# Genetic evidence for complex speciation of humans and chimpanzees

- Introduction
  - t_genome: average genetic divergence
  - t_species: speciation time
  - t(x): genetic divergence at position x
    - t(x) fluctuates throughout the genome and is always equal to or greater than t_species because t_species represents the barrier to coalescence
- Inferring ancient speciation from genetic data
  - One can infer t_species from the distribution of t(x) times but there are several issues that need to be counted for:
    1. Genetic divergence should be corrected for local variation in mutation rates, which can be done by comparing divergence with an outgroup
    2. Genetic divergence should be corrected for recurrent mutations
    3. To study variation in genetic divergence one needs measurements for large subsets of the genome
  - The most common topology is HCG but 18–29% of the genome has a HGC or CGH which implies that humans and chimps must have speciated before gorillas—ie we get these discordant topologies through ILS
- Genome comparisons of five primates
  - Generated HCGM and HCGOM alignments
  - Need to correct for recurrent mutations on divergence times/branch lengths estimates
    - Specifically HG and CG topologies need corrections as estimates without correcting for multiple mutations would leave to overestimates of branch lengths
      - That is most of these sites are produced by recurrent mutation while a small fraction is ILS
    - This can be done by utilizing the HCGOM alignment as some topologies are impossible to explain without recurrent mutation
- Large variation in divergence time across genome
  - Calculated the average autosomal HC divergence and then study subsets of the genome as a function of distance from specific site patterns:

- Distance from HC sites: relative HC divergence decreases as you approach HC sites
  - This makes sense since you are likely to have low divergence near t_species(HC)
- Distance from HG/CG sites: relative HC divergence increases as you approach HG/CG sites
  - This makes sense because the normal topology should be HCG, but if you are near an HG/CG site humans and chimps are inherently more distantly related
    - After accounting for recurrent mutations increases, this relative estimate as recurrent mutations will look like smaller divergence
- Large reduction in divergence time on chromosome X
  - Relative HC divergence among autosomal chromosomes largely reflect the autosomal average
  - Relative HC divergence is significantly less than expected on the X
    - HG divergence on the X closely mirrors the expectation which invalidates a slowdown of mutations on the X as it would also result in a lower HG X chromosome divergence
  - Absence of HG and CG sites at comparable rates to the autosomes on the X chromosome provides another line of evidence for decreased HC divergence on the X chromosome
    - The presence of HG/CG sites would inherently increase HC divergence
- Implications for current models of human–chimp speciation
  - The existing claims based on the fossil record about HC speciation times seem to be at odds with the large variance in HC autosomal divergence and reduced divergence on the X chromosome
  - Based on inferred divergence times between HO and HM it is unlikely that the fossil record reflects the true HC speciation time
  - The relative ratio (R) of X_div:Autosomal_div is expected to be 0.75
    - R roughly equals 0.75 among humans, chimpanzees, HG, and HB
    - R is between 0–0.29 for HC
      - This could be explained by strong selection on the X chromosome
- Possible hybridization in the human–chimp lineage
  - Divergence with subsequent gene flow could explain the patterns of HC divergence

    - This would explain the inconsistency with the fossil record, where fossils that predate human speciation still have hominid features
    - This would explain the large variance in HC divergence
        - Where some loci reflect the initial HC separation (older HC divergence) and other loci reflect the more recent HC hybridization (younger HC divergence)
    - This would explain abnormally low HC divergence on the X chromosome
- One would expect rapid selection on the X chromosome under a model of divergence with secondary contact
    - Haldane's rule predicts that the majority of hybrid incompatibilities would be on the X chromosome or X-linked
    - Selection against multiple hybrid incompatibility loci would result in an overall lower divergence on the X chromosome
    - Hypothetical scenario:
        1. Human and chimp ancestors diverged and then interbred hybrid males would be sterile as predicted by Haldane's rule
        2. A viable population could then only persist through fertile hybrid females backcrossing with on of the parental populations
        3. The subsequent male progeny of F1 female x parental species, would carry X chromosomes derived nearly from the ancestral population