

Evolution of Old World monkeys and great apes links to massive and directional shrinkage of the dinucleotide short tandem repeat compartment

M Arabfard

Baqiyatallah University of Medical Sciences

M Salesi

Baqiyatallah University of Medical Sciences

I Arabipour

Islamic Azad University

M Ohadi (✉ ohadi.mina@yahoo.com)

University of Social Welfare and Rehabilitation Sciences

Research Article

Keywords: dinucleotide, short tandem repeat, directional, Old World monkey, great ape, rodent, non-random

Posted Date: April 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-441796/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The evolutionary trend of short tandem repeats (STRs) at the crossroads of speciation remains largely elusive and attributed to random evolution for the most part. To explore this trend, we selected nine species, which shared sequential chronological ancestors, including rat, mouse, olive baboon, gelada, macaque, gorilla, chimpanzee, bonobo, and human, and collected three sets of data on the abundance of all classes of dinucleotide STRs (≥ 6 -repeats) for three regions of every chromosome, each region spanning 10 Mb of DNA.

Results: In all three datasets, we found directional shrinkage of the dinucleotide STR compartment as follows: rodents>Old World monkeys>great apes ($P=0.000$). The decremented gradient observed for the dinucleotide STRs was not detected for a number of other classes of STRs, such as mono and trinucleotide STRs.

Conclusion: We report the first instance of massive and directional gradient of STRs, which may link with the evolution of Old World monkeys and great apes.

Introduction

With an exceptionally high degree of polymorphism and plasticity, short tandem repeats (STRs) are a spectacular source of variation required for speciation and evolution¹⁻³. The impact of STRs on speciation is supported by their various functional implications, such as gene expression, alternative splicing, and translation³⁻¹⁰.

While a limited number of studies indicate that purifying selection and drift can shape the structure of STRs at the inter- and intra-species levels¹¹⁻¹⁶, the global trend of STR evolution at the crossroads of primate speciation remains largely unknown.

The most common STRs in the human genome are dinucleotide repeats¹⁷. Here we analyzed the evolutionary trend of this category of STRs in nine selected species encompassing rodent, Old World monkey, and great apes.

Materials And Methods

Extraction of STRs from genomic sequences.

The abundance of dinucleotide STRs of ≥ 6 -repeats was studied in nine selected species, which shared sequential chronological ancestors, including mouse, rat, gelada, olive baboon, macaque, gorilla, chimpanzee, bonobo, and human. Those species were classified into three classes as follows: Class 1 (rodent): rat and mouse, Class 2 (Old World monkeys): gelada, olive baboon, macaque, and Class 3 (great apes): gorilla, bonobo, chimpanzee, human.

We designed a software package in C# environment (Suppl. 1) (<https://github.com/arabfard/STR-Finder>). All possibilities of dinucleotide motifs, consisting of AC, AG, AT, CA, CG, CT, GA, GC, GT, TA, TC, and TG were included. The written program was based on perfect (pure) STRs. By using the REST API service¹⁸ from Ensembl 101 (<https://asia.ensembl.org>)^{19,20}, data of three arbitrary regions of every chromosome, each region spanning 10 Megabase (Mb) of genomic DNA, were accessed in the nine species (Fig. 1). In each chromosome, the STR abundances were calculated and compared on a chromosome-to-chromosome basis (Suppl. 2). Subsequent to collecting the entire data, we also differentiated the STRs based on their length into two classes of 6–20 repeats and > 20-repeats, and studied their abundance in the selected species. Finally, the data of the selected regions of the chromosomes in the nine species were aggregated and analyzed.

Additionally, to compare the trend of dinucleotide STRs with other classes of STRs, we used the STR-Finder tool to screen the selected regions and species for mononucleotide STRs (T, G, A, and C) of ≥ 6 -repeats and trinucleotide STRs (GCC/ GGC) of ≥ 6 -repeat lengths.

Statistical analysis

The dinucleotide STR abundance trend in the nine selected species was compared across datasets 1, 2, and 3, by correlation coefficient and repeated measurements analysis (Table 1).

Table 1
Overall dinucleotide STR count across datasets 1, 2, and 3.

	Bonobo	Human	Chimpanzee	Gorilla	Gelada	Baboon	Macaque	Mouse	Rat
6–20 Repeats	14047	19422	16459	17620	18407	19802	22965	27951	39381
> 20 Repeats	715	1067	614	845	383	434	819	6948	12806
6–20 Repeats	14583	15743	18215	18061	19371	19661	20230	39189	41443
> 20 Repeats	836	1143	781	833	443	572	967	10222	13722
6–20 Repeats	17634	19157	19046	18326	21585	19925	19902	41658	41480
> 20 Repeats	905	1324	773	838	511	586	968	11087	13983
	Bonobo	Human	Chimpanzee	Gorilla	Gelada	Baboon	Macaque	Mouse	Rat
DS1	14762	20489	17073	18465	18790	20236	23784	34899	52187
DS2	15419	16886	18996	18894	19814	20233	21197	49411	55165
DS3	18539	20481	19819	19164	22096	20511	20870	52745	55463
Sum	48720	57856	55888	56523	60700	60980	65851	137055	162815
P-values across datasets 1, 2, and 3: 0.146									
Across Species Classes 1, 2, and 3: 0.000									
Within Class 1: 0.174									
Within Class 2: 0.288									
Within Class 3: 0.198									
Class 1 = Rodents. Class 2 = Old World monkeys, Class 3 = Great apes.									
DS: Dataset									

Comparisons of within and between classes were analyzed using one and two-way Anova tests. These analyses were confirmed by nonparametric tests

Results

Overall directional shrinkage of the dinucleotide STR compartment in Old World monkeys and great apes vs. rodents.

In three independent analyses, we studied the distribution of dinucleotide STRs in respect of their abundance across rodents and primates. The observed trend was strikingly decremental as follows: rodents > Old World monkeys > great apes $P = 0.000$ (Table 1, Fig. 2). That trend was replicated in datasets 1, 2, and 3 ($P = 0.80$).

Differential gradient of dinucleotide STRs based on their length.

The directional gradient, rodents > Old World monkeys > great apes, was found to be predominantly laid in the dinucleotide STRs of 6–20 repeats (Fig. 3). While the > 20-repeat compartment was the most dramatically affected in respect of shrinkage in primates vs. rodents, this compartment was the largest in human, in comparison with the remaining six primate species studied.

Differential gradient of STR classes in rodents vs. Old World monkeys vs. great apes.

To examine whether the observed trend in dinucleotide STRs can be generalized to other STR classes, we analyzed the abundance trend of mononucleotide STRs (G, A, T, and C) ≥ 6 -repeats (Fig. 4) and trinucleotide STRs (GGC/GCC) ≥ 6 -repeats (Fig. 5) in the selected species. While the trend in dinucleotide STRs was a decremented gradient (rodents > Old World monkeys > great apes), a similar trend was not detected in the mono (Fig. 4) and tri STR (Fig. 5) compartment. In fact, the dramatic excess of the dinucleotide compartment observed in rodents was not observed for the mono and trinucleotide STRs.

Discussion

It is largely unknown whether at the crossroads of speciation, STRs evolved as a result of purifying selection, genetic drift, and/or in a directional manner. In an attempt to resolve part of the picture, we selected multiple species that shared sequential chronological ancestors, and investigated all possible dinucleotide STRs of all possible lengths (≥ 6 -repeats). Our analysis revealed an overall directional gradient in the abundance of dinucleotide STRs during primate speciation, evidenced by the following trend: rodents > Old World monkeys > great apes (Fig. 6).

Dinucleotide STRs are the most abundant class of STRs in the vertebrate genomes, and their global pattern of abundance may shed light on a vastly unknown aspect of evolutionary biology. The replicated trends observed in our three datasets seem to be independent of the genome size of the selected species. Mouse and rat have the highest abundance of dinucleotide STRs in comparison to the seven selected primate species, and yet their genomes are smaller than those species. This finding is in line with the previous reports of lack of relationship between genome size and abundance of STRs^{21,22}.

An alternative hypothesis to the directional shrinkage of the dinucleotide STR compartment in primates vs. rodents is that this compartment has expanded excessively in rodents. Indeed, rodent genomes appear to be significantly rich in STRs in comparison to several other mammals¹¹. However, our present findings indicate that the above property cannot be generalized to all classes of STRs, as, at least, mono and trinucleotide STRs did not show the dramatic excess observed in the dinucleotide compartment in rodents. Furthermore, the decremented gradient of the

dinucleotide STR compartment in the following order: rodents > Old World monkeys > great apes supports the shrinkage hypothesis for the dinucleotide compartment.

It is possible that there is a mathematical threshold required for the abundance of STRs in various orders of species (Fig. 6). This is in line with the hypothesis that STRs function as scaffolds for biological computers²³.

Certain STRs located in the protein-coding gene core promoters have been subject to contraction in the process of human and non-human primate evolution²⁴. A number of those STRs are identical in formula in primates vs. non-primates, and the genes linked to those STRs are involved in characteristics that have diverged primates from other mammals, such as craniofacial development, neurogenesis, and spine morphogenesis. It is likely that those STRs functioned as evolutionary switch codes for primate speciation. In line with the above, structural variants are enriched near genes that diverged in expression across great apes^{25,26}. It is speculated that STR variants are more likely than single-nucleotide variants to have epistatic interactions, which can have significant consequences in complex traits, in human as well as model organisms^{27,28}. Future studies such as large-scale genome-editing of STRs²⁹ in embryonic stem cells and investigation of their differentiation into various cell lineages may be candidate approaches to investigate how the massive and dramatically diverged trend of dinucleotide STRs links to primate speciation and evolution.

Conclusion

In conclusion, we propose that massive and directional shrinkage of the dinucleotide STR compartment links to, and probably had a determining impact on primate speciation. This is a prime instance of non-random STR gradient in multiple speciation.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Bioinformatics tools are available at: https://figshare.com/articles/dataset/Di-Finder_Data/13603274.

Competing interests

Authors have no conflict of interest to declare.

Funding

This research was funded by the University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

Authors' contributions

MA collected data and performed the bioinformatic analyses. MS performed the biostatistics analyses. IA contributed to data collection. MO conceived, designed, and supervised the project, and wrote the manuscript.

Acknowledgements

Not applicable

References

- 1 Mohammadparast, S., Bayat, H., Biglarian, A. & Ohadi, M. Exceptional expansion and conservation of a CT-repeat complex in the core promoter of PAXBP1 in primates. *Am. J. Primatol.* **76**, 747-756 (2014).
- 2 Bushehri, A., Barez, M. R. M., Mansouri, S. K., Biglarian, A. & Ohadi, M. Genome-wide identification of human-and primate-specific core promoter short tandem repeats. *Gene* **587**, 83-90 (2016).
- 3 Nikkhah, M., Rezazadeh, M., Khorshid, H. R. K., Biglarian, A. & Ohadi, M. An exceptionally long CA-repeat in the core promoter of SCGB2B2 links with the evolution of apes and Old World monkeys. *Gene* **576**, 109-114 (2016).
- 4 Jakubosky, D. *et al.* Properties of structural variants and short tandem repeats associated with gene expression and complex traits. *Nature Communications* **11**, 1-15 (2020).
- 5 Valipour, E. *et al.* Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes. *Gene* **531**, 175-179 (2013).
- 6 Ranathunge, C. *et al.* Transcribed microsatellite allele lengths are often correlated with gene expression in natural sunflower populations. *Mol. Ecol.* (2020).
- 7 Press, M. O., Hall, A. N., Morton, E. A. & Queitsch, C. Substitutions are boring: Some arguments about parallel mutations and high mutation rates. *Trends Genet.* **35**, 253-264 (2019).
- 8 Fotsing, S. F. *et al.* The impact of short tandem repeat variation on gene expression. *Nat. Genet.* **51**, 1652-1659 (2019).
- 9 Arabfard, M., Kavousi, K., Delbari, A. & Ohadi, M. Link between short tandem repeats and translation initiation site selection. *Human genomics* **12**, 47 (2018).
- 10 Yap, K. *et al.* A short tandem repeat-enriched RNA assembles a nuclear compartment to control alternative splicing and promote cell survival. *Mol. Cell* **72**, 525-540 (2018).
- 11 Srivastava, S., Avvaru, A. K., Sowpati, D. T. & Mishra, R. K. Patterns of microsatellite distribution across eukaryotic genomes. *BMC Genomics* **20**, 153 (2019).
- 12 Pavlova, A. *et al.* Purifying selection and genetic drift shaped Pleistocene evolution of the mitochondrial genome in an endangered Australian freshwater fish. *Heredity (Edinb.)* **118**, 466-476 (2017).
- 13 Jorde, P. E. *et al.* Genetically distinct populations of northern shrimp, *Pandalus borealis*, in the North Atlantic: adaptation to different temperatures as an isolation factor. *Mol. Ecol.* **24**, 1742-1757 (2015).

- 14 Legrand, D., Chenel, T., Campagne, C., Lachaise, D. & Cariou, M. L. Inter-island divergence within *Drosophila mauritiana*, a species of the *D. simulans* complex: Past history and/or speciation in progress? *Mol. Ecol.* **20**, 2787-2804 (2011).
- 15 Sun, G. *et al.* Global genetic variation at nine short tandem repeat loci and implications on forensic genetics. *Eur. J. Hum. Genet.* **11**, 39-49 (2003).
- 16 Abe, H. & Gemmell, N. J. Evolutionary footprints of short tandem repeats in avian promoters. *Sci. Rep.* **6**, 1-11 (2016).
- 17 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. (2001).
- 18 Yates, A. *et al.* The Ensembl REST API: Ensembl data for any language. *Bioinformatics* **31**, 143-145 (2015).
- 19 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754-D761 (2018).
- 20 Howe, K. L. *et al.* Ensembl genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689-D695 (2020).
- 21 Neff, B. D. & Gross, M. R. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution* **55**, 1717-1733 (2001).
- 22 Park, J. Y. *et al.* Evolutionary constraints over microsatellite abundance in larger mammals as a potential mechanism against carcinogenic burden. *Sci. Rep.* **6**, 1-5 (2016).
- 23 Herbert, A. Simple Repeats as Building Blocks for Genetic Computers. *Trends Genet.* **36**(10):739-750. (2020).
- 24 Ohadi, M. *et al.* Core promoter short tandem repeats as evolutionary switch codes for primate speciation. *Am. J. Primatol.* **77**, 34-43 (2015).
- 25 Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360** (2018).
- 26 Bilgin Sonay, T. *et al.* Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. *Genome Res.* **25**, 1591–1599 (2015).
- 27 Bagshaw AT, Horwood LJ, Fergusson DM, Gemmell NJ, Kennedy MA. Microsatellite polymorphisms associated with human behavioural and psychological phenotypes including a gene-environment interaction. *BMC Med Genet.* 2017 Feb 3;18(1):12. doi: 10.1186/s12881-017-0374-y. PMID: 28158988; PMCID: PMC5291968.
- 28 Press, M. O., Carlson, K. D. & Queitsch, C. The overdue promise of short tandem repeat variation for heritability. *Trends Genet.* **30**, 504-512 (2014).
- 29 Smith, C. J. *et al.* Enabling large-scale genome editing at repetitive elements by reducing DNA nicking. *Nucleic Acids Res.* **48**, 5183-5195 (2020).

Table

Table 1. Overall dinucleotide STR count across datasets 1, 2, and 3.

	Bonobo	Human	Chimpanzee	Gorilla	Gelada	Baboon	Macaque	Mouse	Rat
6-20 Repeats	14047	19422	16459	17620	18407	19802	22965	27951	39381
>20 Repeats	715	1067	614	845	383	434	819	6948	12806
6-20 Repeats	14583	15743	18215	18061	19371	19661	20230	39189	41443
>20 Repeats	836	1143	781	833	443	572	967	10222	13722
6-20 Repeats	17634	19157	19046	18326	21585	19925	19902	41658	41480
>20 Repeats	905	1324	773	838	511	586	968	11087	13983
	Bonobo	Human	Chimpanzee	Gorilla	Gelada	Baboon	Macaque	Mouse	Rat
DS1	14762	20489	17073	18465	18790	20236	23784	34899	52187
DS2	15419	16886	18996	18894	19814	20233	21197	49411	55165
DS3	18539	20481	19819	19164	22096	20511	20870	52745	55463
Sum	48720	57856	55888	56523	60700	60980	65851	137055	162815

P-values across datasets 1, 2, and 3: 0.146

Across Species Classes 1, 2, and 3: 0.000

Within Class 1: 0.174

Within Class 2: 0.288

Within Class 3: 0.198

Class 1=Rodents. Class 2=Old World monkeys, Class 3=Great apes.

DS: Dataset

Figures



Figure 1

Schematic representation of data collection of the mono, di, and trinucleotide STR compartments. All chromosomes were screened across the nine selected species in three datasets 1, 2, and 3. Only one chromosome is depicted as an example.

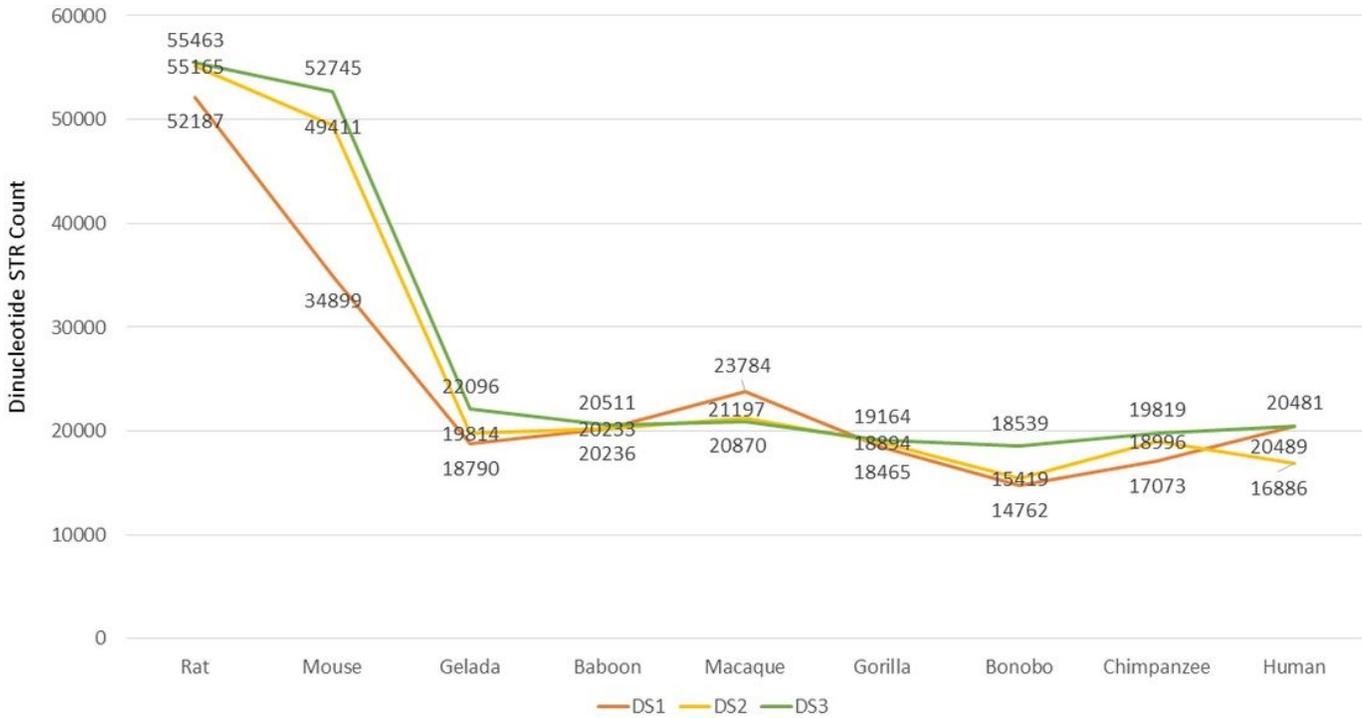


Figure 2

Massive directional shrinkage of the dinucleotide STR compartment in primates, replicated in datasets 1, 2, and 3.

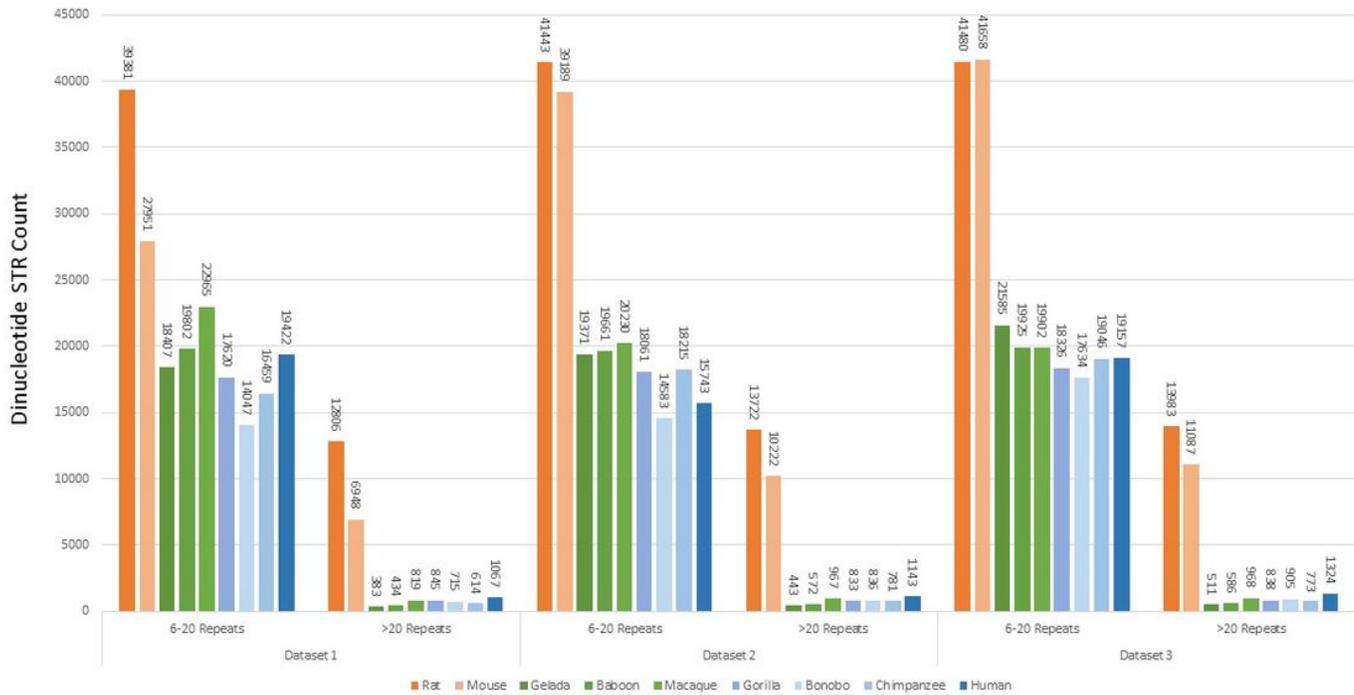


Figure 3

Abundance of various dinucleotide STR lengths across the nine selected species. The directional decremented gradient was predominantly laid in the 6-20 length compartment. While the >20-repeat compartment was the most dramatically affected as a result of shrinkage in primates in comparison with rodents, human had the highest abundance of the >20 repeat STRs in comparison with all other primates studied.

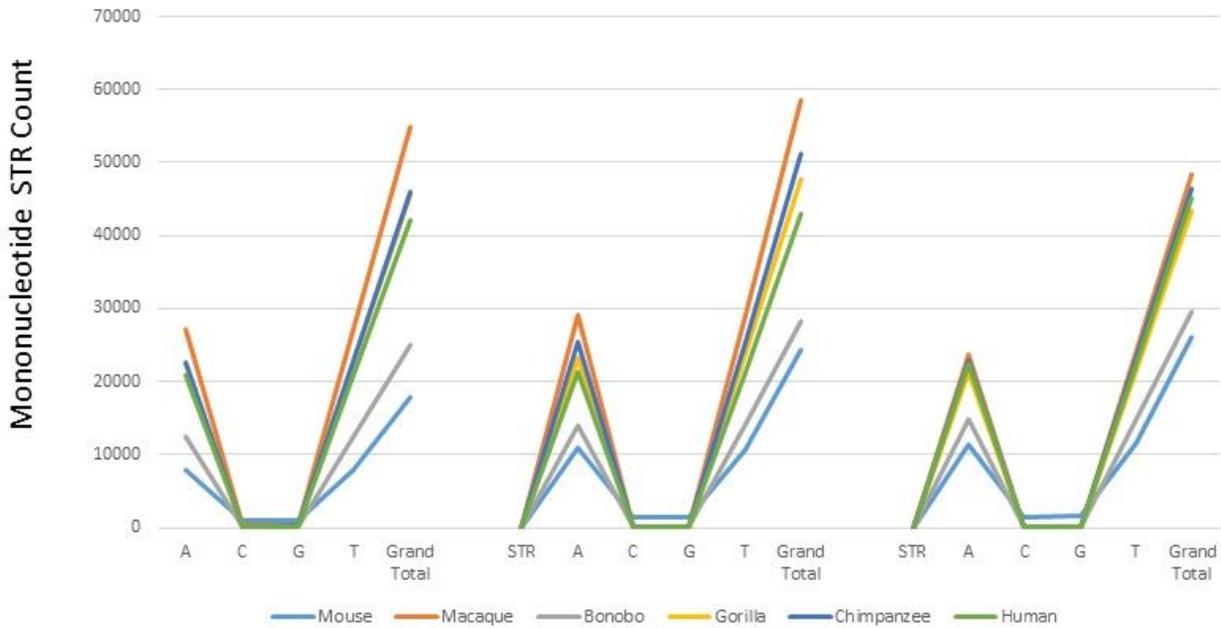


Figure 4

Mononucleotide STR trend across rodent and primates. In contrast to the dinucleotide compartment, the trend in the trinucleotide STRs was not decremented in primates vs. mouse.

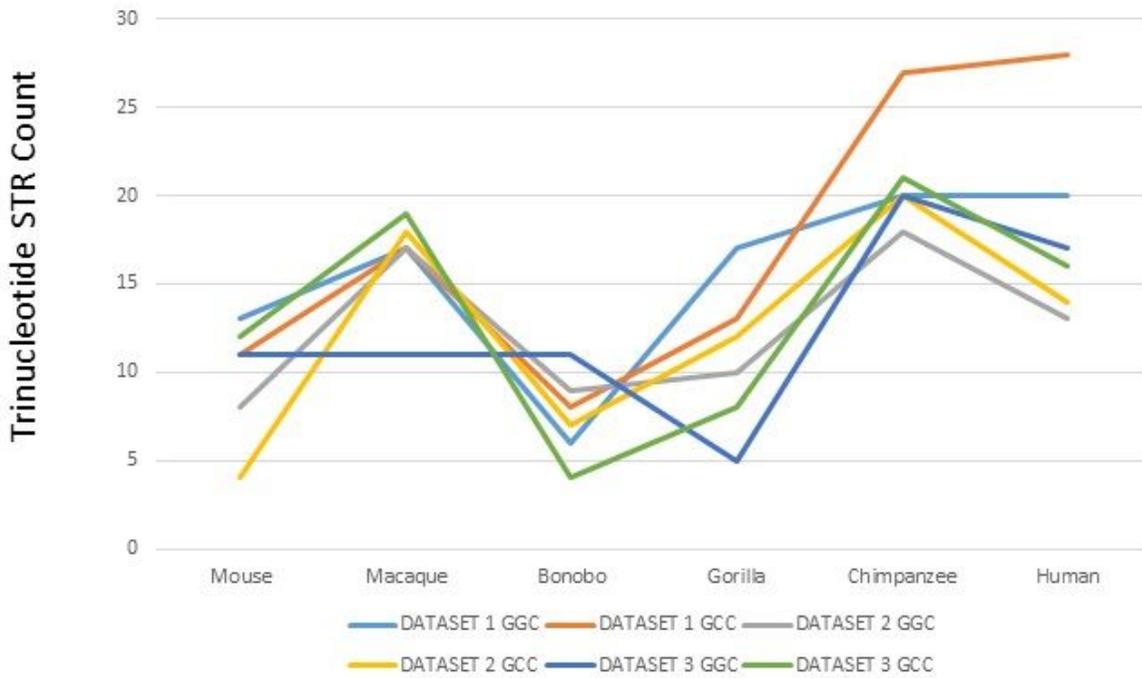


Figure 5

Trinucleotide STR trend across rodent and primates. In contrast to the dinucleotide compartment, the trend in the trinucleotide STRs was not decremented in primates vs. mouse.

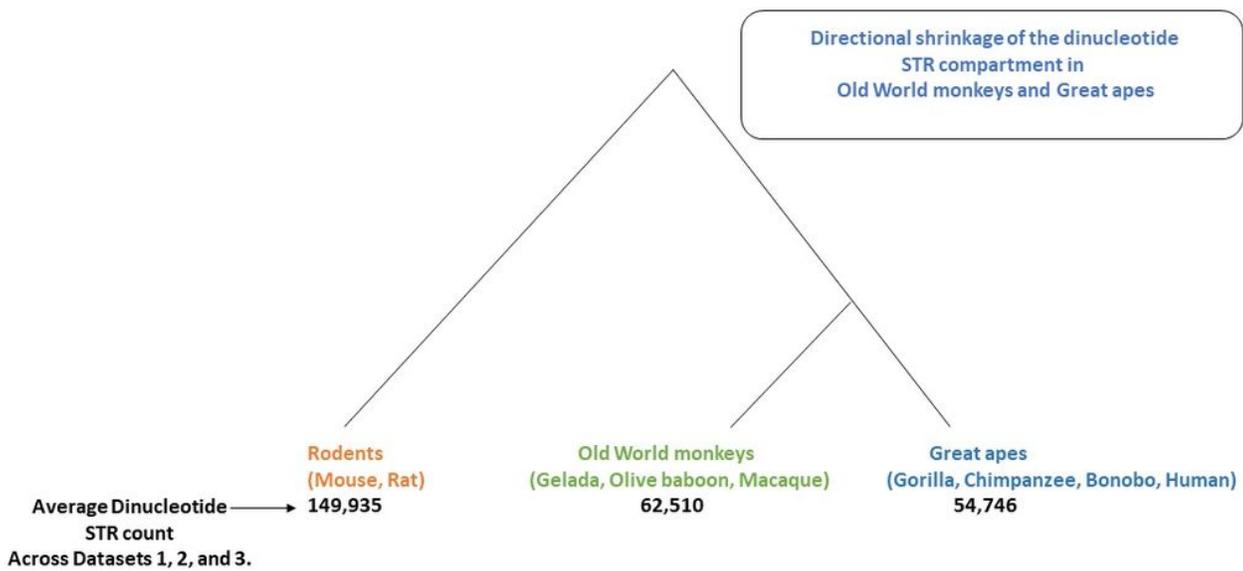


Figure 6

Proposed trend of dinucleotide STRs in primate speciation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Suppl.1.STRFinder.docx](#)
- [Suppl.2.Ninespeciesdatasets.docx](#)