

Phylogeny and Evolution

Area Co-Chairs

Dannie Durand,

Carnegie Mellon University

Olivier Gascuel,

Centre National de la Recherche Scientifique, Montpellier

Evolution in the Genomic Era

Sequence

Genes

Genomes

Organisms

Populations

Species

What are the relationships between taxa?

What events can occur and what are the rates?

How is evolution related to function and morphology?

Computational Challenges

Sequence

Genes

Genomes

Organisms

Populations

Species

Scale

Accuracy

*Modeling
Complexity*

Evolution in the Genomic Era

Sequence

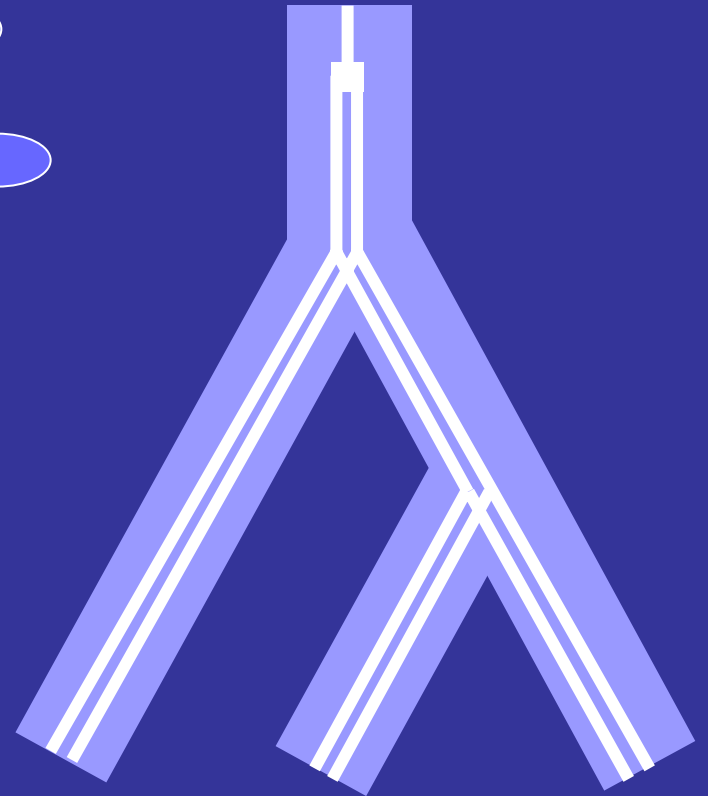
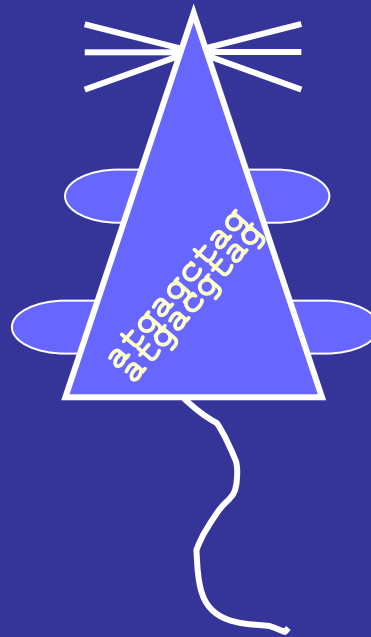
Genes

Genomes

Organisms

Populations

Species



Sequence Evolution

Relate observed change to phenotypic differences

→ between species

*Identification of Nine **Human-specific Frameshift Mutations** by Comparative Analysis of the Human & the Chimpanzee Genome Sequences, Hahn & Lee*



Human and chimpanzee orthologs have >99% amino acid similarity, on average.

King and Wilson, *Science*, 1975

*Identification of Nine **Human-specific Frameshift Mutations** by Comparative Analysis of the Human & the Chimpanzee Genome Sequences, Hahn & Lee*

What sequence features account for human-specific traits?

- **Goal:** Identify human specific mutations
- **Frameshift errors:** Small changes in DNA sequence → large changes in protein
- **Challenge:** Distinguish mutations from sequencing errors in draft Chimpanzee sequence

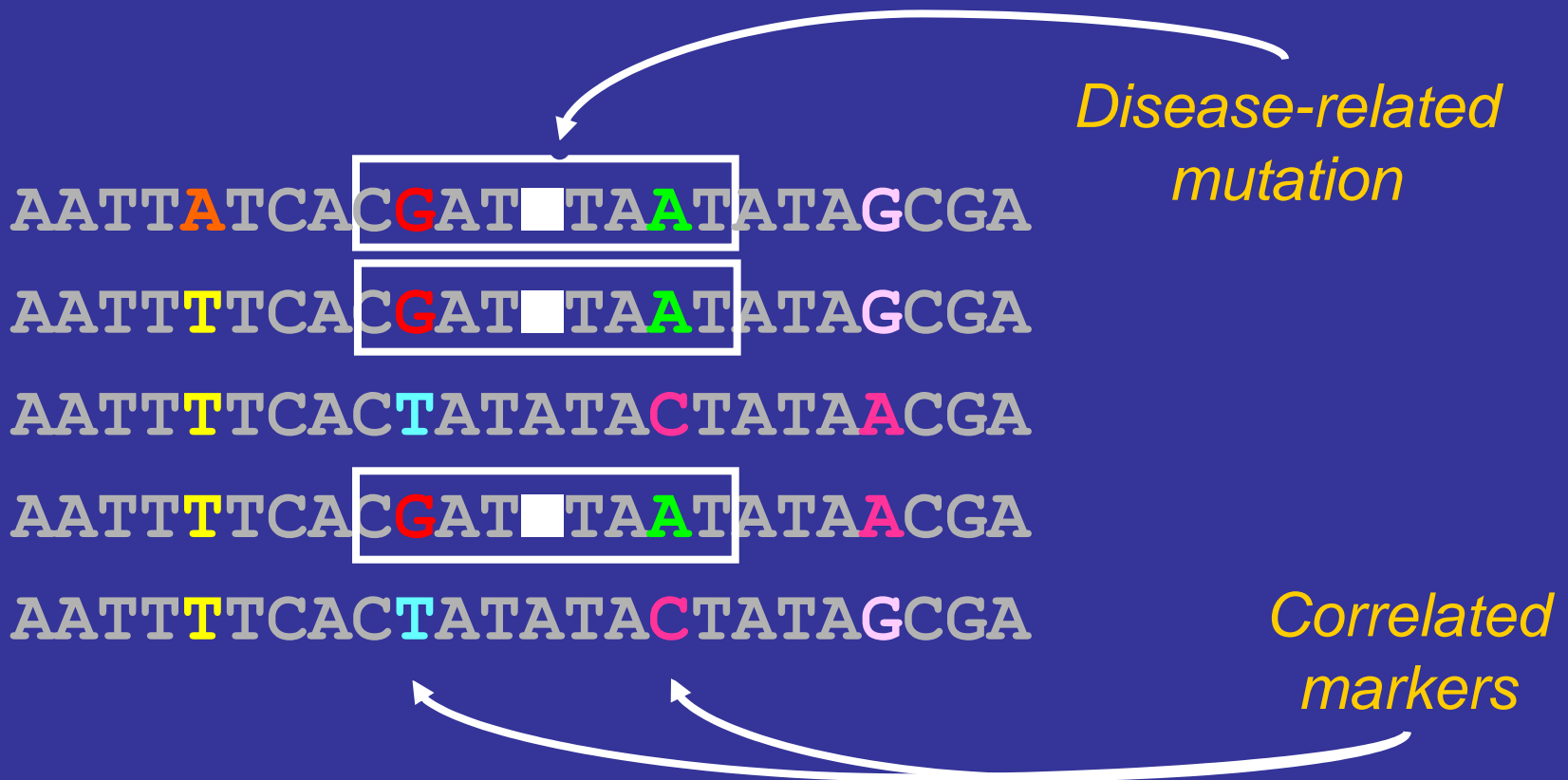
Sequence Evolution

Relate observed change to phenotypic differences

→ between species

→ between individuals within a population

Correlated sites: Association Studies using Single Nucleotide Polymorphisms



Correlated sites: Association studies

- Search for neutral markers (SNPs) correlated with mutations that cause disease.

AATT**T**TCAC**G**AT■TA**A**TATAGCGA

- Compare SNP frequencies in affected and un-affected individuals
 - Most effective if most disease-related mutations are linked to neutral markers
- **Recombination should not be too frequent**

*Efficient Computation of Close Lower and Upper
Bounds on the **Minimum Number of Recombinations**
in Biological Sequence Evolution, Song, Wu and
Gusfield*

*Recombination
event*

AATTATCACGAT■TAATATAGCGA

AATTTCACGAT■TAATATAGCGA

AATTTCACGATATACTATAACGA

AATTTCACTAT■TAATATAACGA

AATTTCACTATATACTATAGCGA



Sequence Evolution

Relate observed change to phenotypic differences

→ between species

→ between individuals within a population

Dependence between sites

→ due to common ancestry

Sequence Evolution

Relate observed change to phenotypic differences

→ between species

→ between individuals within a population

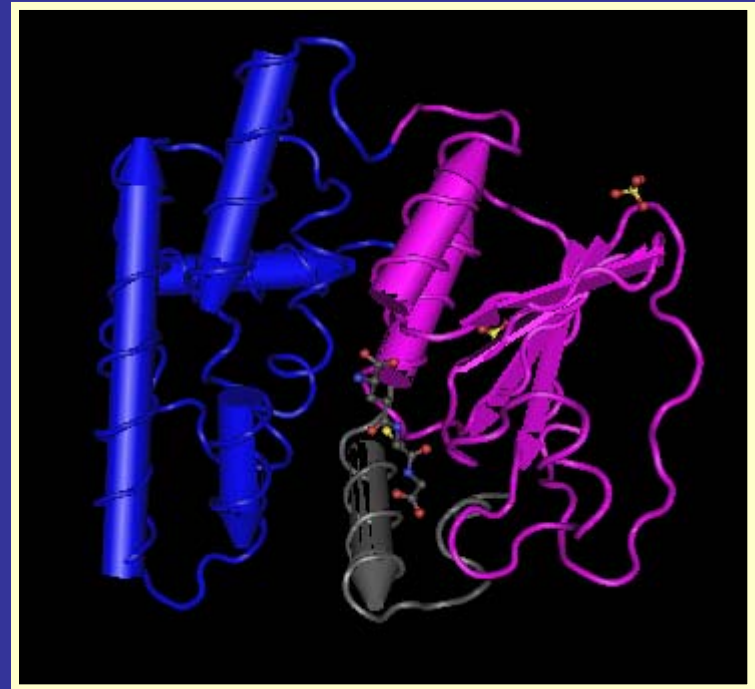
Dependence between sites

→ due to common ancestry

→ due to functional selection

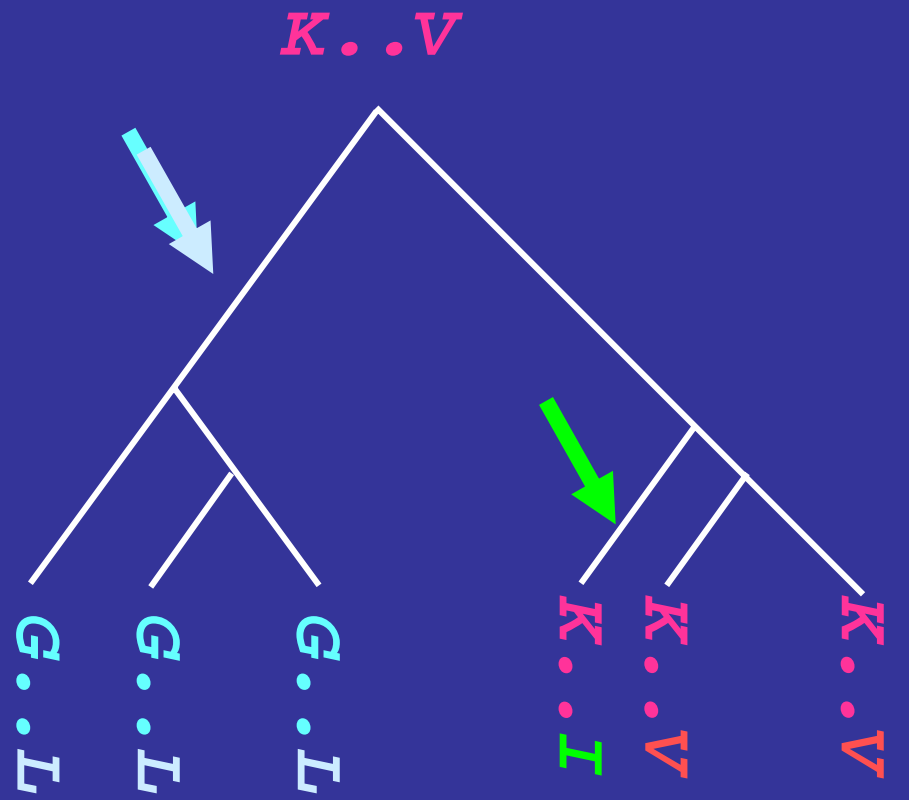
Sequence correlations can be evidence of functional selection or

... LGEENFKALV ...
... VGEEHFIGLV ...
... LGEQYFKGLV ...
... LKEETFKAIV ...
... LKEETFKAVV ...
... LKEETFKAVV ...



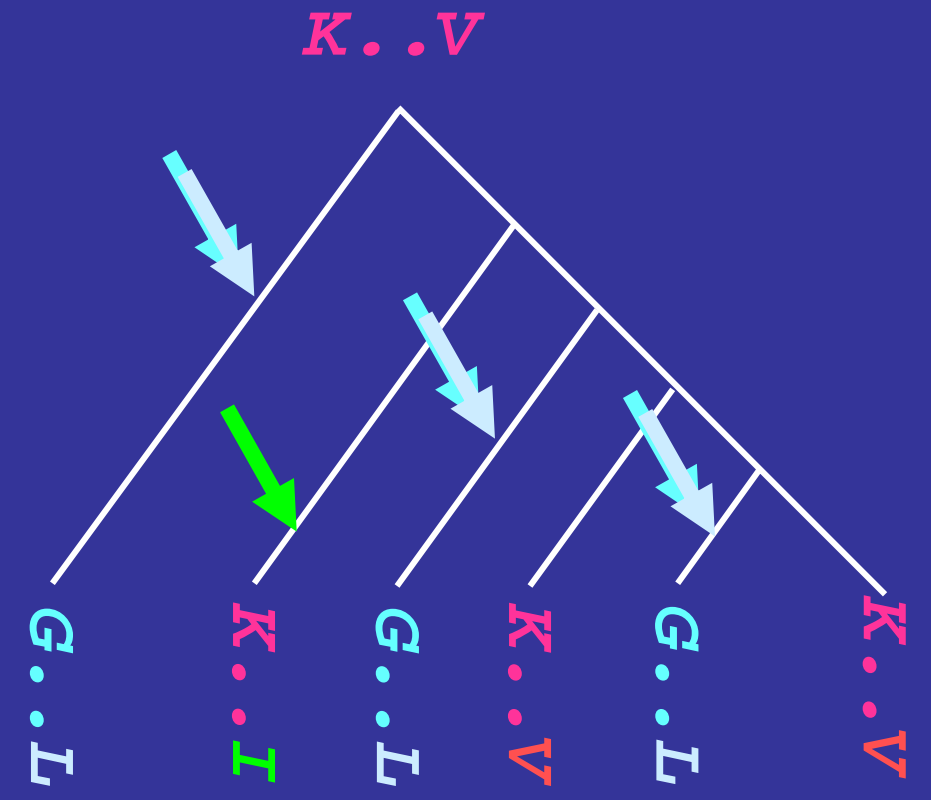
Sequence correlations can be evidence of functional selection or shared ancestry

... LGEENFKALV ...
... VGEEHFIGLV ...
... LGEQYFKGLV ...
... LKEETFKAIV ...
... LKEETFKAIV ...
... LKEETFKAIV ...



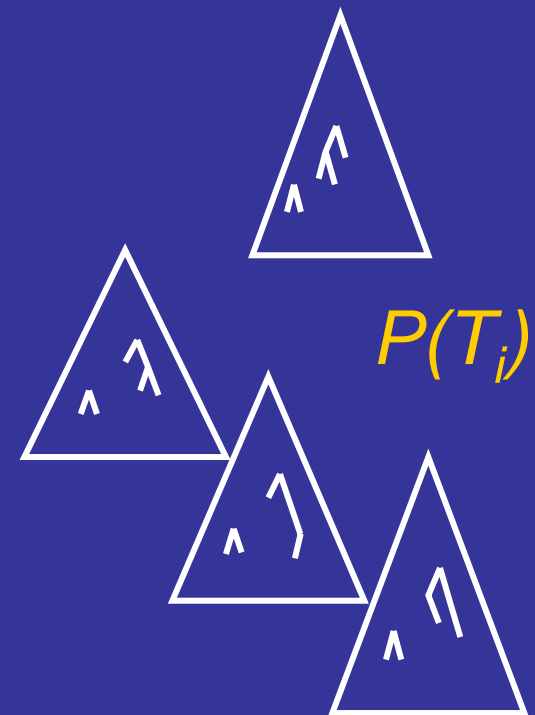
Sequence correlations can be evidence of functional selection or shared ancestry

... LGEENFKALV ...
... VGEEHFIGLV ...
... LGEQYFKGLV ...
... LKEETFKAIV ...
... LKEETFKAVV ...
... LKEETFKAVV ...



Detecting Coevolving Amino Acid Sites using Bayesian Mutational Mapping, Dimmic, Hubisz, Bustamente, Nielsen

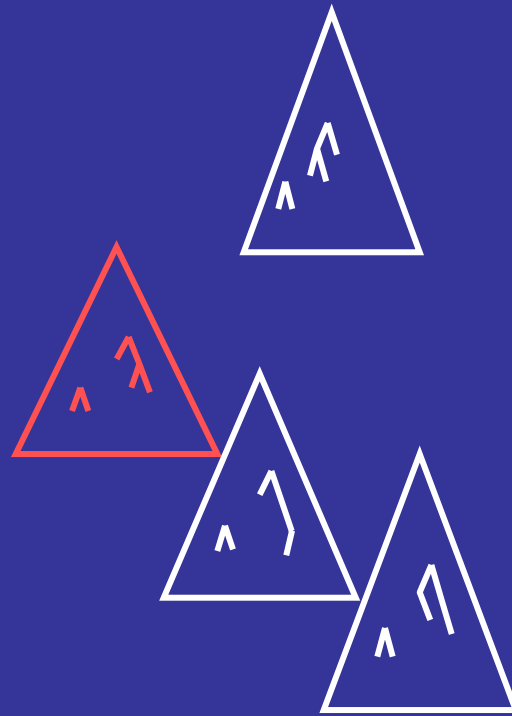
... LGEENFKALV ...
... VGEEHFIGLV ...
... LGEQYFKGLV ...
... LKEETFKAIV ...
... LKEETFKAIV ...
... LKEETFKAIV ...



$$\sum_i \text{CoEvolution}(T_i) P(T_i)$$

Maximum Likelihood of Evolutionary Trees: Hardness and Approximation, Chor & Tuller

$$L = P(\text{Data} \mid \text{Tree, rates})$$

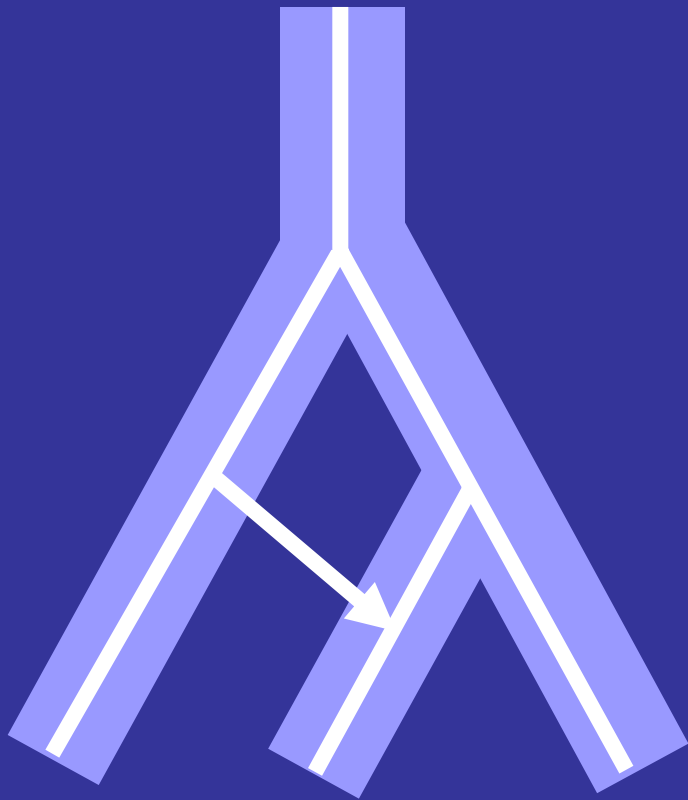


$P(T_i)$

Phylogenetics: Emerging Problems

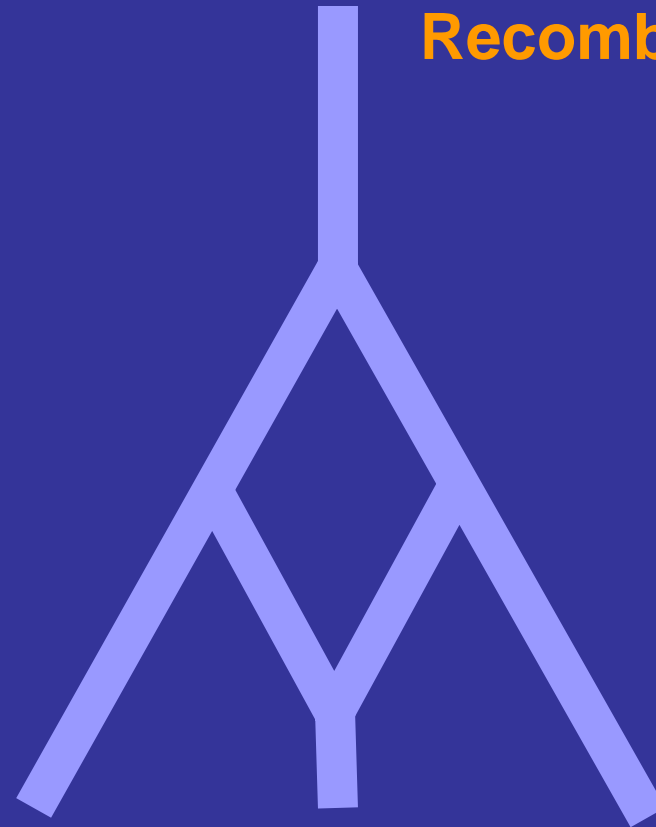
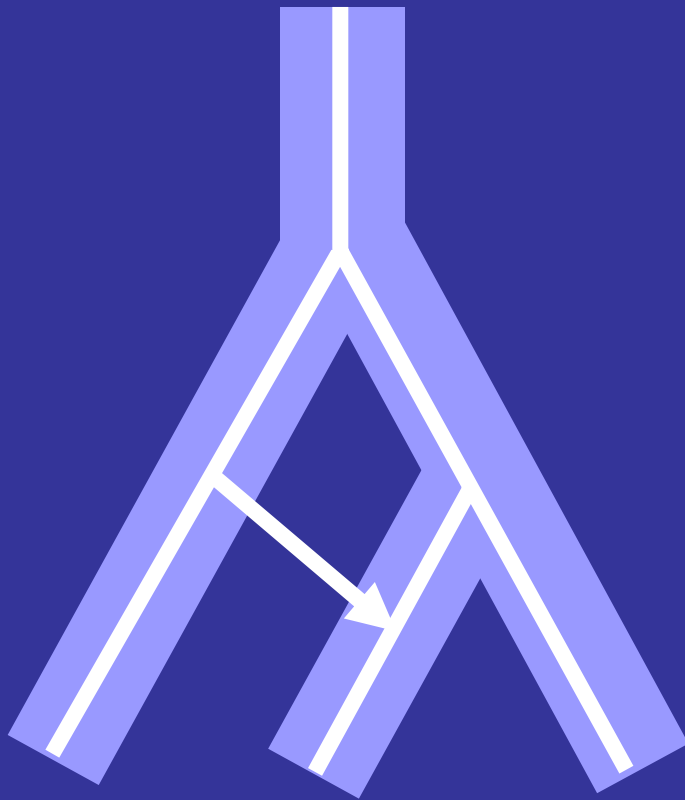
- Probability distribution over tree space
- Inferring reticulated networks

Evolutionary Networks: Is evolution tree like?



Evolutionary Networks: Is evolution tree like?

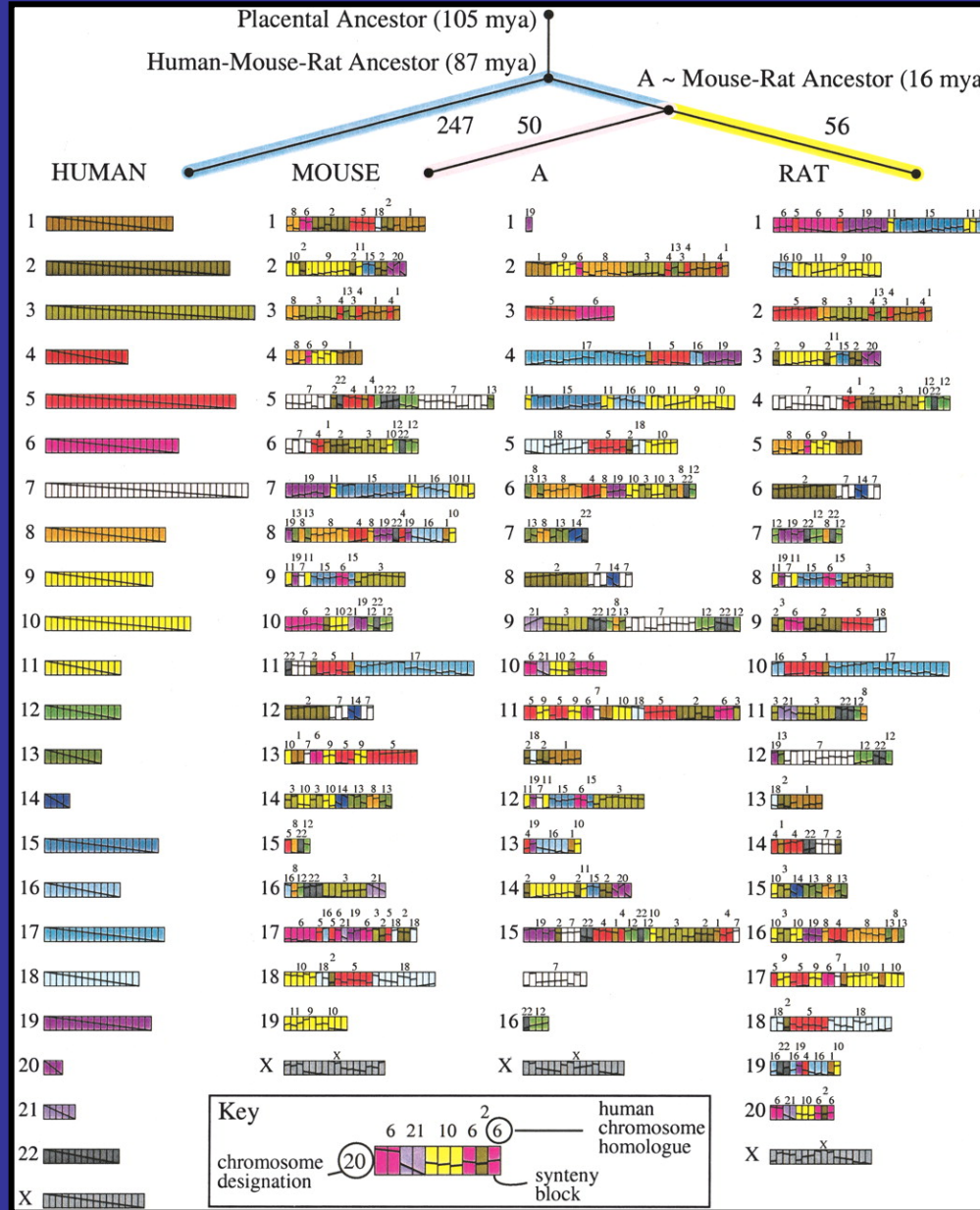
Horizontal Gene Transfer
Genome Hybridization
Recombination



Phylogenetics: Emerging Problems

- Probability distribution over tree space
- Inferring reticulated networks
- Species phylogeny reconstruction:
 - Multi-gene phylogenies
 - Breakpoint distance
 - Presence and absence of genes

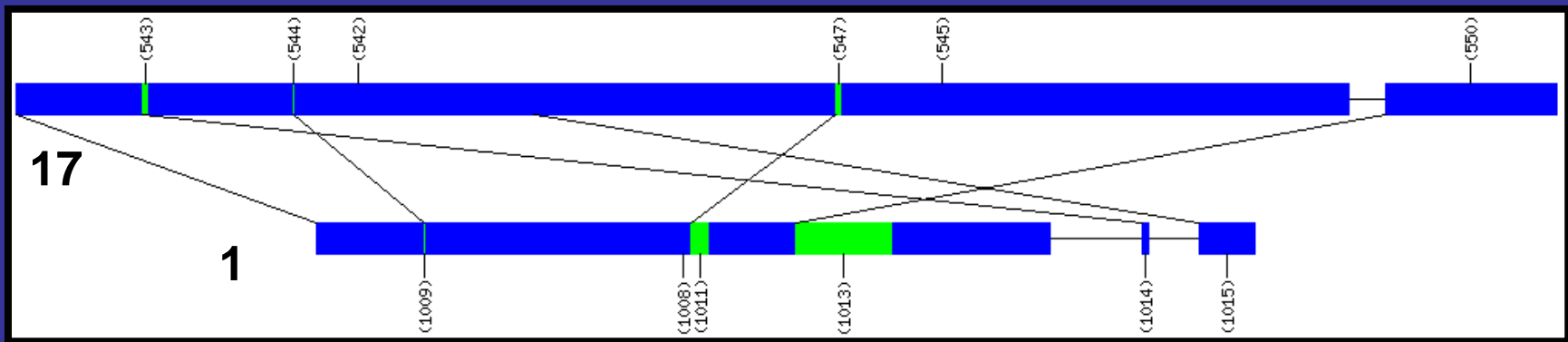
Ancestral Murid Rodent Genome & Evolutionary Tree



Guillaume Bourque et al. Genome Res. 2004; 14: 507-516



Genome Evolution and Organization: Emerging Problems



McLysaght, Hokamp, Wolfe, *Nature Genetics*, 2002

- Paralogy versus orthology
- Identifying conserved homologous regions
- Identifying breakpoints

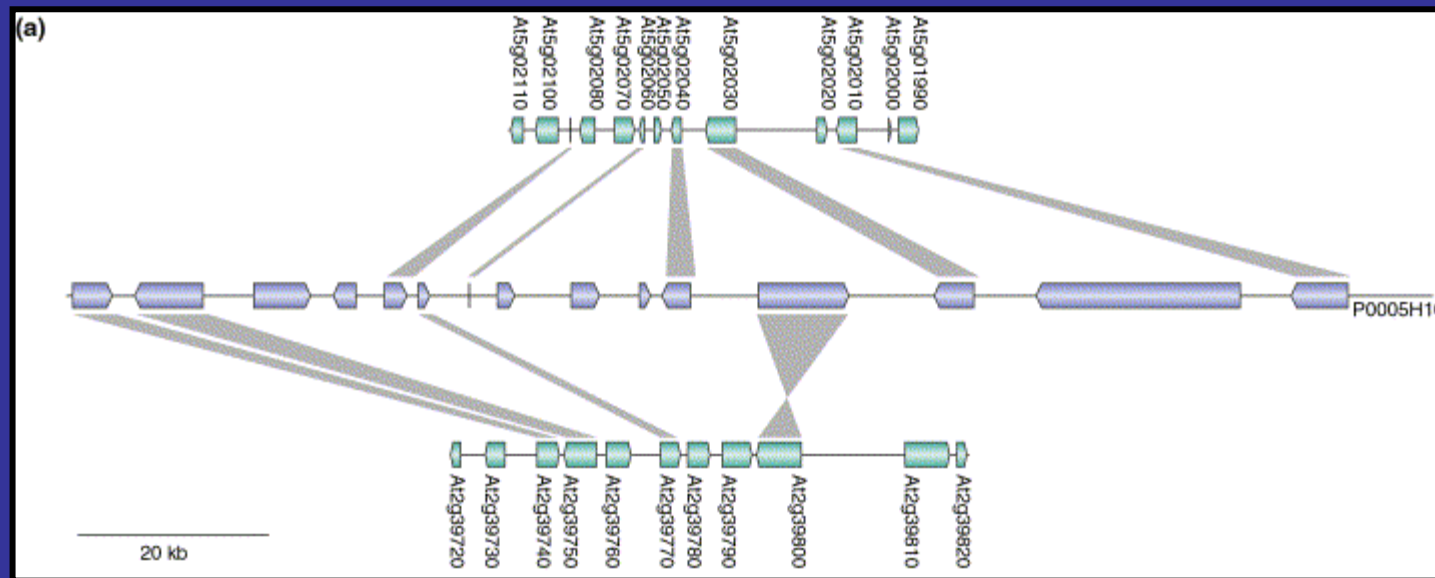
- Inferring ancestral maps
- Rates of rearrangement
- Breakpoint reuse

Putative mammalian ancestor



International Chicken Genome Sequencing Consortium, Nature, 2004

Genome Evolution and Organization: Emerging Problems



Vandepoele *et al.*, *Trends Gen.*, 2002

- Multi-genome homologies

Thanks

- Mathieu Blanchette
- Guillaume Bourque
- David Bryant
- Rick Desper
- Dannie Durand
- Martin Farach
- **Olivier Gascuel**
- Manolo Gouy
- Mike Hendy
- Ian Holmes
- Susan Holmes
- Laurence Hurst
- Jens Lagergren
- James McInerney
- Vincent Moulton
- Christian N. S. Pedersen
- Edward Susko
- Simon Whelan
- Ken Wolfe