Significance

- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

Significance

- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

Significance

- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

Significance

- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

Significance

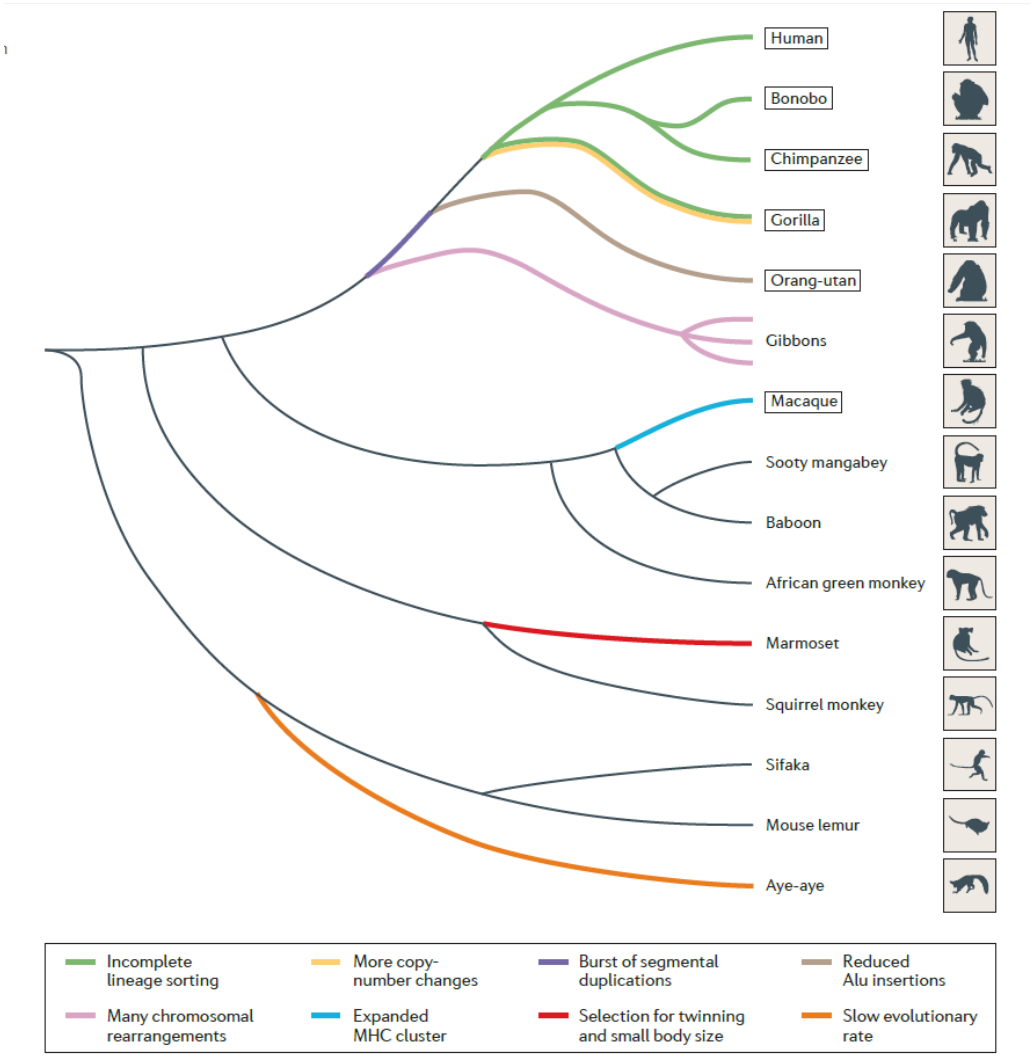
- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

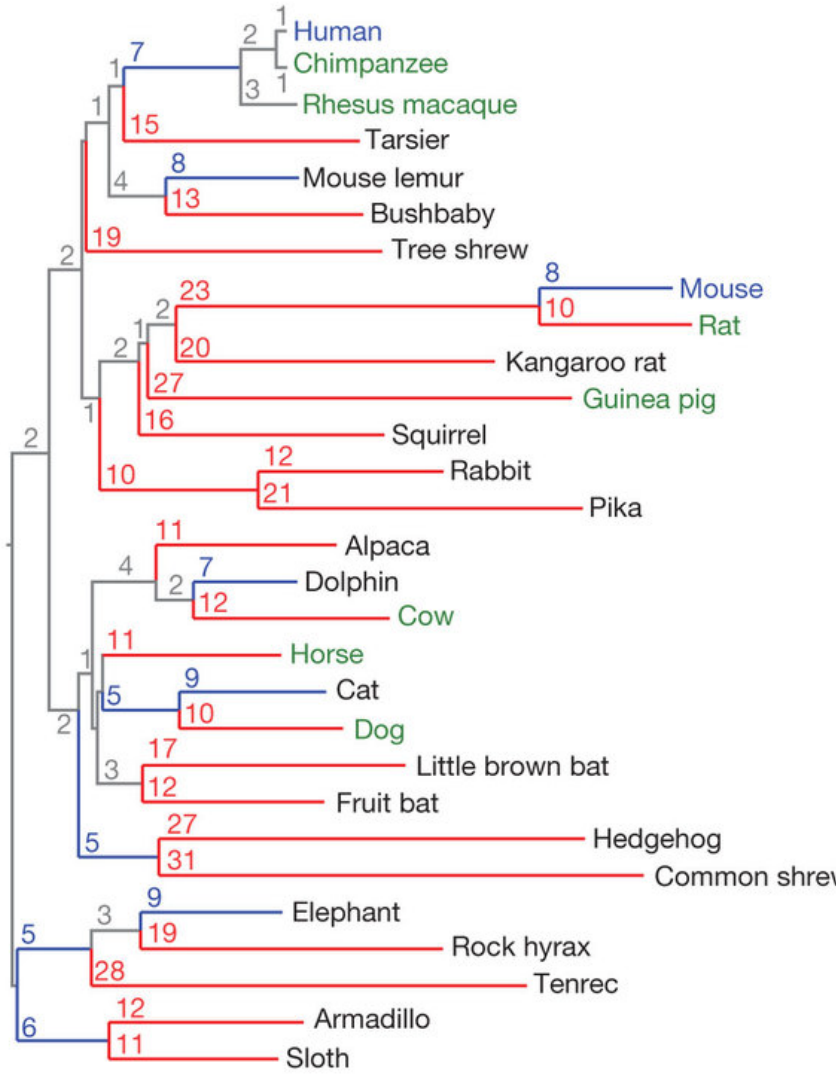
Significance

- Evolutionary processes are the foundation to understanding mutation and variation
- Genotype-phenotype correlations make sense in the context of evolution—provides an unbiased framework for discovery and prioritization of regions
- NHGRI blazed the trail in comparative genomics of complex genomes (mammalian and vertebrate genomes)..expertise/computation/resources/consortium

Achievements

- >60 vertebrate genomes and alignment with human data revealed >3 million specific evolutionarily conserved elements (~4.6% of the genome) –**Work in progress (average primate N50= ~25 kbp)**
- Evolutionary reconstructing the origin of the human genome, distinguishing ancestral features from more recent innovations---**species lacking eg. prosimians**
- Catalogs of intra-species variation in multiple primate genomes (e.g. rhesus macaque, African green monkey); genome sequences of ~200 inbred *Drosophila* strains provides basis of further quantitative genetic studies)
- Human Genome Reference Consortium—improvement of human reference**-complex structural variation, intermediate size SV and gaps remain unresolved.**





Few finished high quality assemblies

Rogers and Gibbs, 2014; Lindblad-Toh, 2011

Goal 1: *De novo* sequence and assemble high quality genomes. Advance sequencing technologies to assemble a genome for $10k at a quality exceeding current human assembly (with long-range haplotype phasing).

- Apply this technology to obtain additional human reference genomes (e.g. African) that better represent human diversity (n=50 references)
- Obtain a comprehensive assessment of all genetic variation, including intermediate-size and complex structural variants.
- Understand missing portions of the human genome—"platinum" genome assembly

Goal 2: What makes us human?

- Assign every human-lineage genomic change to a specific branch on the evolutionary tree.
- Catalog all human-specific changes with functional consequences including gene innovations.
- High-quality *de novo* assembly of non-human primates to annotate attributes of the genome that discriminate human from our closest ancestors (eg. 16 primate genomes at level of human genome reference)
- Determine where human mutational processes deviate from non-human primates, differences in ILS, gene flow, etc.
- Did evolution's first response to crises result in human-specific compromises to our genome?

Goal 3: Obtain nucleotide-level resolution of every conserved functional element in humans.

- Sequence with existing technology additional genomes of 200 mammals as a pilot, later to scale up (eventually 5400)
- Identify lineage-specific constraints on all functional elements (including genes, regulatory elements, ncRNAs, etc.) at the single basepair level
- Quantify the selective constraints on each element across mammals and integrate with ENCODE
- Advance computational technologies for inferring constraint trained on comparative functional genomics data

Goal 4: Leverage the power of model organisms for functional genomics.

- Apply large scale genomic sequencing to reference panels (e.g. Diversity Outbred mice).
- Produce and validate large scale panels of human mutations (via CRISPR) in the mouse/zebrafish model and phenotype them for functional inference.
- Assess Genotype x Environment interaction at scale across these panels.
- Use comparative genomics to identify informative mutations to introduce into human cell lines.

Goal 5: Develop informatics infrastructure to produce, display and quantify multiple species genome alignments.

- Produce algorithms and software for alignment (local and whole genomes), and for computational tracing of lineage-specific functional elements.
- Develop browsers on multiple coordinate systems
- Devise methods for analysis of of complex chromosomal rearrangements and duplications.
- Produce benchmarks and quality control metrics to assess accuracy of comparative genome analysis.

Summary

- Evolution is the single most powerful unifying principle in biology including genetics.
- We need to develop technologies to obtain high quality *de novo* references sequences and apply to multiple humans (**Goal 1**).
- Target multiple primate genomes to infer with high confidence all human-specific genome attributes (**Goal 2**).
- Sequencing 200 mammals with current technologies will allow inference of human conserved functional elements (**Goal 3**).
- Model organisms are still of enormous utility for understanding context-dependence of variant functions. Sequencing reference panels of model organisms will accelerate our understanding of the mapping from genotypic to phenotypic variation (**Goal 4**).
- There continues to be an urgent need for better software for genome alignment and for providing user interfaces to understand and manipulate comparative genome data (**Goal 5**).
- NHGRI leadership role is critical —if not NHGRI who else?