Great ape mutation spectra vary across the phylogeny and the genome due to distinct mutational processes that evolve at different rates

Michael E. Goldberg[1] and Kelley Harris[1]

[1]University of Washington, Seattle, WA

Introduction

- A **mutation spectrum** comprises the relative rates at which mutations occur in different sequence contexts. We define the triplet mutation spectrum to be the distribution of 96 unique reverse-complemented mutations classified by their ancestral and derived alleles and 3' and 5' neighboring nucleotides
- Differences in mutational mechanisms lead to differences detectable when comparing mutation spectra in the soma between cells or tumors or in the germline between individuals, populations, or species
- Germline mutagenesis has evolved rapidly among great apes, leading to a species-specific mutation spectrum, or **species signature**
- Mutation spectra across a genome could be modified in *cis* (e.g., a CpG site being methylated) or in *trans* (e.g., modification of methyltransferase, affecting CpG sites genome-wide)

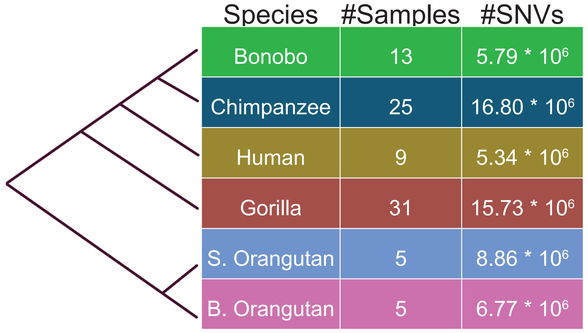
Questions

1. What mutational processes active in the germline are evolving in great apes?

2. Are the species signatures uniformly distributed across the genome?

3. Is the evolutionary landscape of mutation spectra dominated by *cis*- or *trans*-acting modifiers?

Mutation data



Species	#Samples	#SNVs
Bonobo	13	$5.79 * 10^6$
Chimpanzee	25	$16.80 * 10^6$
Human	9	$5.34 * 10^6$
Gorilla	31	$15.73 * 10^6$
S. Orangutan	5	$8.86 * 10^6$
B. Orangutan	5	$6.77 * 10^6$

- We have leveraged variation data (SNVs) from the Great Ape Genome Project (Prado-Martinez & Sudmant et al., 2013)
- Assigned ancestral states based on published annotation and excluded singletons
- Mean coverage: 23.0, short read sequencing
- SNVs are aligned to hg18; therefore, we excluded SNVs occurring in regions with no homology to humans

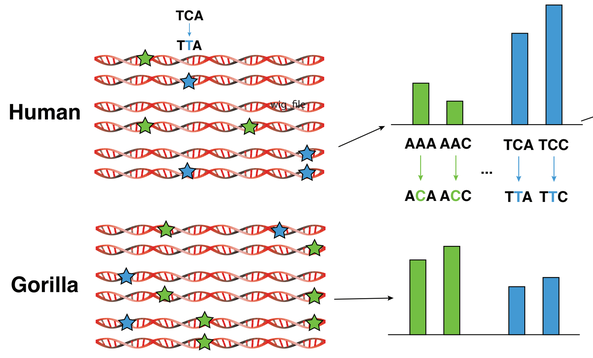
Conclusions

1. Genome-wide, the germline mutation spectra of great apes have evolved and diverged rapidly

2. The mutational processes linked to repetitive elements and genome replication are conserved

3. Different *trans*-acting mutational modifiers that affect the entire genome may evolve quickly and explain rapid divergence of mutation rate; these mutational modifiers could be either genetic or environmental

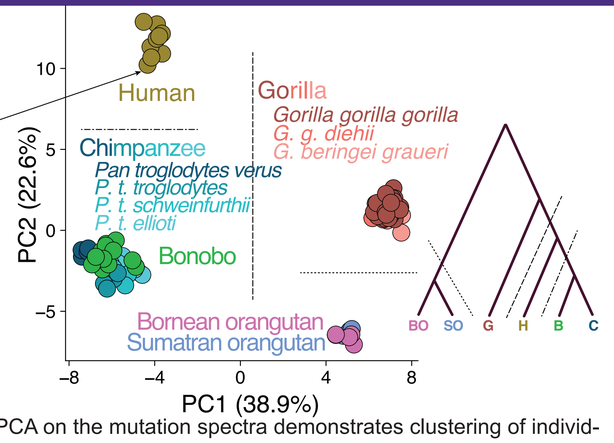
Great apes have species-specific mutational profiles



- For each individual in the GAGP, we calculated the **mutation spectrum** of non-conserved, non-repetitive (NCNR) genomic regions, or **compartment**

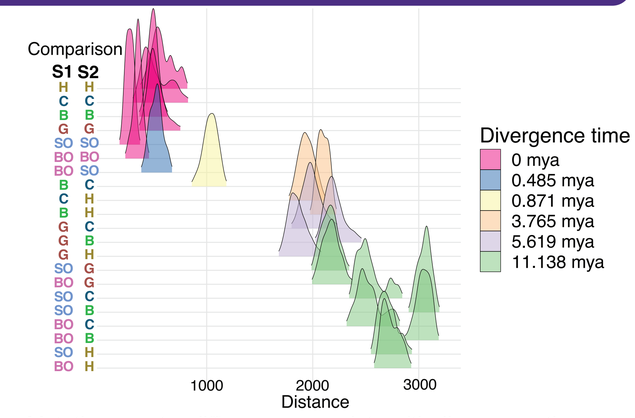
  -NCNR (1.23 Gb, 1.84*107 unique SNVs) defined as genomic regions excluding repeatMasker-defined repeats, CpG islands, phastCons-defined conserved regions, and expressed exons

  -Ancestral alleles were determined using a parsimony approach



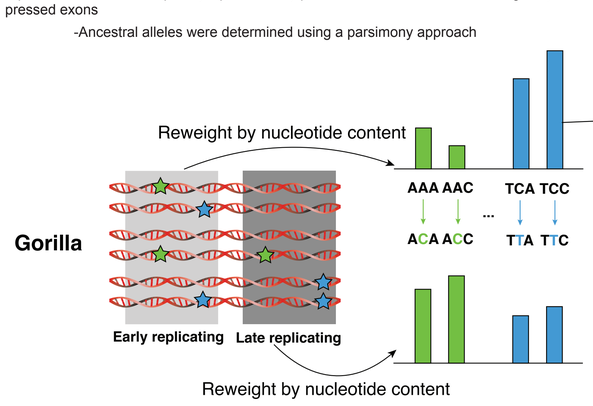
- PCA on the mutation spectra demonstrates clustering of individuals by species that reflects phylogeny

  -Calculated mutation fraction for each individual by dividing counts by total number of derived alleles



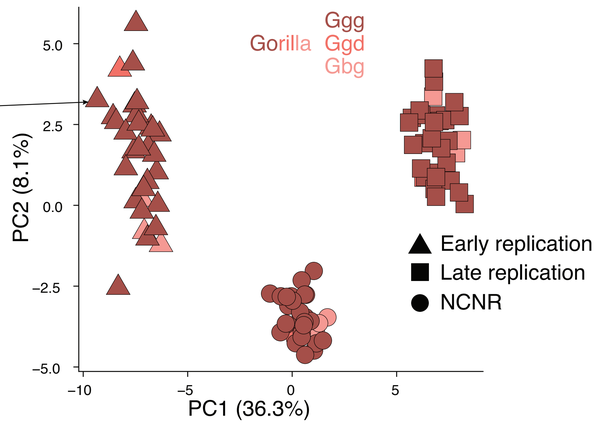
- Mutation spectra differences correlate with divergence times

  -Distribution of Euclidean distances for all intra- and interspecific comparisons

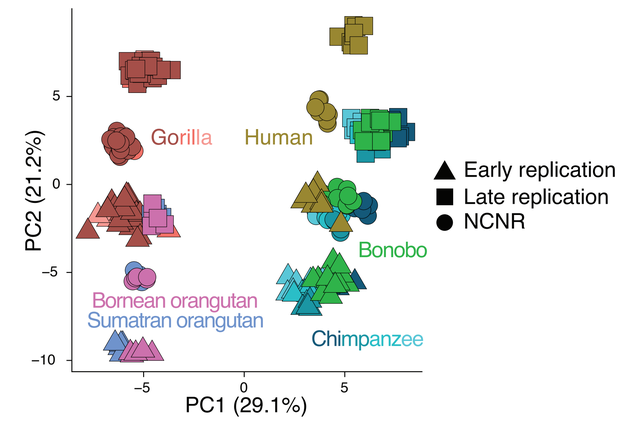


- Calculated separate mutation spectra for each individual in compartments that replicate early or late in S1 phase; reweighted spectra by nucleotide content of compartment

  -Reweighted mutation spectra in a given compartment C by multiplying each mutation fraction by the ratio of the counts of the mutating triplet in NCNR to the counts in compartment C

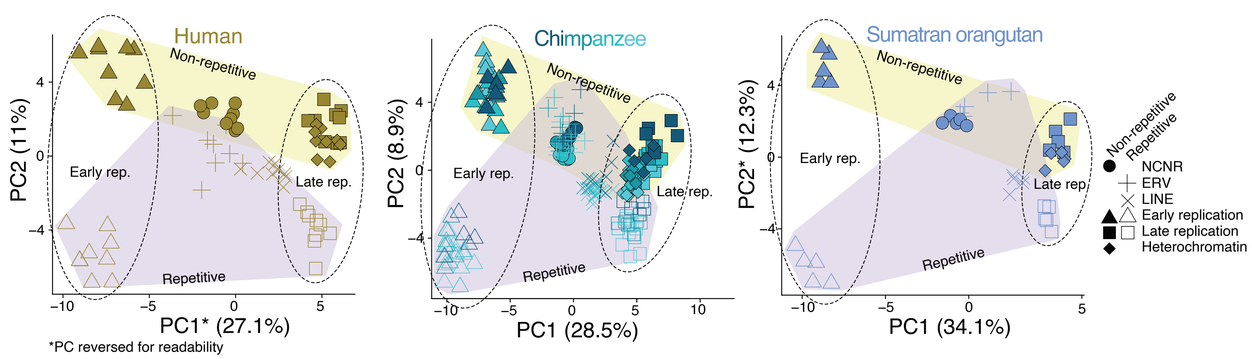


- Each point in the PCA represents the mutation spectrum from a single individual's NCNR, early replicating, or late replicating compartment
- Clustering by compartment implies differences in mutation spectra based on replication timing



- Clustering of individual mutation spectra in PCA along orthogonal "phylogeny" and "replication timing" axes; mutation signature associated with late replication timing appears conserved among all great apes

All great ape genetic variation appears to be shaped by a conserved landscape of *cis*-acting mutational modifiers



*PC reversed for readability

- Results above show that late replication timing is associated with a conserved mutational signature across great apes
- We defined eight overlapping functional compartments to test for presence of evolution of mutation spectrum modifiers along axes of chromatin accessibility, replication timing, and repetitive content
- Ran PCA for each species independently including the individual mutation spectra for all eight compartments
- PC1 and PC2 separate compartments along gradients that correspond to replication timing and repetitive content, respectively (dotted lines vs. shaded polygons)
- Similarities of independent PCAs across all species implies conservation of *cis*-acting mutational signatures