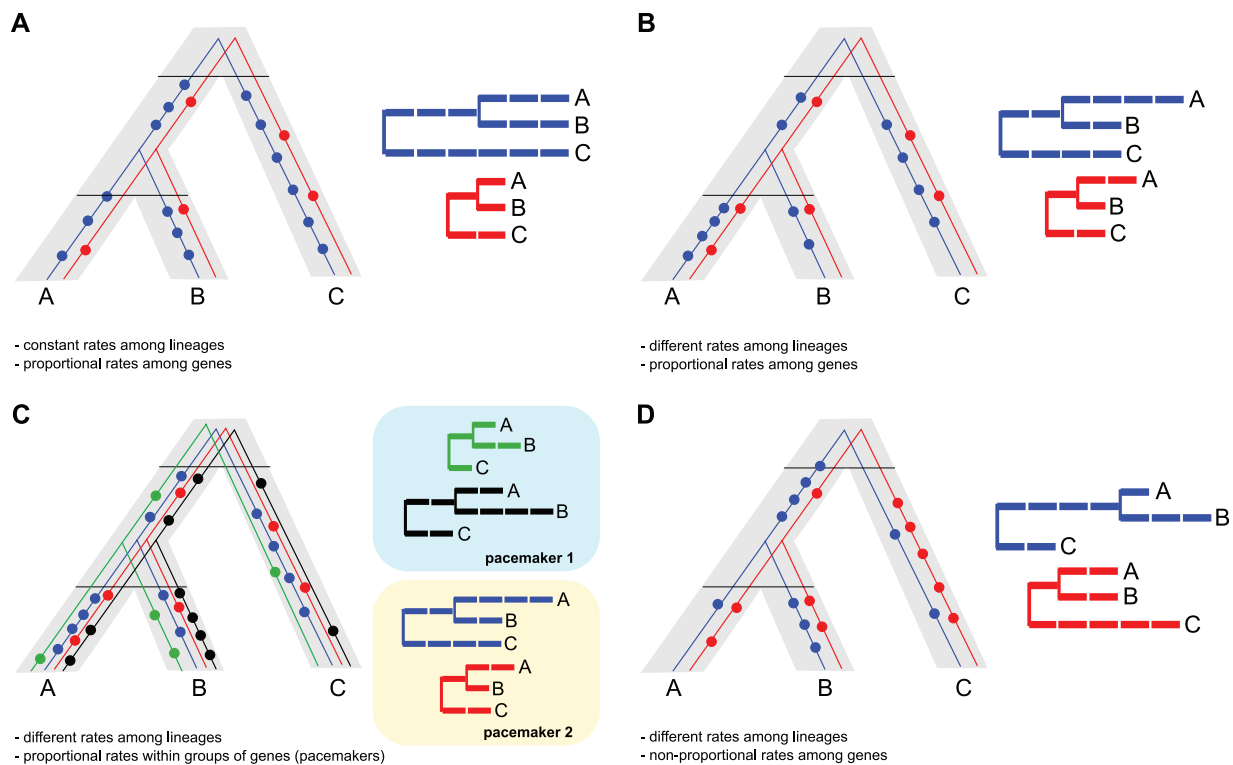


Figure 1. Scenarios of substitution rate variation among lineages and loci. The gray area represents the phylogeny of species A, B, and C, with speciation times defined by the horizontal lines. The genealogy of each gene (gene tree) is shown inside phylogenies with different colors. Nucleotide/ amino acid substitutions in each gene are depicted by colored circles. In (A) a strict molecular clock governs the evolution of genes; (B) a single pacemaker exists, in which evolutionary rates are not constant, but they are proportional between genes; (C) multiple pacemakers; (D) rates differ among genes and are not proportional between lineages, the degenerate pacemaker.

heterogeneity of gene tree topologies should be taken into account when estimating pacemakers, instead of enforcing the species phylogeny, as previously done to infer mammalian pacemakers.²⁰ Enforcing the same tree topology for different regions across the genome will bias the estimation of evolutionary rates,^{28,29} resulting in spurious assignments of substitutions that stretch branch lengths in an attempt to accommodate topological discordance. This was denominated by SPILS (Substitutions Produced by ILS).²⁸ Thus, failure to account for gene trees/species tree topological discordance potentially leads to a biased recognition of the number and models of pacemakers, as it relies upon an accurate estimate of branch lengths and, consequently, substitution rates.

Another factor that complicates the inference of evolutionary pacemakers is that coalescence times of gene trees may differ significantly from speciation times due to the stochasticity of the coalescent process along the species tree.³⁰ Even under the strict molecular clock, gene trees will likely not exhibit proportional branch lengths (Figure 2). This may ultimately yield spuriously inferred pacemakers if relative branch lengths, instead of evolutionary rates, are used to estimate the pacemaker model.²⁰ Therefore, population-level processes have a twofold influence on pacemaker evaluation: gene trees/species tree topological discordance should be accounted for; and, the

decoupling between gene coalescence times and speciation times must be explicitly incorporated.

The evolutionary pacemaker of great apes was analyzed with genomic coding sequences accounting for populational phenomena that leads to gene tree/species tree discordance. We used an approach that ruled out the uncertainty in fossil calibrations on the evolutionary rate estimation, avoiding the controversies related to the age and phylogenetic position of fossil hominids.³¹ Moreover, relative rates were inferred with no a priori assumptions about the model of rate evolution or the branching process in the phylogenetic tree, which are known to impact molecular dating^{32–37} (see “Tree building and inference of evolutionary rates” section under the “Material and methods” section). The inferred pacemaker model for great apes corroborates the hominoid slowdown hypothesis^{24,38} and suggest that life history traits have a central role on the evolutionary rates across the genome. Finally, our findings reinforce that fully accounting for populational level phenomena is crucial to a more robust study of genome evolution.

Material and Methods

To infer a pacemaker model for great apes, we have downloaded from the OrthoMaM v10 database^{39,40} (accessed on February 4, 2019), a total of 11595 alignments of 1-to-1 orthologous

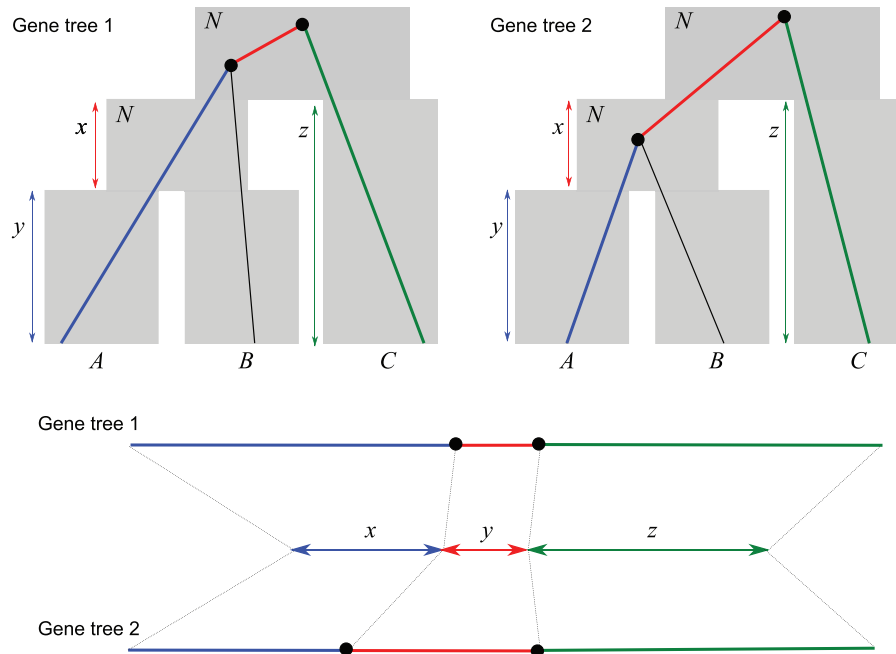


Figure 2. Even under the strict molecular clock, the relative proportions of branch lengths vary because of the variance in coalescence times under the multispecies coalescent process. The relative length of the internal branch, ancestral to species A and B, is greater in gene tree 2 than in gene tree 1, although the evolutionary rate was constant among all branches (N =effective population size).

coding genes containing sequences of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo abeli*), and gibbon (*Nomascus leucogenys*). The gibbon sequence was used as outgroup throughout the analyses. Alignments that contained non-unique sequences for any of the species were discarded, leading to a total of 11 491 gene alignments. OrthoMaM database entries for coding sequences used are available in the supplementary material.

Tree building and inference of evolutionary rates

Topological inference was carried out for each gene independently under the maximum likelihood framework as implemented in MEGA X⁴¹ under the substitution model chosen by the Akaike information criterion. Candidate substitution models were the default models tested by MEGA. Variation of evolutionary rates across alignment sites and invariant sites were also tested during the model selection procedure. The topology search criterion used was NNI (nearest neighbor interchange), with an initial tree constructed under NJ (Neighbor Joining).⁴² To estimate the pacemakers, we have analyzed only those gene trees that matched the species phylogeny, (((*Homo*, *Pan*), *Gorilla*), *Pongo*), comprising 6835 trees. This was done to avoid incorrect inference of pacemakers that may arise from SPILS.²⁸ We also used MEGA to implement a likelihood ratio test to evaluate the strict molecular clock hypothesis for each gene. The null hypothesis of the strict molecular was not rejected for 81% of these gene trees. This analysis was therefore used as a confirmatory result for evaluating the performance of pacemaker estimation. Since most genes failed to reject the strict

clock, we expect that at least one of the inferred pacemakers will exhibit similar relative rates in each branch.

Because definition of pacemakers ultimately relies on branch lengths, for the sake of comparison, we also performed maximum likelihood phylogenetic analysis in IQ-TREE.⁴³ This program implements sophisticated substitution models of sequence evolution that were chosen by the ModelFinder feature.⁴⁴ All the DNA substitution models implemented in IQ-TREE were used as candidates for the model selection procedure, including the treatment for base frequencies and rate heterogeneity across sites.⁴⁴ The topology search was performed with default IQ-TREE search, with an initial tree constructed under 100 parsimony trees + BIONJ tree.⁴³ In IQ-TREE, 6712 gene trees matched the species phylogeny.

The variation in ILS across chromosomes showed that the higher proportion of genes that match the species tree is located on the X chromosome, as well as this chromosome harbored the lower percentages of ILS (Supplementary material). This is in agreement with previous results,^{45–48} highlighting that genomic evolutionary patterns of ILS are consistent among Hominidae.

To estimate pacemakers, evolutionary rates for each branch must be calculated. Because pacemakers are defined by the relative ratio of rates among lineages (tree branches) across genes, the estimation of relative evolutionary rates is preferable than absolute rates. Using this approach, both systematic and sampling errors of absolute rate estimates, which are inherent of molecular dating with fossil calibrations, were avoided. Relative evolutionary rates for each branch were calculated using the RelTime method,^{49,50} as implemented in the MEGA

X software.⁴¹ This algorithm was shown successful for analyzing phylogenies with evolutionary rate heterogeneity.^{50,51} In RelTime, relative rates were computed from branch lengths for both IQ-TREE and MEGA inferred gene trees, using gibbon as the outgroup⁵² and considering that the evolutionary rate of the ancestral of great apes was one.

Pacemaker estimation

The number of pacemakers, as well as the assignment of genes to each pacemaker, was estimated by unsupervised statistical learning, using clustering algorithms available in the R programming environment. Difference of the relative rates between genes was calculated using the Euclidian distances between each gene's vector of relative branch rates. We employed K -means clustering to investigate the number of pacemakers that reduced the within-cluster sum of squared (WSS) Euclidian distances of rates between genes. The K -means algorithm was used to allocate genes into K clusters (pacemakers), varying from $K=1$ to $K=20$. Choice of the value of K that significantly reduced the total WSS was carried out with the gap statistic employing 1000 bootstrap replicates to estimate standard errors.⁵³ The gap statistic was calculated with the `clusGap` function of the `cluster` R package using the `scaledPCA` for deriving the null distribution, and the `firstSEmax` method for selecting the optimum number of clusters. This procedure was performed for the (1) gene trees estimated in MEGA that matched the species tree; (2) gene trees estimated in MEGA that matched the species tree and failed the strict molecular clock test; and (3) gene trees estimated in IQ-TREE that matched the species tree. To test whether very high relative rate values were impacting our results, we performed additional clustering analyses excluding genes that presented relative rates higher than 2, 5, 10, and 20 for at least one branch of the estimated phylogenetic tree.

To investigate whether our approach to estimate pacemakers would generate spurious results (false positives), we conducted a simulation in which gene trees were evolved under a species tree that mimicked the evolution of great apes, using parameters available in the literature. Speciation times were set according to the `timetree.org` database: *Homo/Pan* (6.7 Ma), *Gorilla* (9.1 Ma), *Pongo* (15.8 Ma), and *Nomascus* (20.2 Ma). Effective population sizes were kept constant along branches and assumed the values of 50 000, 75 000, and 100 000 Wright-Fisher individuals.⁵⁴ The generation time was set to 15 years and the mutation rate used was substitutions/site/generation.^{55–57} For each population size, 10 replicates containing 10 000 genes were simulated. Multispecies coalescent simulations were carried out in the `MCcoal` software.⁵⁸ In all simulations, our approach correctly estimated a single rate category, ie, the correct scenario of a strict molecular clock model was not rejected.

Besides measuring the rate of false positives (type I error), we measured the power of the test to correctly reject the UPM model when trees evolved under 2 and 5 pacemakers. To do so,

Table 1. Number of pacemakers estimated using the relative rate framework when data was simulated under models with 2 and 5 pacemakers (PM).

N_e	MEAN NUMBER OF PACEMAKERS (SD)	
	2-PM MODEL	5-PM MODEL
50 000	2.4 (0.8)	4.3 (1.6)
75 000	3.0 (1.3)	4.2 (1.7)
100 000	3.0 (0.9)	4.1 (2.6)

branch lengths of each gene tree used to calculate the type I error rate were multiplied by a vector of proportionality constants that were sampled from a log-normal distribution as in Duchêne and Ho.²⁰ In the 2-pacemakers model, two vectors were used, whereas five vectors were used in the 5-pacemakers model. Using the pacemaker inference approach described above, we were able to correctly reject the null hypothesis in all pacemaker scenarios (power=100%). The accuracy of our approach was impacted by the effective population size (N_e). In general, larger N_e s resulted in lower accuracy (Table 1).

Results

The gap statistic using K -means clustering assigned the genes to 5 and 3 clusters depending on the phylogenetic algorithm used to infer the trees, IQ-TREE or MEGA (Table 2). Therefore, great apes coding genes were allocated to 5 and 3 pacemakers according to IQ-TREE and MEGA results, respectively. Importantly, regardless of the algorithm used, a large cluster (pacemaker) containing more than 99% of the genes was recovered (Table 2). Besides this, we found that high relative rates, which could be considered as outliers, did not impact these results. When ignoring genes that produced high relative rate values, all genes fell within a single pacemaker (regardless we used 2, 5, 10, and 20 as cutoffs for relative rate values). Therefore, we opted for showing results including all genes, regardless the evolutionary rate value. Because of this, mean (Table 2) and median (Figure 3B and C) values of the distribution of relative rates for each branch may be very different. Therefore, most of the results and discussion will be conducted based on median values, which may be a better representation for skewed distributions.

In both IQ-TREE and MEGA analyses, the large cluster of genes was denominated as pacemaker 1 and presented very similar rate profiles (Figure 3B and C). In this pacemaker, the branch leading to orangutan presented median relative rates that were very similar to the ancestral of great apes. The distribution of relative rate values for the *Pongo* lineage presented a symmetrical shape and was centered around one (Figure 3). The internal branch ancestral of *Gorilla* and *Homo + Pan*, denominated as I2, presented a slight rate slowdown tendency when compared with the ancestor of great apes and the orangutan. This lineage exhibited the lowest dispersion of rate

values across genes. In the branch leading to gorillas, relative rates had a median value very close to one, although the dispersion around this value was greater than that calculated for the orangutan lineage. The ancestral of humans and chimpanzees, denominated as I1, showed a rate slowdown, evidenced by the smallest median value among all the branches analyzed. A minor rate slowdown was also recovered for *Homo* and *Pan* lineages, although it was more noticeable in humans than chimpanzees. The chimpanzee branch was the lineage with the greater amount of dispersion of relative rate values around the median (Table 2 and Figure 3), indicating higher levels of deviation from the ancestral rate of the great apes and, consequently, a largest deviation from the molecular clock.

Besides the large pacemaker, clustering analyses based on IQ-TREE phylogenetic reconstruction recovered four additional pacemakers (Table 2). All of them were composed by few genes each (from 9 to 20) and together they accounted for 0.82% of all genes analyzed. Rate patterns revealed by these clusters included a sharp acceleration of rates in chimpanzees (pacemaker 2); a rate increase in chimpanzees and slowdown in gorillas (pacemaker 3); a sharp acceleration of rates in the *gorilla* lineage plus a sharp slowdown in orangutans (pacemaker 4); and an independently speed up of the evolutionary rates in *Homo* and *Pan* lineages plus a slowdown on the *Pongo* lineage (pacemaker 5).

Clustering analysis based on MEGA produced two additional pacemakers. The recovered relative rate pattern for both pacemakers included speed ups of rates in the chimpanzee lineage, although the rate increase was more pronounced in pacemaker 3 than in pacemaker 2. The molecular clock test performed in MEGA failed to reject the strict clock hypothesis for 81% of the gene trees that matched the species tree. When considering only these 1294 non-clock genes, gap statistics analysis assigned genes to 3 clusters (Table 3). The vast majority of the non-clock genes were placed within one large group. On this cluster, *Pan* lineage exhibited much faster evolutionary rates than the remaining branches. The average rate values replicated this pattern, which were better visualized by the distribution of the logarithm of rates (Figure 4). In addition, most non-clock genes presented a slowdown in the human lineage, as well as on the internal branches, mainly on the ancestral of *Homo* and *Pan*. The other pacemakers recovered from non-clock genes retrieved an acceleration of evolutionary rates in the chimpanzee lineage. Such rate increase was high on pacemaker 2, but even more pronounced on pacemaker 3.

Discussion

Our results implied that most genomic coding regions of great apes were statistically assigned to a single pacemaker, using both IQ-TREE and MEGA phylogenetic algorithms. Among the gene trees that matched the species tree, more than 99% fell into this pacemaker, which corroborates the UPM model along hominid evolution. The remaining genes were placed into other pacemaker categories that together accounted for less

Table 2. Number of pacemakers, mean and median (inside parentheses) relative rates for each branch according to IQ-TREE and MEGA phylogenetic analyses.

APPROACH	PACEMAKERS	NUMBER OF GENES (%)	HOMO	PAN	I1	GORILLA	I2	PONGO
IQ-TREE	1	6657 (99.18)	2.7 (0.9)	7.2 (1.0)	1.3 (0.9)	4.0 (1.0)	0.9 (0.9)	1.2 (1.0)
	2	20 (0.30)	222.7 (0.5)	4382.6 (4535.6)	7.1 (0.5)	192.0 (0.1)	0.5 (0.5)	1.2 (1.6)
	3	13 (0.19)	538.9 (0.8)	8108.0 (7444.0)	0.8 (0.7)	0.8 (0.1)	0.8 (0.7)	0.7 (0.7)
	4	13 (0.19)	12.2 (8.0)	1038.8 (9.5)	0.7 (0.7)	6941.5 (6564.6)	0.7 (0.7)	0.04 (0.04)
	5	9 (0.13)	8585.7 (9028.9)	12200.2 (10021.1)	0.9 (0.9)	3.0 (2.0)	0.9 (0.9)	0.02 (0.01)
MEGA	1	6676 (99.14)	2.7 (1.0)	10.3 (1.0)	1.1 (0.9)	11.7 (1.0)	0.9 (0.9)	1.1 (1.0)
	2	56 (0.82)	468.8 (0.6)	6456.3 (5373.8)	3.6 (0.6)	112.5 (0.5)	0.6 (0.5)	1.0 (1.0)
	3	3 (0.04)	0.7 (0.7)	694418.4 (666604.4)	0.7 (0.7)	0.7 (0.7)	0.7 (0.7)	0.7 (0.7)

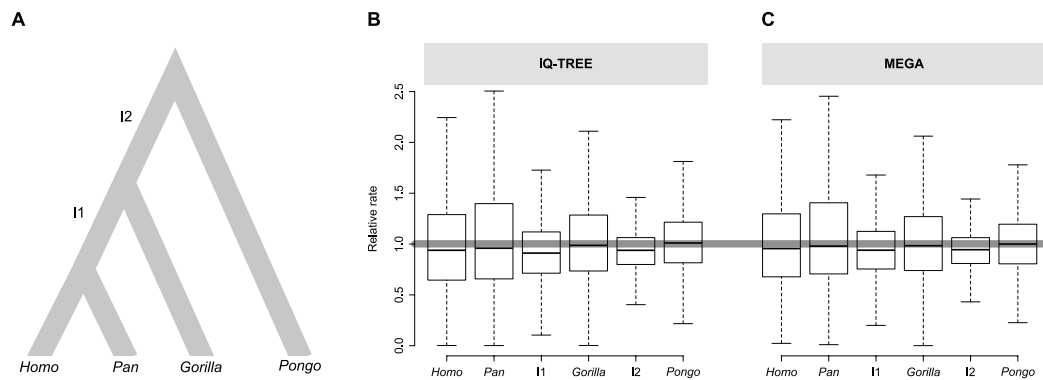


Figure 3. The distribution of relative rates across branches for genes in pacemaker 1. (A) Phylogeny of great apes with the denomination adopted for internal branches; (B) Relative rates recovered based on IQ-TREE phylogenetic analyses for genes in pacemaker 1; (C) Relative rates recovered based on MEGA phylogenetic analyses for genes in pacemaker 1. Outlier values were omitted. The horizontal black line inside the boxplots represents the median values.

Table 3. Number of pacemakers, mean and median (inside parentheses) relative rates for each branch inferred for the non-clock genes.

PACEMAKERS	NUMBER OF GENES (%)	HOMO	PAN	I1	GORILLA	I2	PONGO
1	1235 (95.44%)	9.9 (0.7)	48.4 (1.3)	1.6 (0.6)	58.5 (1.0)	0.7 (0.7)	1.6 (1.0)
2	56 (4.33%)	468.8 (0.6)	6456.3 (5373.8)	3.6 (0.6)	112.5 (0.5)	0.6 (0.5)	1.0 (1.0)
3	3 (0.23%)	0.7 (0.7)	694418.4 (666604.4)	0.7 (0.7)	0.7 (0.7)	0.7 (0.7)	0.7 (0.7)

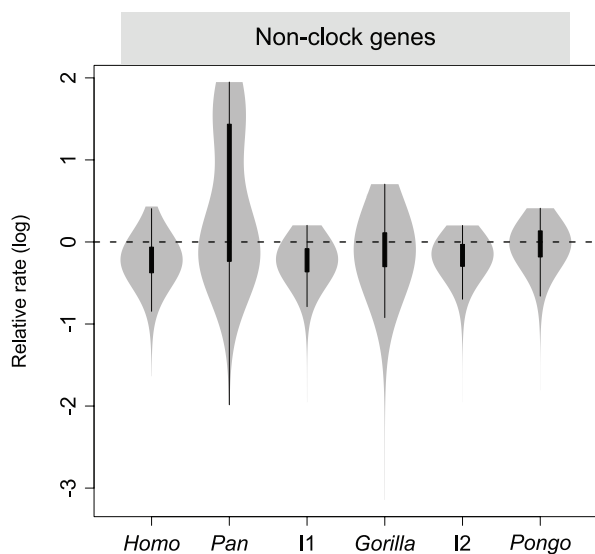


Figure 4. The distribution of relative rates for each lineage in the largest pacemaker recovered for non-clock genes. Outlier values were omitted.

than 1% of all coding sequences analyzed. We consider such percentage to be exceedingly low, indicating that only a few genomic regions exhibited modes of evolution better explained by gene effects. Therefore, we have not found evidences to support the MPM model and hypothesize that *lineage effects* played a major role during the evolutionary process of great apes. Our results also indicate that employing less sophisticated substitution models did not impact branch length estimation substantially, with distinct phylogenetic algorithms producing consistent results.

In addition, it was not possible to reject the strict molecular clock hypothesis for most coding regions from hominid genomes. In fact, the observation that great apes' evolutionary rates generally support a strict molecular clock was reported before.^{25,26} This massive indication of evolutionary rate constancy corroborates our results, since most genomic coding regions were statistically assigned to a pacemaker model of rate evolution that is closely related to the strict molecular clock model. However, slight deviations from a constant rates scenario were detected when analyzing the distribution of rates (Figure 3), such as the rate slowdown on the internal branches (I1 and I2) and on the *Homo* lineage. This corroborates the hypothesis of a decrease in the evolutionary rates of the human lineage, as previously sustained by several studies, and indicates that such rate slowdown also occurred in the ancestral branches leading to *Homo*.^{25,26,59–62} Our results were also in agreement with previous works that reported slower substitution rates in humans when compared with chimpanzees, which are, in turn, slower than the substitution rates in gorillas.^{25,59} Therefore, our inference of a pacemaker model supports the idea that life history traits drove rate variation in hominids, suggesting that these traits were major drivers of evolutionary rate evolution across the genome.⁶³ Specifically, the increase in generation times has been suggested as the main cause for such slowdown in hominoids.^{27,64,65} Importantly, other studies already have shown that such decrease is not recovered when CpG transitions are analyzed, since such mutations correlates with calendar years instead of replication cycles.^{25,60}

Few genes fell within pacemakers other than pacemaker 1. In all cases, both IQ-TREE and MEGA phylogenetic algorithms resulted in alternative pacemakers that presented extremely large rate accelerations (Tables 2 and 3). As already mentioned, this could be mainly due to gene effects. But, we should not rule out other explanations for that. Misidentification of orthologous regions is one possible reason for the presence of unrealistic branch lengths and, consequently, extremely high relative rate estimates. However, OrthoMaM is a high-quality 1-to-1 database of coding regions, decreasing the probability of such error. Another explanation would be the use of incorrect substitution models, which would also bias branch length estimation. In fact, this was one of the main reasons we adopted two different phylogenetic algorithms to perform model selection and phylogeny reconstruction. Therefore, there is a considerable probability that the evolution of genes excluded from pacemaker 1 have been governed by gene effects, which makes such loci ideal to test the effects of selective constraints.

The estimation of a pacemaker model was previously carried out for mammals by Duchêne and Ho,²⁰ who suggested that mammalian genes could be assigned to 14 distinct pacemakers. This was in contrast to previous studies that found a UPM model for bacteria, archaea, *Drosophila*, and yeast.^{18,19,23} Duchêne and Ho's²⁰ analysis, however, did not account for gene tree/species tree discordance, which probably impacted their results due to the presence of SPILS.²⁸ Importantly, the number of species analyzed was large²⁹ and, as the number of terminals increase, the number of gene tree topologies also increases exponentially. This may further aggravate the bias caused by spurious substitutions. In our pacemaker estimation analysis, gene trees that did not match the species tree were discarded; mainly because enforcing their topology to the species tree would have misled our result.

Moreover, the approach employed in Duchêne and Ho²⁰ did not estimate evolutionary rates for each branch to infer the pacemaker models. Instead, they implemented a procedure that scaled the total length of each gene tree to 1, making branch lengths fractions of this total length. Evolutionary rates and times were thus not entirely decoupled, making the comparison of branch lengths misleading. This is particularly true for internal branches, where time elapsed for each branch is completely unknown. In contrast, we have specifically estimated relative evolutionary rates for all branches of the gene tree using the rate at the hominid's ancestor as the unit. This strategy was advantageous because we did not assume any model of rate evolution among branches, or the branching process between lineages. Both factors are known to impact molecular dating and, consequently, evolutionary rate estimation.^{32,36,66} Hence, we consider the evolutionary rate estimation used in this study to be less sensitive to uncertainties imposed by model assumptions from most molecular dating analyses.

In summary, our estimation of a pacemaker model for great apes corroborated the hominoid slowdown hypothesis by employing entirely different approaches from those performed up to date. The results suggested that the vast majority of genes follow a pace of evolution that is compatible with the strict molecular clock, although rate slowdowns were recovered. The estimation of relative rates relied on a calibration-free approach that took gene tree discordance into account, which had never been carried out for pacemaker estimation. We have investigated only orthologous coding sequences that resulted in gene trees matching the species phylogeny. However, given the large number of loci analyzed (>6000), the inferred pacemaker model may consist of a reasonable representation of great apes' genomes. We stress that ignoring populational level phenomena on ancestral lineages by enforcing the species tree topology to gene trees may impact evolutionary rate estimation and, consequently, pacemaker inference. Finally, the UPM model is supported by our analyses, indicating that lineage effects had a fundamental role in shaping the evolution of substitution rates across the genome.

Author Contributions

BM and CGS conceived the idea, designed research, performed analyses, discussed results and wrote the manuscript.

Supplemental Material

Supplemental material for this article is available online.

ORCID iD

Beatriz Mello  <https://orcid.org/0000-0002-4511-2757>

REFERENCES

- Zuckerandl E, Pauling L. Molecular disease, evolution, and genic heterogeneity. In: Kasha M, Pullman B, eds. *Horizons in Biochemistry*. New York, NY: Academic Press; 1962:189–225.
- Zuckerandl E, Pauling L. Molecules as documents of evolutionary history. *J Theor Biol*. 1965;8:357–366.
- Thorne JL, Kishino H, Painter IS. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol*. 1998;15:1647–1657.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4:e88.
- Lepage T, Bryant D, Philippe H, Lartillot N. A general comparison of relaxed molecular clock models. *Mol Biol Evol*. 2007;24:2669–2680.
- Heath TA, Holder MT, Huelsenbeck JP. A Dirichlet process prior for estimating lineage-specific substitution rates. *Mol Biol Evol*. 2012;29:939–955.
- Kumar S. Molecular clocks: four decades of evolution. *Nat Rev Genet*. 2005;6:654–662.
- Ayala FJ. Vagaries of the molecular clock. *Proc Natl Acad Sci U S A*. 1997;94:7776–7783.
- Gillespie JH. The molecular clock may be an episodic clock. *Proc Natl Acad Sci*. 1984;81:8009–8013.
- Gu X, Li W-H. Higher rates of amino acid substitution in rodents than in humans. *Mol Phylogenet Evol*. 1992;1:211–214.
- Bousquet J, Strauss SH, Doerksen AH, Price RA. Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. *Proc Natl Acad Sci U S A*. 1992;89:7844–7848.
- Thorne JL, Kishino H. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol*. 2002;51:689–702.
- Gillespie JH. Lineage effects and the index of dispersion of molecular evolution. *Mol Biol Evol*. 1989;6:636–647.

14. Ho SYW. The changing face of the molecular evolutionary clock. *Trends Ecol Evol.* 2014;29:496–503.
15. Grishin NV, Wolf YI, Koonin EV. From complete genomes to measures of substitution rate variability within and between proteins. *Genome Res.* 2000;10:991–1000.
16. Wolf YI, Novichkov PS, Karev GP, Koonin EV, Lipman DJ. The universal distribution of evolutionary rates of genes and distinct characteristics of eukaryotic genes of different apparent ages. *Proc Natl Acad Sci U S A.* 2009;106:7273–7280.
17. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell.* 2008;134:341–352.
18. Snir S, Wolf YI, Koonin EV. Universal pacemaker of genome evolution. *PLoS Comput Biol.* 2012;8:e1002785.
19. Snir S, Wolf YI, Koonin EV. Universal pacemaker of genome evolution in animals and fungi and variation of evolutionary rates in diverse organisms. *Genome Biol Evol.* 2014;6:1268–1278.
20. Duchêne S, Ho SY. Mammalian genome evolution is governed by multiple pacesetters. *Bioinformatics.* 2015;31:2061–2065.
21. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A.* 1987;84:9054–9058.
22. Gaut B, Yang L, Takuno S, et al. The patterns and causes of variation in plant nucleotide substitution rates. *Annu Rev Ecol Syst.* 2011;42:245–266.
23. Wolf YI, Snir S, Koonin EV. Stability along with extreme variability in core genome evolution. *Genome Biol Evol.* 2013;5:1393–1402.
24. Goodman M. Serological analysis of the systematics of recent hominoids. *Hum Biol.* 1963;35:377–436.
25. Moorjani P, Amorim CEG, Arndt PF, Przeworski M. Variation in the molecular clock of primates. *Proc Natl Acad Sci U S A.* 2016;113:10607–10612.
26. Besenbacher S, Hvilson C, Marques-Bonet T, et al. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol.* 2019;3:286–292.
27. Scally A. Mutation rates and the evolution of germline structure. *Philos Trans R Soc Lond B Biol Sci.* 2016;371:20150137.
28. Mendes FK, Hahn MW. Gene tree discordance causes apparent substitution rate variation. *Syst Biol.* 2016;65:711–721.
29. Martin CH, Hohna S, Crawford JE, Turner BJ, Richards EJ, Simons LH. The complex effects of demographic history on the estimation of substitution rate: concatenated gene analysis results in no more than twofold overestimation. *Proc Biol Sci.* 2017;284:20170537.
30. Pamilo P, Nei M. Relationships between gene trees and species trees. *Mol Biol Evol.* 1988;5:568–583.
31. Fleagle JG. *Primate Adaptation and Evolution.* Cambridge, MA: Academic Press.
32. Condamine FL, Nagalingum NS, Marshall CR, Morlon H. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evol Biol.* 2015;15:65. doi:10.1186/s12862-015-0347-8.
33. Mello B, Tao Q, Kumar S. Molecular dating for phylogenies containing a mix of populations and species [published online ahead of print January 31, 2019]. *bioRxiv.* doi:10.1101/536656.
34. Battistuzzi FU, Filipinski A, Hedges SB, Kumar S. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol Biol Evol.* 2010;27:1289–1300.
35. Christin P-A, Spriggs E, Osborne CP, Stromberg CA, Salamin N, Edwards EJ. Molecular dating, evolutionary rates, and the age of the grasses. *Syst Biol.* 2014;63:153–165.
36. Foster CSP, Sauquet H, van der Merwe M, et al. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Syst Biol.* 2016;66:338–351.
37. Pacheco MA, Matta NE, Valkiunas G, et al. Mode and rate of evolution of haemosporean mitochondrial genomes: timing the radiation of avian parasites. *Mol Biol Evol.* 2018;35:383–403.
38. Yi SV. Morris Goodman's hominoid rate slowdown: the importance of being neutral. *Mol Phylogenet Evol.* 2013;66:569–574.
39. Douzery EJP, Scornavacca C, Romiguier J, et al. OrthoMaM v8: a database of orthologous exons and coding sequences for comparative genomics in mammals. *Mol Biol Evol.* 2014;31:1923–1928.
40. Scornavacca C, Belkhir K, Lopez J, et al. OrthoMaM v10: scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. *Mol Biol Evol.* 2019;36:861–862. doi:10.1093/molbev/msz015.
41. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–1549. doi:10.1093/molbev/msy096.
42. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4:406–425.
43. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32:268–274.
44. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–589.
45. Hobolth A, Christensen OF, Mailund T, Schierup MH. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet.* 2007;3:e7.
46. Scally A, Duthel JY, Hillier LW, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature.* 2012;483:169–175.
47. Pease JB, Hahn MW. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution.* 2013;67:2376–2384.
48. Duthel JY, Munch K, Nam K, Mailund T, Schierup MH. Strong selective sweeps on the X chromosome in the human–chimpanzee ancestor explain its low divergence. *PLoS Genet.* 2015;11:e1005451.
49. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipinski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A.* 2012;109:19333–19338.
50. Tamura K, Tao Q, Kumar S. Theoretical foundation of the RelTime method for estimating divergence times from variable evolutionary rates. *Mol Biol Evol.* 2018;35:1770–1782.
51. Mello B, Tao Q, Tamura K, Kumar S. Fast and accurate estimates of divergence times from big data. *Mol Biol Evol.* 2017;34:45–50.
52. Mello B. Estimating timetrees with MEGA and the TimeTree resource. *Mol Biol Evol.* 2018;35:2334–232. doi:10.1093/molbev/msy133.
53. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc Ser B Stat Methodol.* 2001;63:411–423.
54. Schrago CG. The effective population sizes of the anthropoid ancestors of the human–chimpanzee lineage provide insights on the historical biogeography of the great apes. *Mol Biol Evol.* 2014;31:37–47.
55. Yang Z. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics.* 2002;162:1811–1823.
56. Wong WSW, Solomon BD, Bodian DL, et al. New observations on maternal age effect on germline de novo mutations. *Nat Commun.* 2016;7:10486. doi:10.1038/ncomms10486.
57. Capellari RT, Costa-Paiva EM, Schrago CG. Appropriate assignment of fossil calibration information minimizes the difference between phylogenetic and pedigree mutation rates in humans. *Life (Basel).* 2018;8:49.
58. Rannala B, Yang Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics.* 2003;164:1645–1656.
59. Elango N, Thomas JW, Yi SV. Variable molecular clocks in hominoids. *Proc Natl Acad Sci U S A.* 2006;103:1370–1375.
60. Kim S-H, Elango N, Warden C, Vigoda E, Yi SV. Heterogeneous genomic molecular clocks in primates. *PLoS Genet.* 2006;2:e163.
61. Steiper ME, Young NM, Sukarna TY. Genomic data support the hominoid slowdown and an early oligocene estimate for the hominoid–cercopithecoid divergence. *Proc Natl Acad Sci U S A.* 2004;101:17021–17026.
62. Yi S, Ellsworth DL, Li W-H. Slow molecular clocks in old world monkeys, apes, and humans. *Mol Biol Evol.* 2002;19:2191–2198.
63. Sayres MAW, Venditti C, Pagel M, et al. Do variations in substitution rates and male mutation bias correlate with life-history traits? a study of 32 mammalian genomes: life history traits and genome evolution. *Evolution.* 2011;65:2800–2815.
64. Jones JH. Primates and the evolution of long, slow life histories. *Curr Biol.* 2011;21:R708–717.
65. Amster G, Sella G. Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc Natl Acad Sci U S A.* 2016;113:1588–1593.
66. dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PC, Yang Z. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol.* 2015;25:2939–2950.