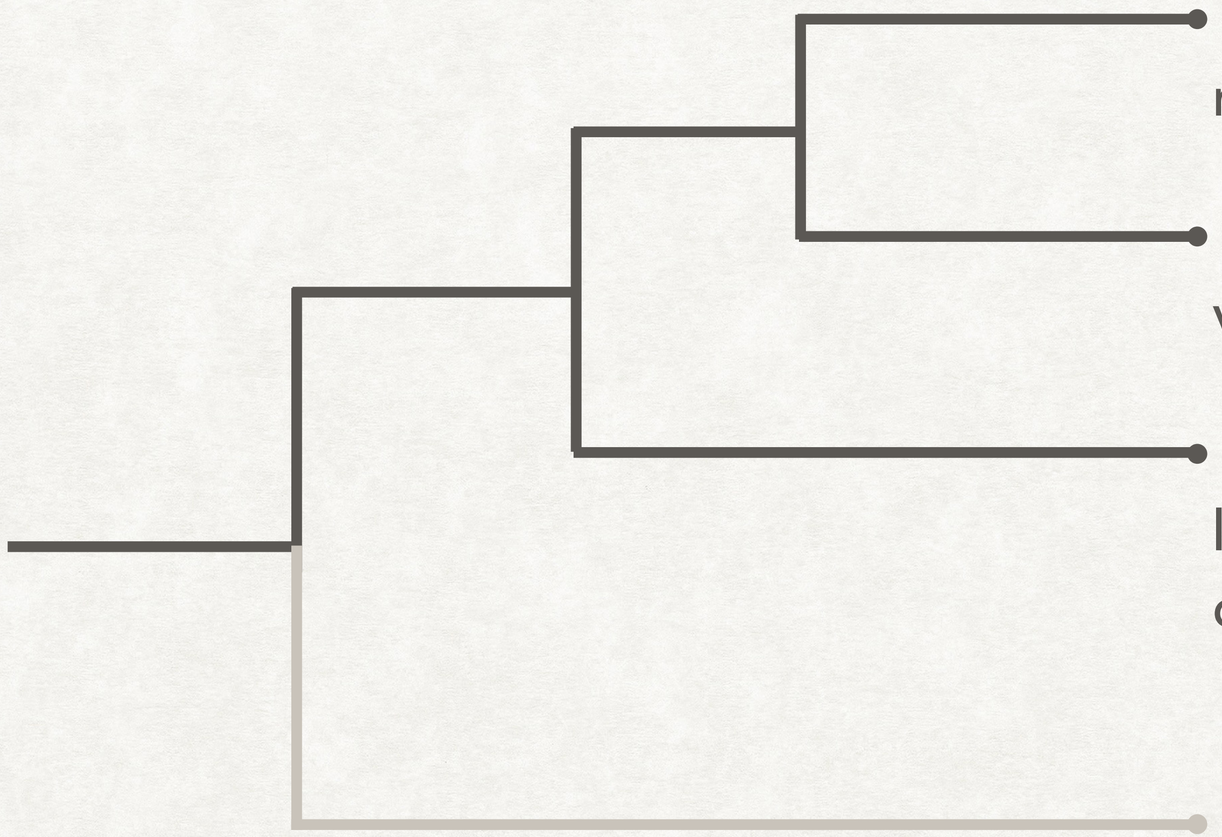
A QUICK INTRO TO:

PHYLOGENOMICS

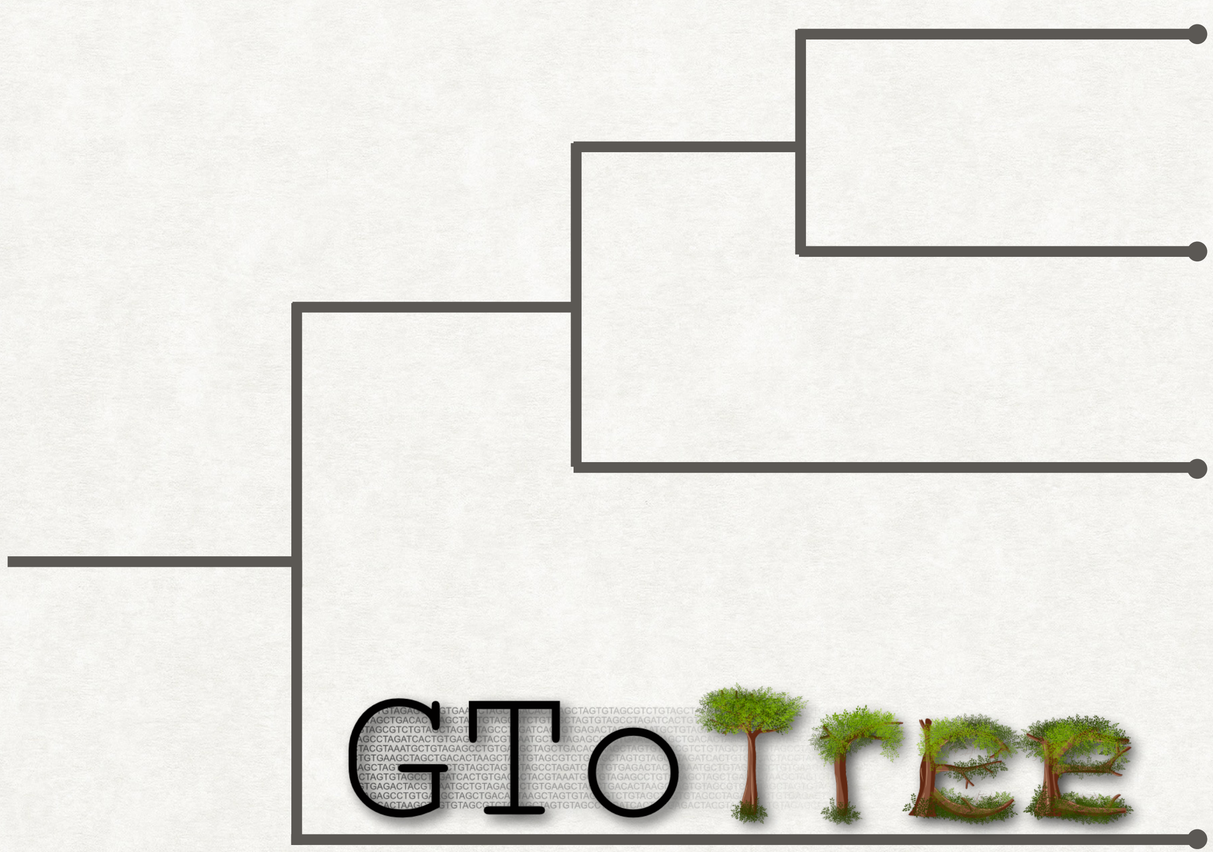
MARIA VALDEZ CABRERA - POSTDOCTORAL FELLOW

OBJECTIVES FOR THE HOUR



- Getting phylogenetic trees on your mind.
- Giving you a statistician's point of view of these objects.
- Discussing the potential and limitations of phylogenetic tree construction for microbiome data
- Introducing you to the fabulous GToTree workflow by Mike Lee

OBJECTIVES FOR THE HOUR

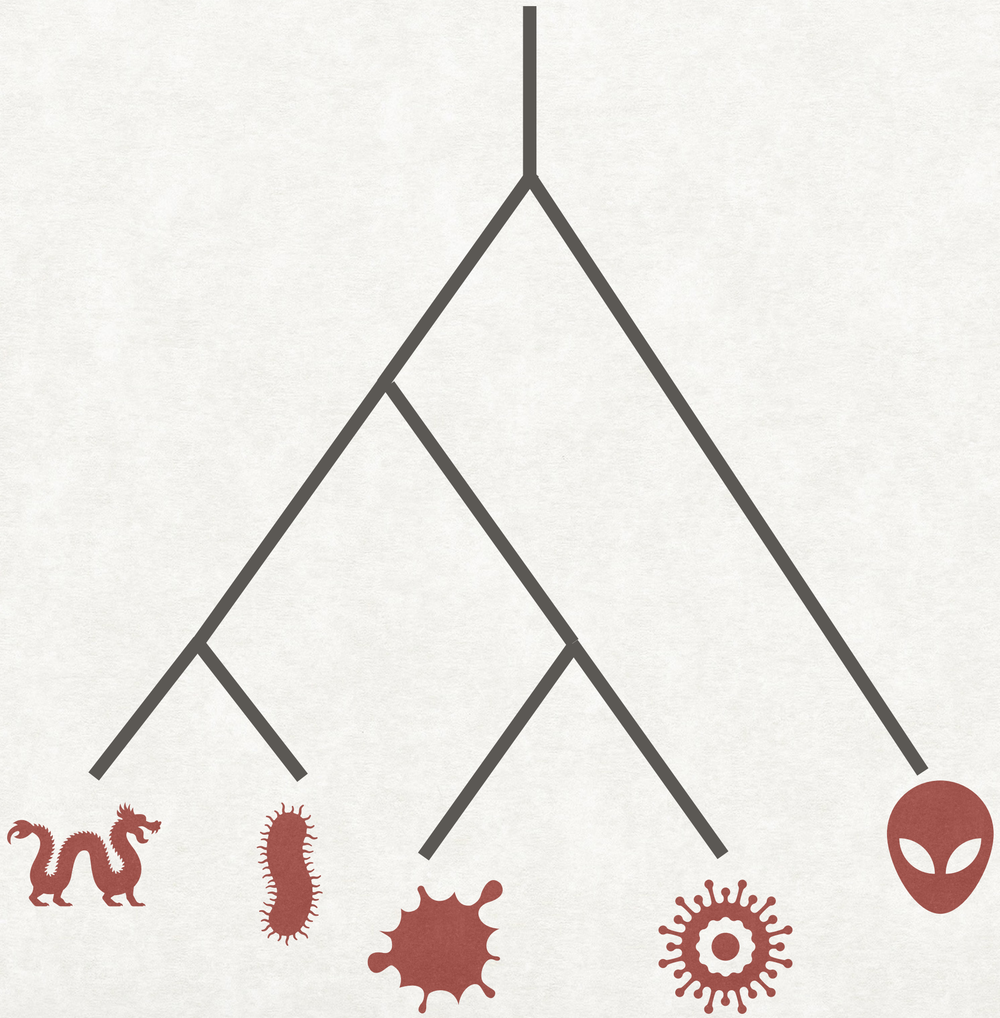


- Getting phylogenetic trees on your mind.
- Giving you a statistician's point of view of these objects.
- Discussing the potential and limitations of phylogenetic tree construction for microbiome data
- Introducing you to the fabulous GToTree workflow by Mike Lee

GToTree

@AstrobioMike

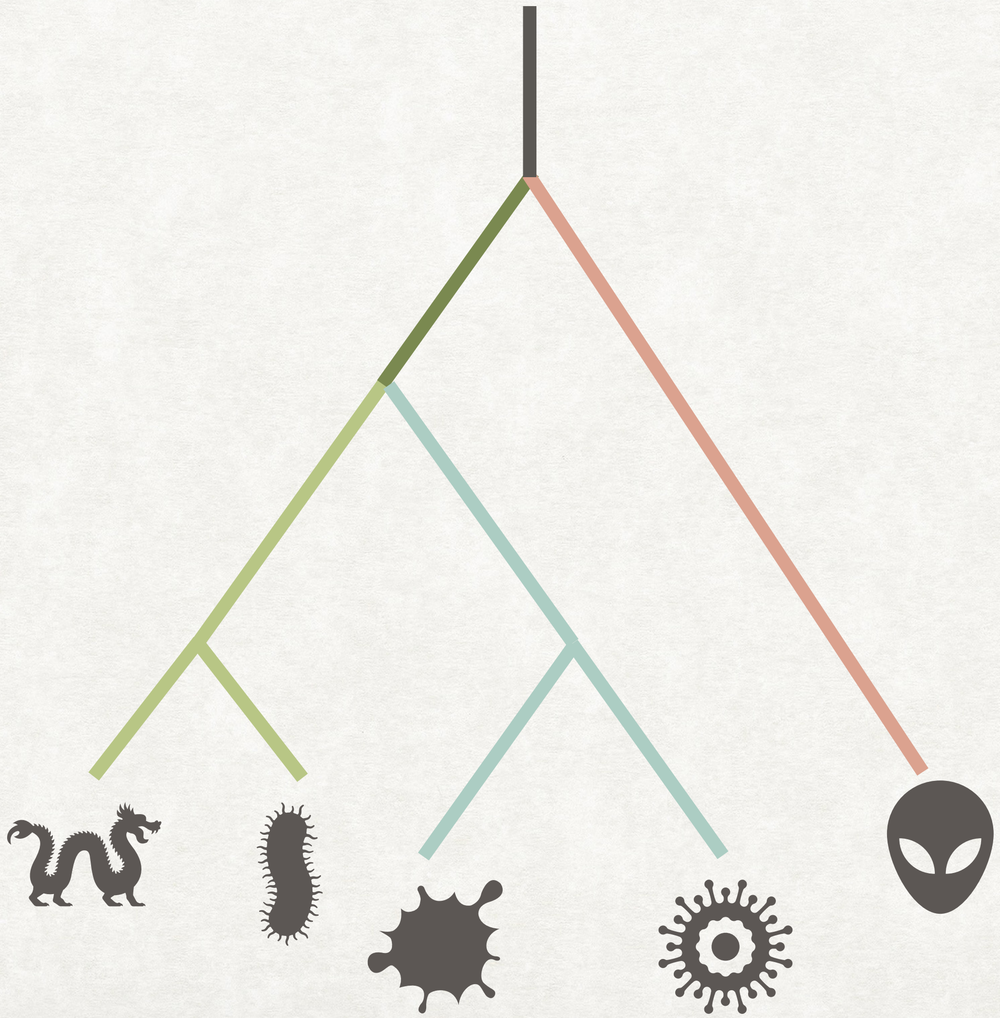
WHAT EVEN IS A PHYLOGENETIC TREE?



Components:

- Leaves
- Shape (topology)
- Branch lengths
- Root

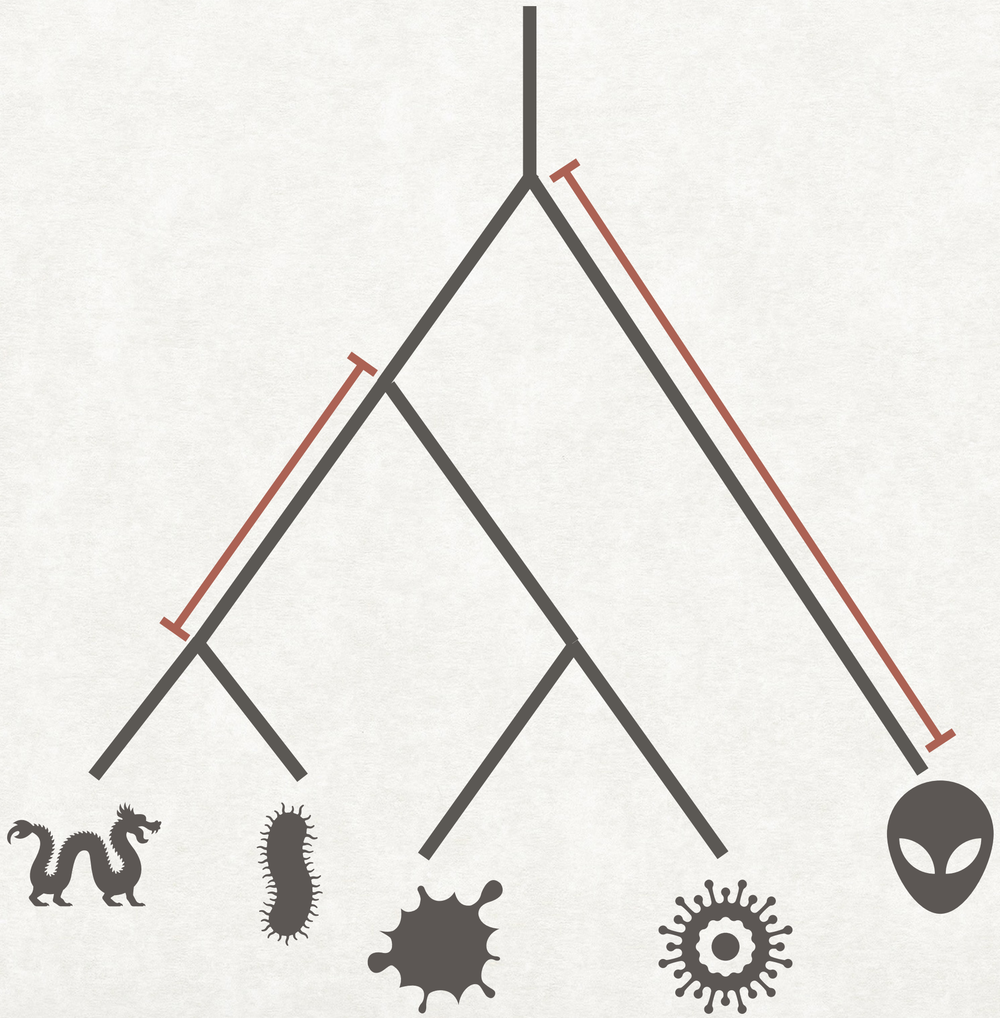
WHAT EVEN IS A PHYLOGENETIC TREE?



Components:

- Leaves
- Shape (topology)
- Branch lengths
- Root

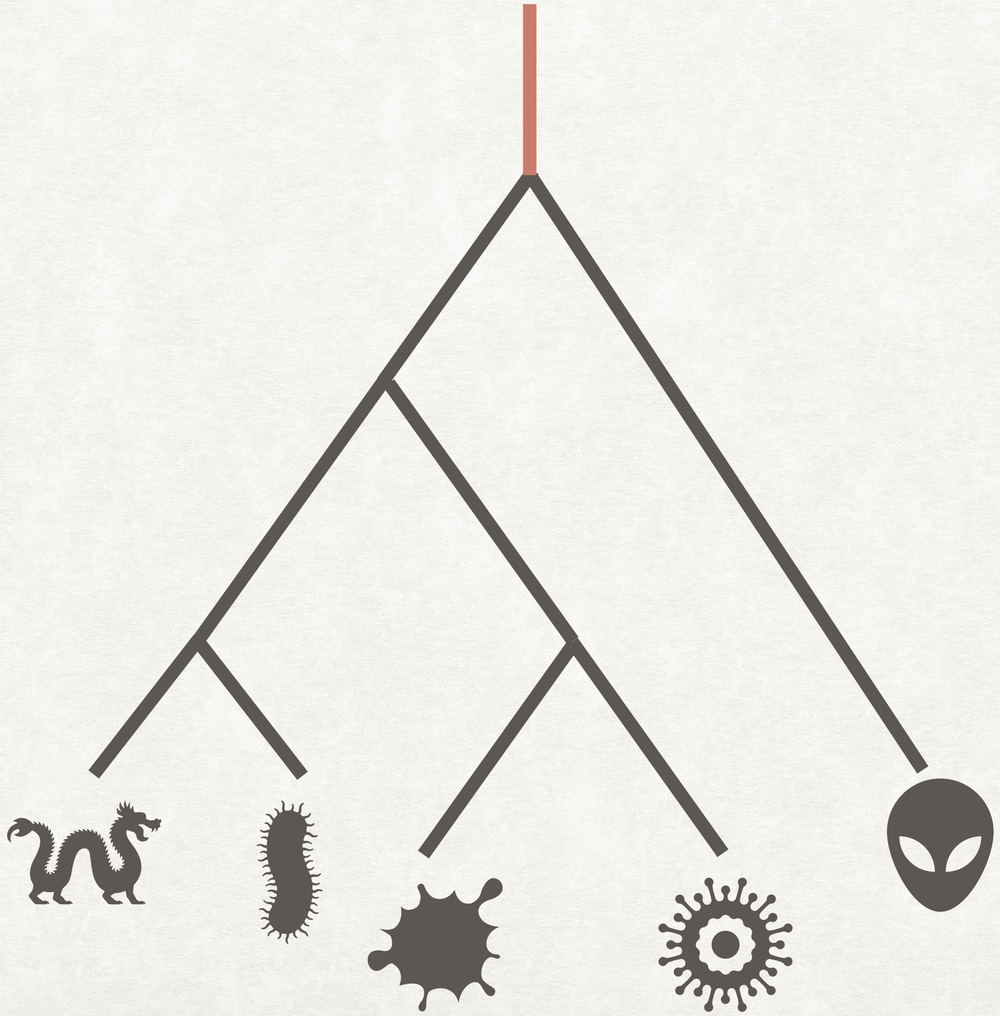
WHAT EVEN IS A PHYLOGENETIC TREE?



Components:

- Leaves
- Shape (topology)
- **Branch lengths**
- Root

WHAT EVEN IS A PHYLOGENETIC TREE?



Components:

- Leaves
- Shape (topology)
- Branch lengths
- **Root**

ACTIVITY ALERT

What scientific questions can be explored with Trees?

ACTIVITY ALERT

What scientific questions can be explored with Trees?

Taxonomy

Contents lists available at ScienceDirect

Diagnostic Microbiology and Infectious Disease

journal homepage: www.elsevier.com/locate/diagmicrobio

Review Article

Taxonomic update on proposed nomenclature and classification changes for bacteria of medical importance, 2015

J. Michael Janda *

Public Health Laboratory, Department of Public Health, Kern County, Bakersfield, CA 93306-3302

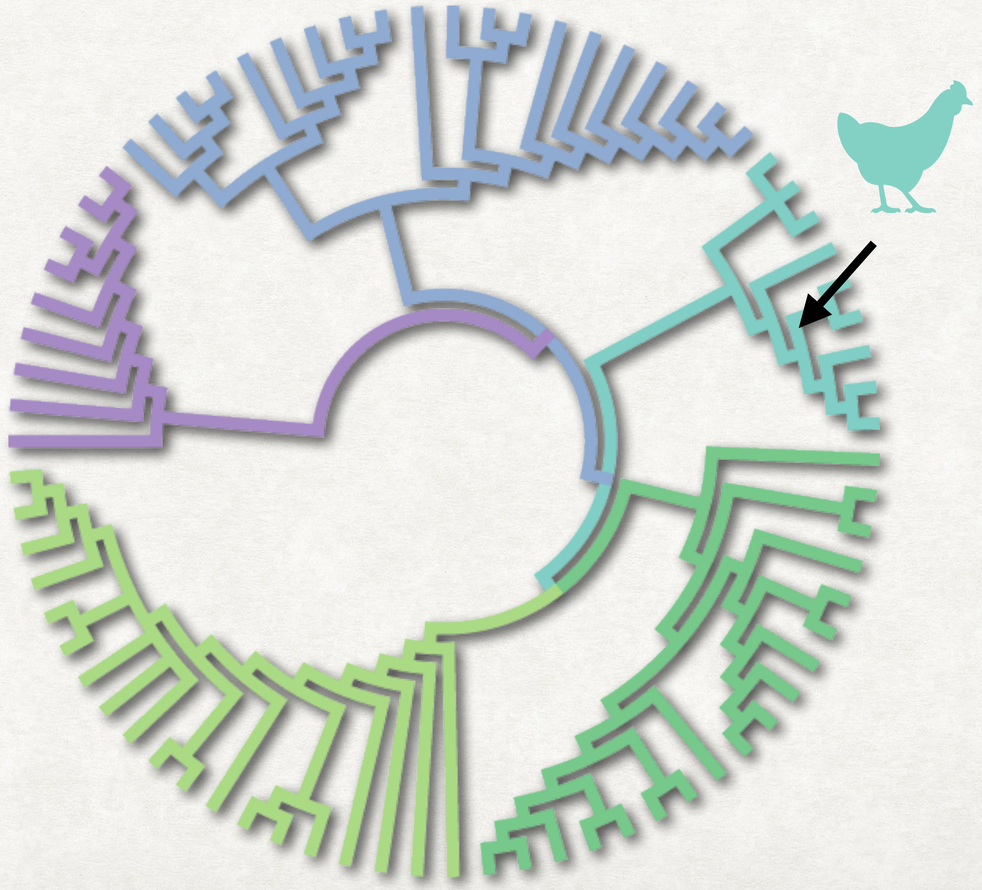
ARTICLE INFO

Article history:
Received 10 May 2016
Received in revised form 28 June 2016
Accepted 30 June 2016
Available online 4 July 2016

Keywords:
Nomenclature
Bacterial taxonomy
Classification
Updates

ABSTRACT

A key aspect of medical, public health, and diagnostic microbiology laboratories is the accurate and rapid reporting and communication regarding infectious agents of clinical significance. Microbial taxonomy in the age of molecular diagnostics and phylogenetics creates changes in taxonomy at a rapid rate further complicating this process. This update focuses on the description of new species and classification changes proposed in 2015.



ACTIVITY ALERT

What scientific questions can be explored with Trees?

ARTICLE

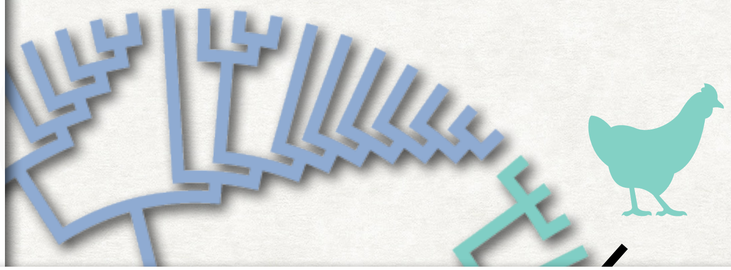
https://doi.org/10.1038/s41467-019-13443-4 OPEN

Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea

Qiyun Zhu[1,19], Uyen Mai[2,19], Wayne Pfeiffer[3], Stefan Janssen[1,4], Pedro Belda-Ferre[1], Gabriel A. Al-Ghalith[6], Evguenia Kopylova[1], D John B. Yin[8,9], Shi Huang[1,10], Nimaichand Salam[11], Jian-Yu Jiao[11], Zijun Yimeng Yang[6], Erfan Sayyari[8], Maryam Rabiee[2], James T. Morton Wen-Jun Li[11], Curtis Huttenhower[14,15], Nicola Segata[5], Larry Rob Knight[1,2,16,18*]

Available online 4 July 2016

Keywords:
Nomenclature
Bacterial taxonomy
Classification
Updates

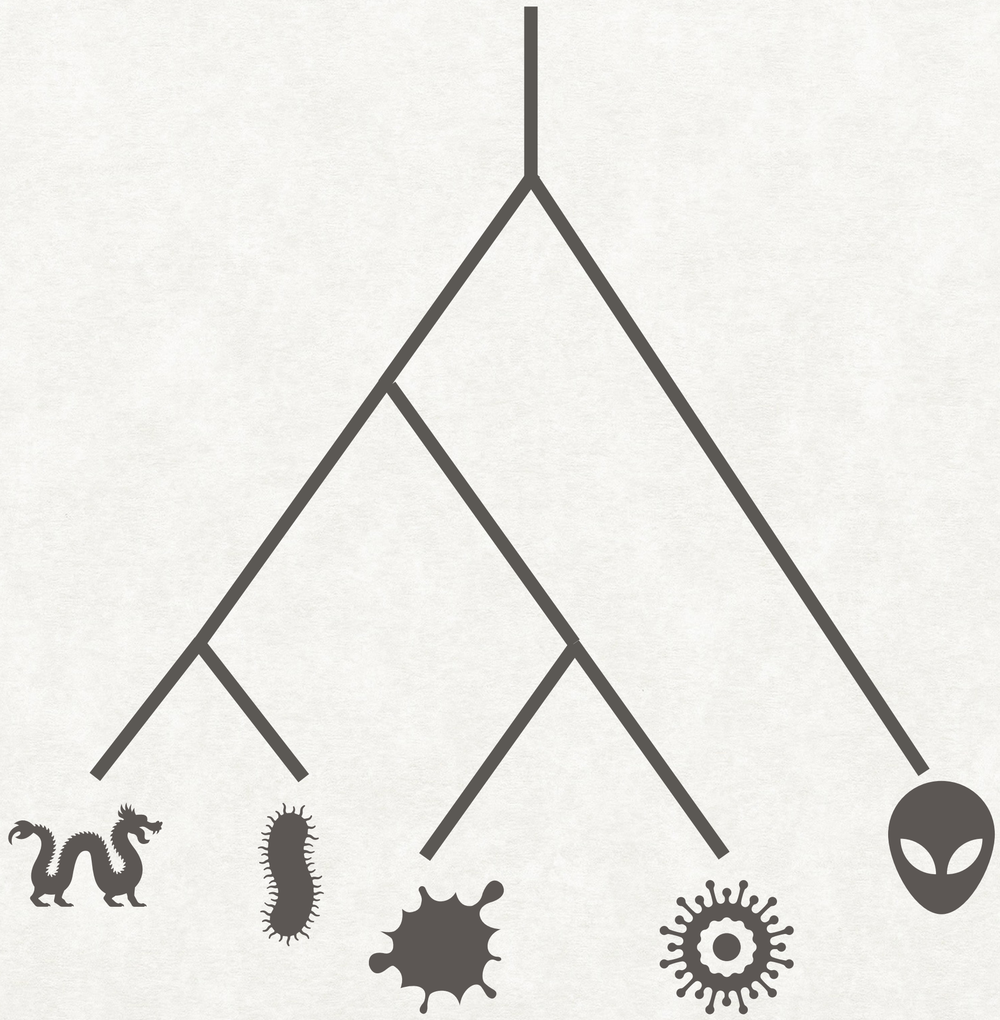


Insights into the phylogeny and coding potential of microbial dark matter

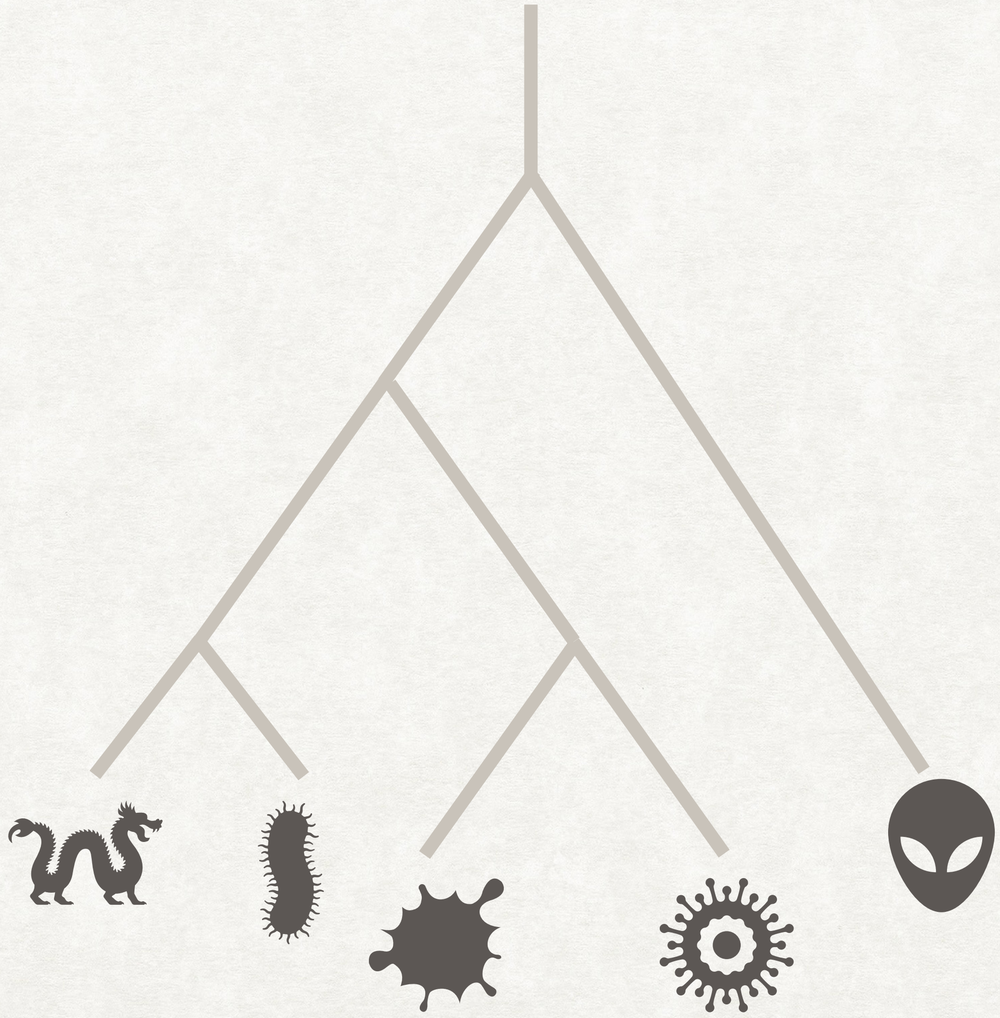
Authors: Christian Rinke, Patrick Schwientek, Alexander Sczyrba, Natalia N. Ivanova, Iain J. Anderson and Jan-Fang Cheng
Date: July 25, 2013
From: Nature(Vol. 499, Issue 7459)
Publisher: Nature Publishing Group
Document Type: Report
Length: 5,379 words
DOI: http://dx.doi.org/10.1038/nature12352

Abstract:

WHAT EVEN IS A PHYLOGENETIC TREE?



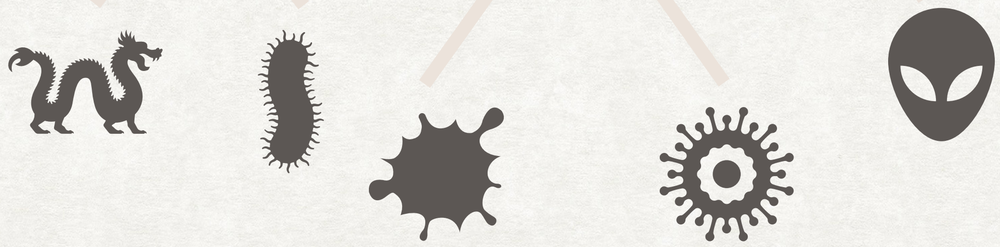
WHAT EVEN IS A PHYLOGENETIC TREE?



WHAT EVEN IS A PHYLOGENETIC TREE?

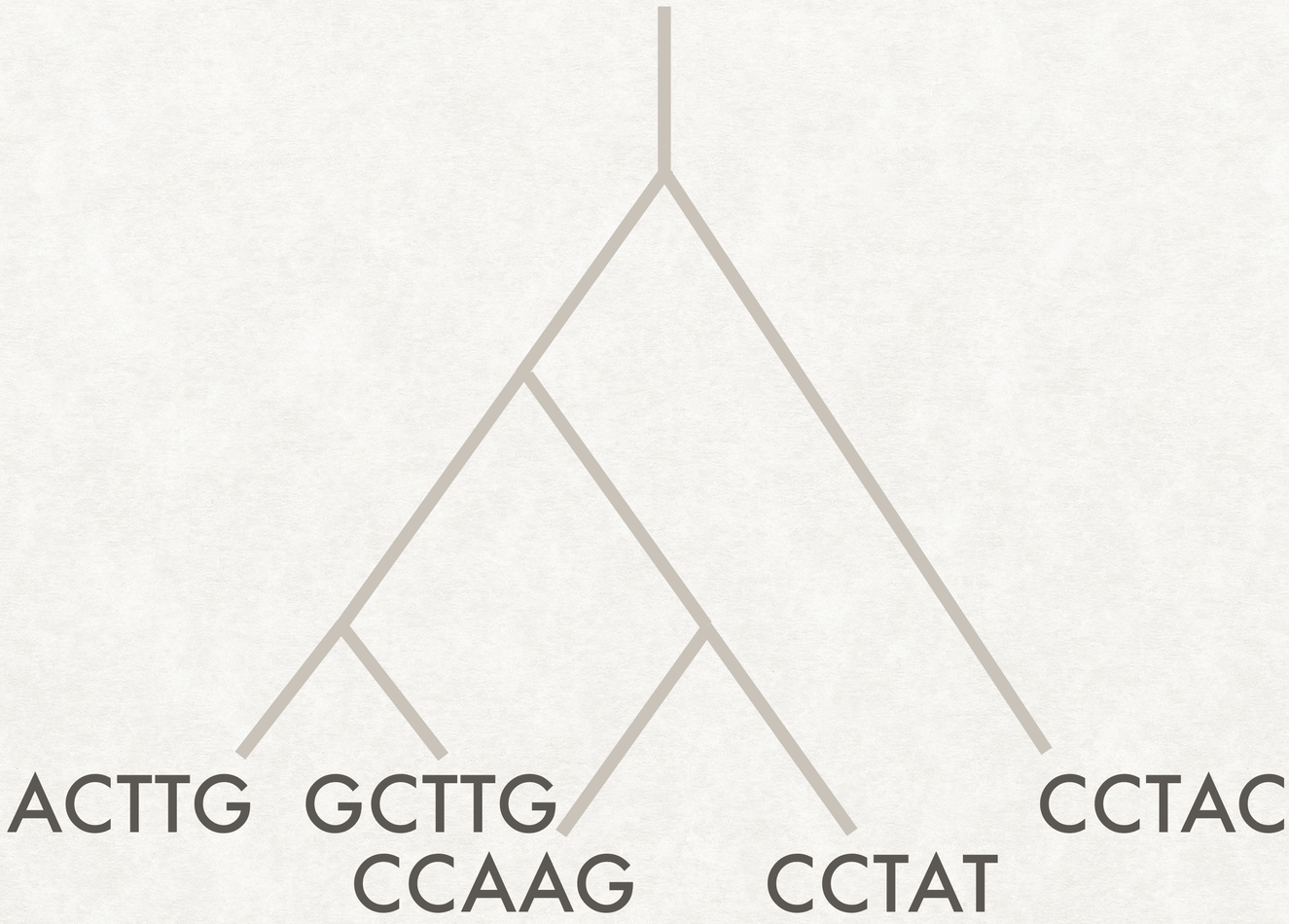
AN ESTIMATE!!

BUT OF WHAT EXACTLY?*...

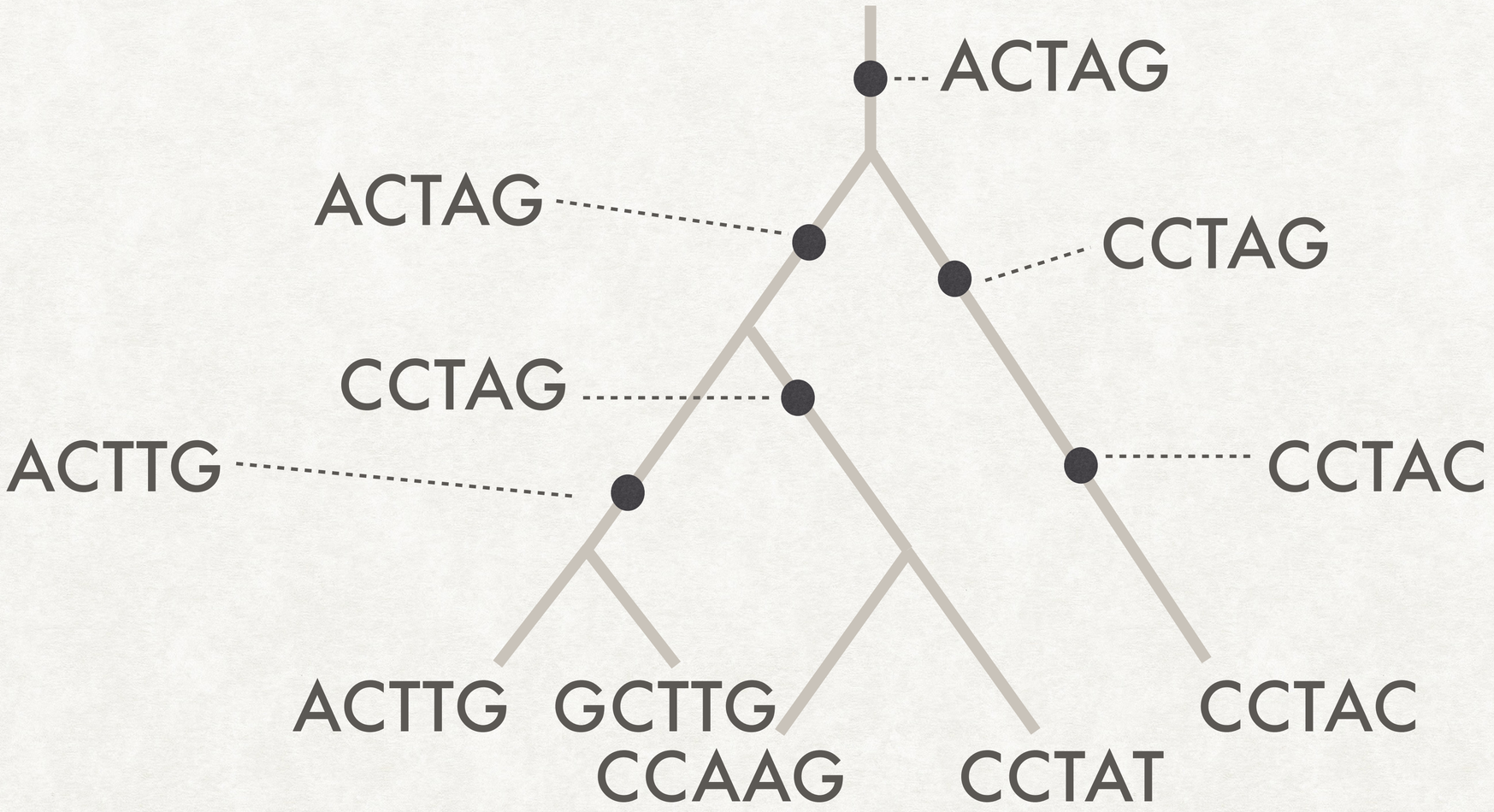


*Not a one correct answer, and very complex when talking about microorganisms

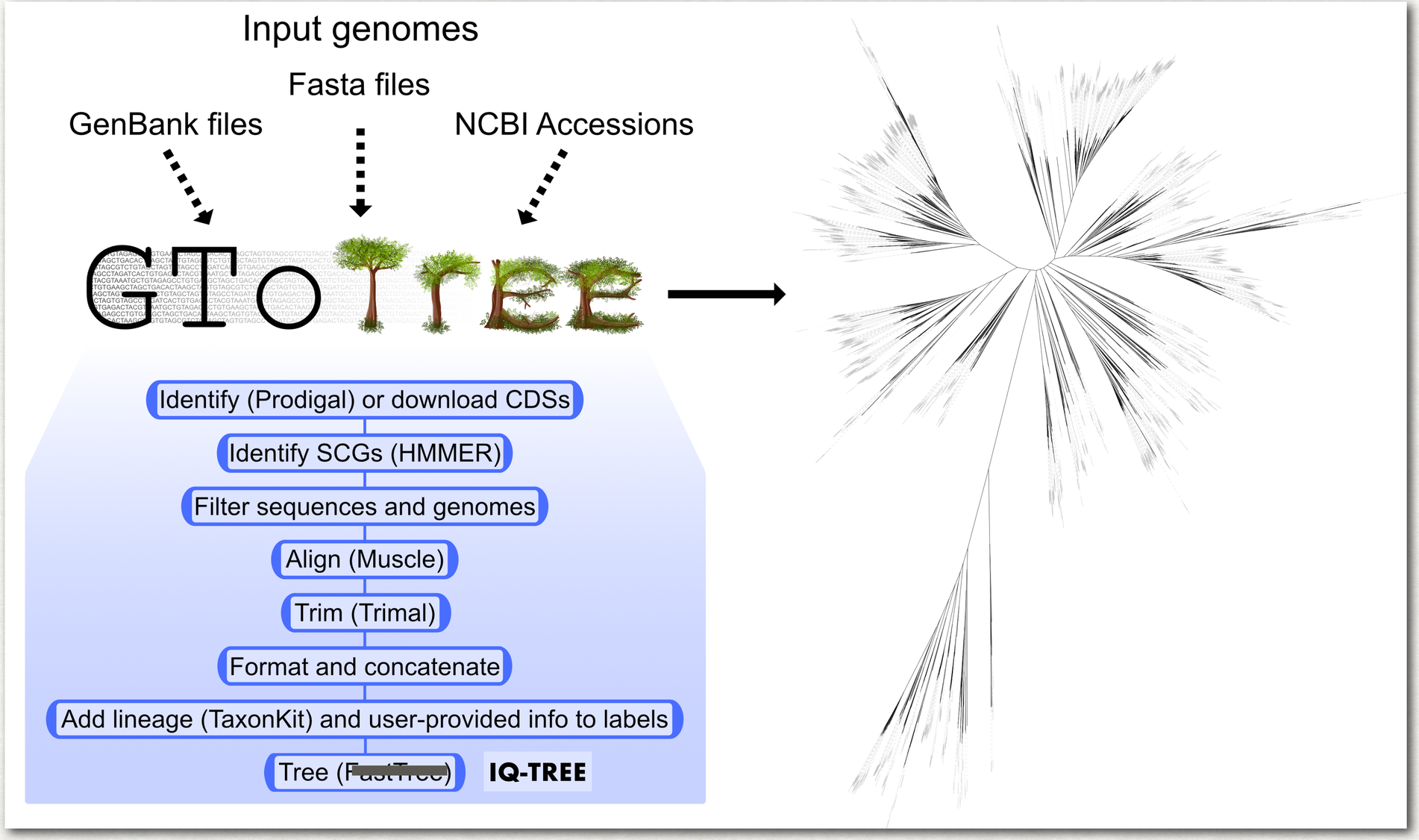
TREE ESTIMATION



TREE ESTIMATION

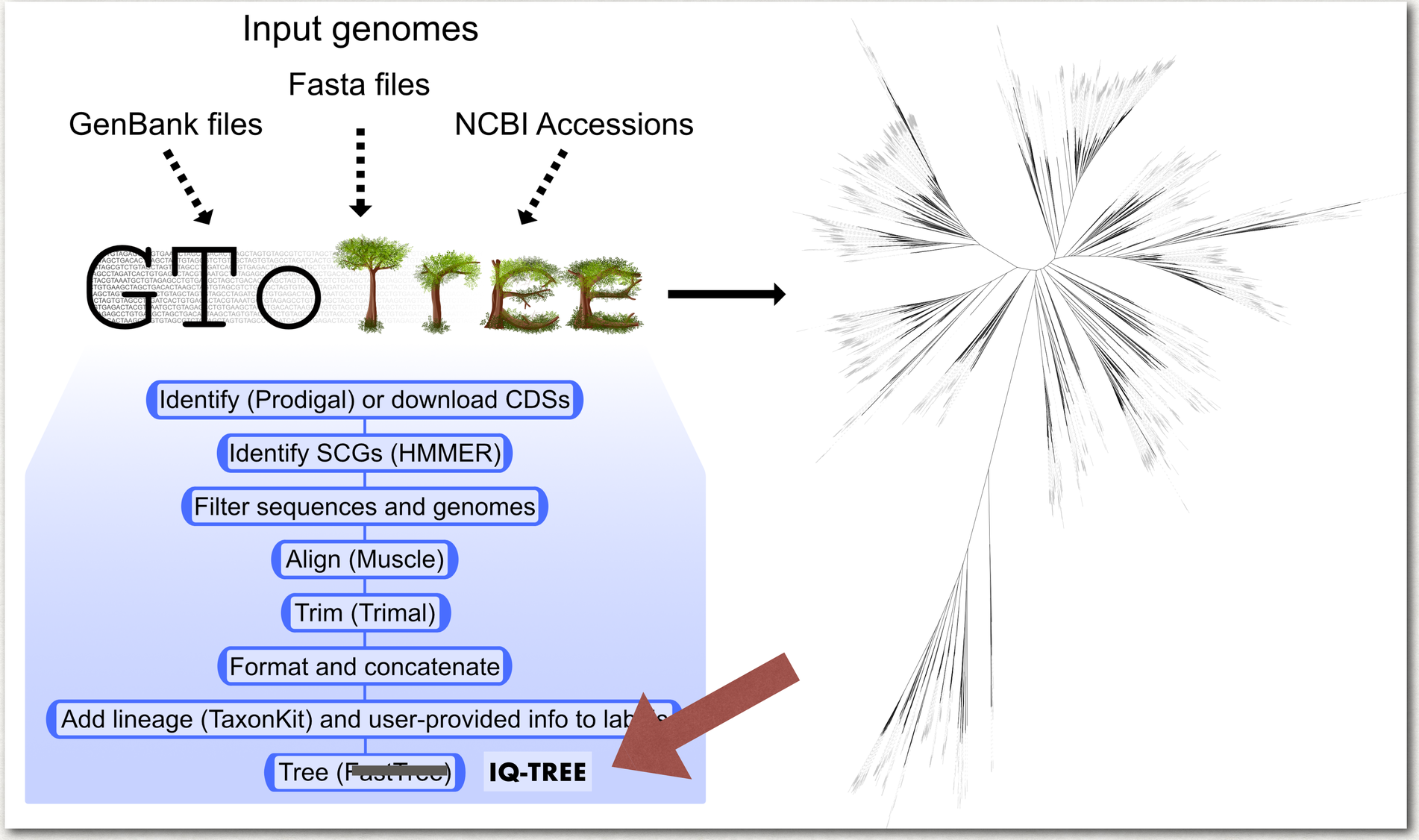


TREE ESTIMATION



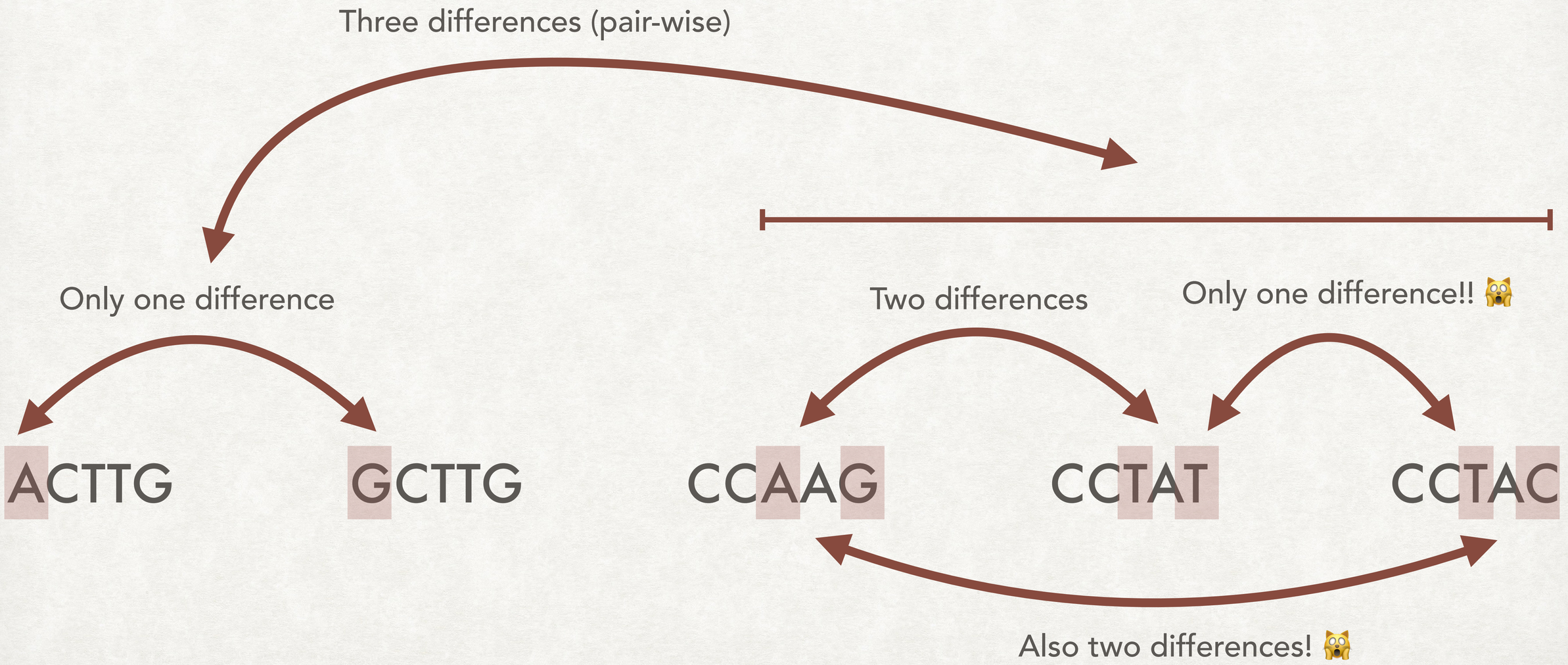
Access Lab instructions here

TREE ESTIMATION

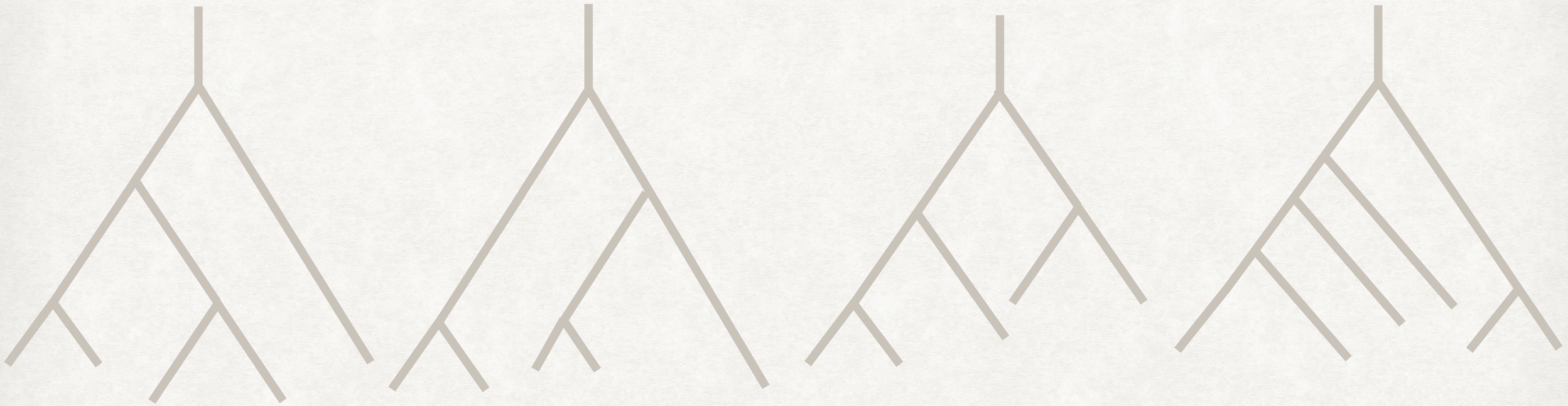


TREE ESTIMATION

Three differences (pair-wise)



TREE ESTIMATION

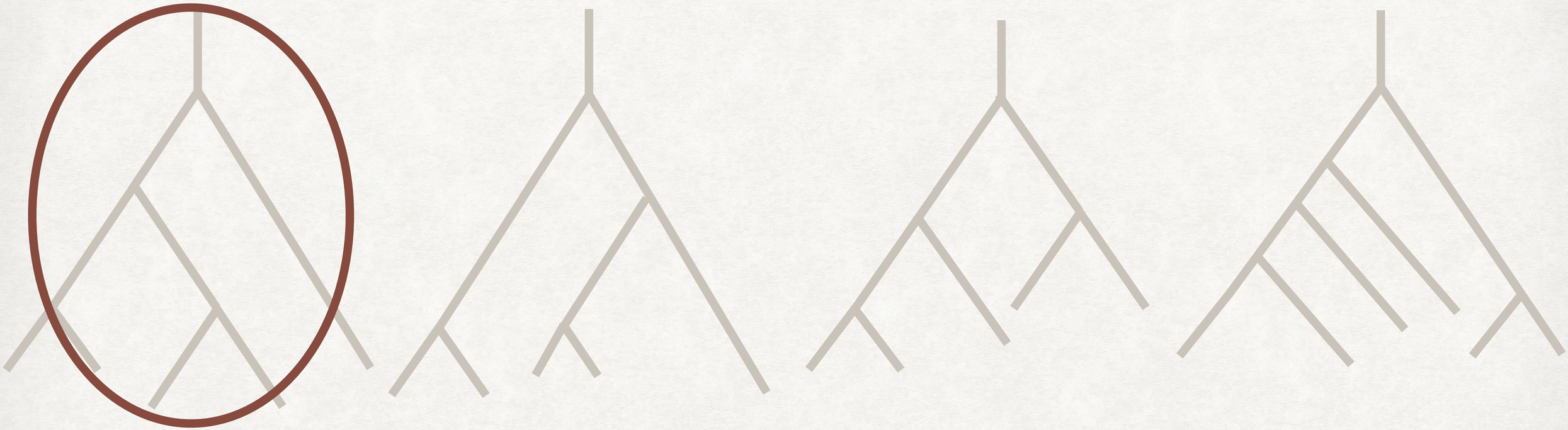


ACTTG GCTTG CCAAG CCTAT CCTAC

TREE ESTIMATION

Which of these trees is more “likely" to have produced the genomes?

P(🧬,🧬,…,🧬| 🌳, Θ)



P = 0.004

P = 0.0023

P = 0.00051

P = 0.000084

ACTTG

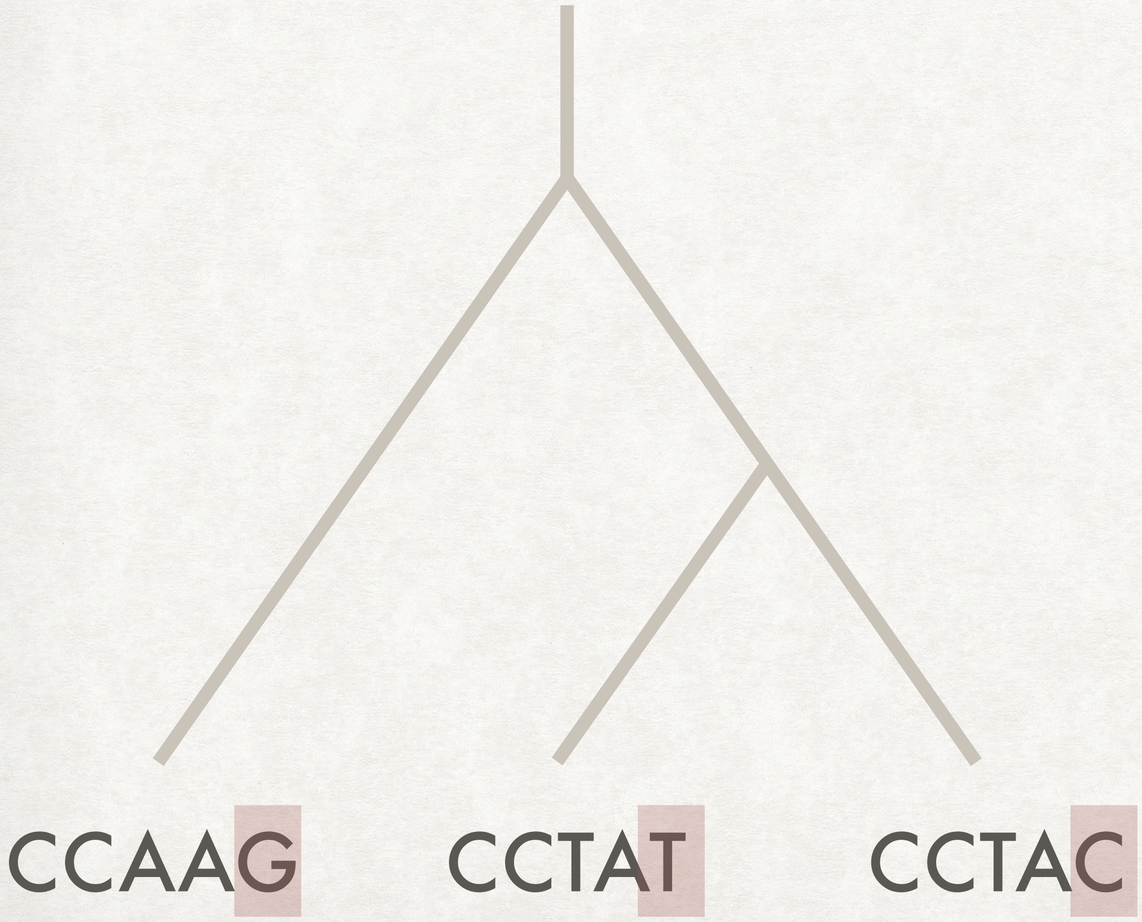
GCTTG

CCAAG

CCTAT

CCTAC

TREE ESTIMATION

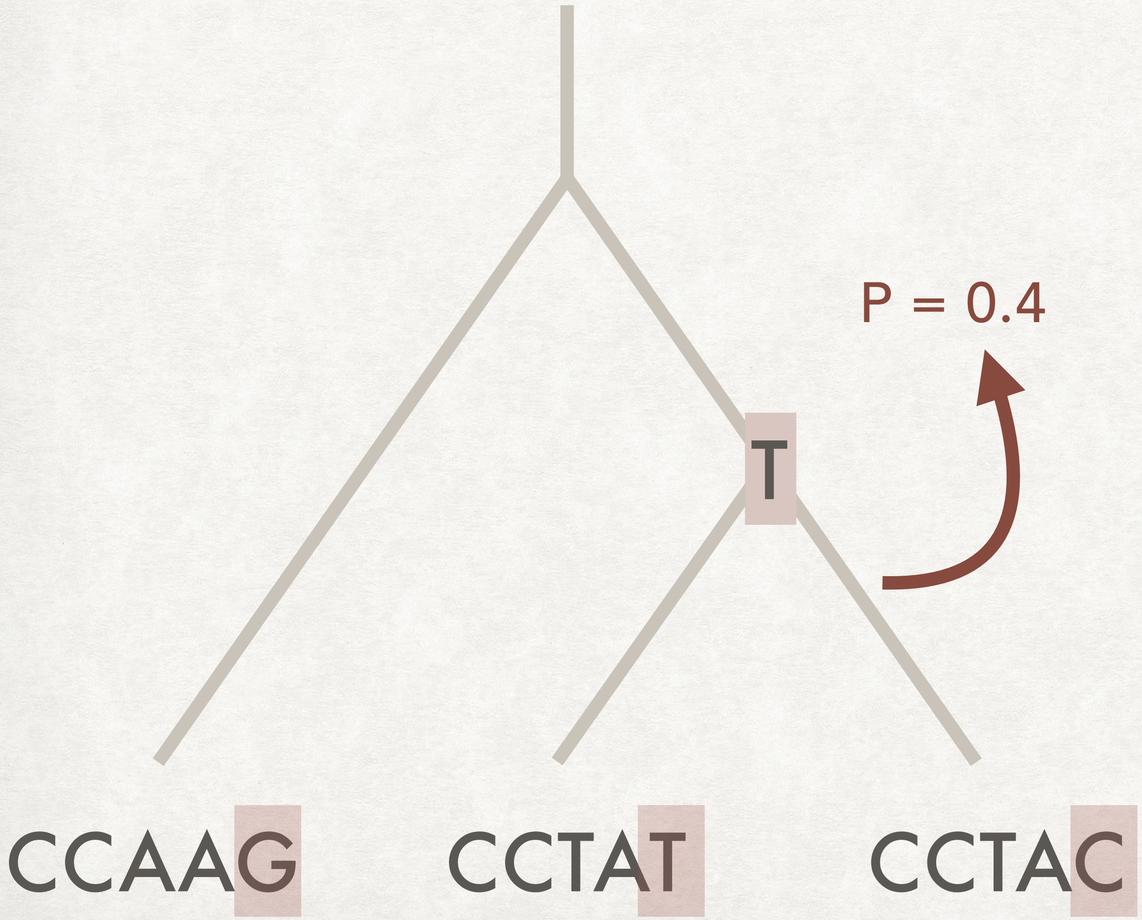


μ : Rate of change (μ x branch length = expected changes in one site)

r(A ←→ C), r(A ←→ G), …, r(G ←→ T): Proportion of change between different nucleotides

πA, πC, πG, πT: “Equilibrium" frequencies of each nucleotide.

TREE ESTIMATION



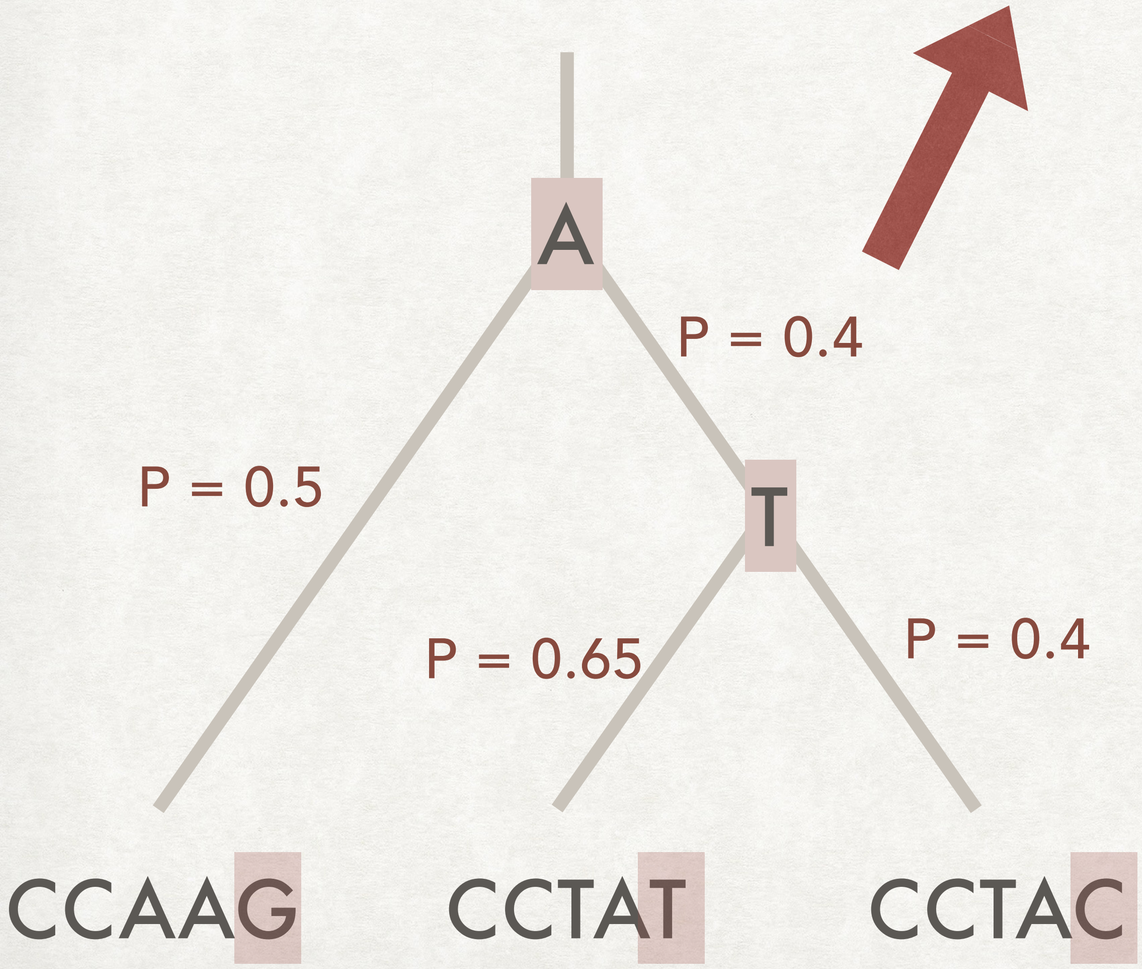
μ : Rate of change (μ x branch length = expected changes in one site)

r(A ←→ C), r(A ←→ G), …, r(G ←→ T): Proportion of change between different nucleotides

πA, πC, πG, πT: “Equilibrium" frequencies of each nucleotide.

TREE ESTIMATION

Total Prob = 0.4 x 0.4 x 0.65 x 0.5 = 0.052



Θ ☂

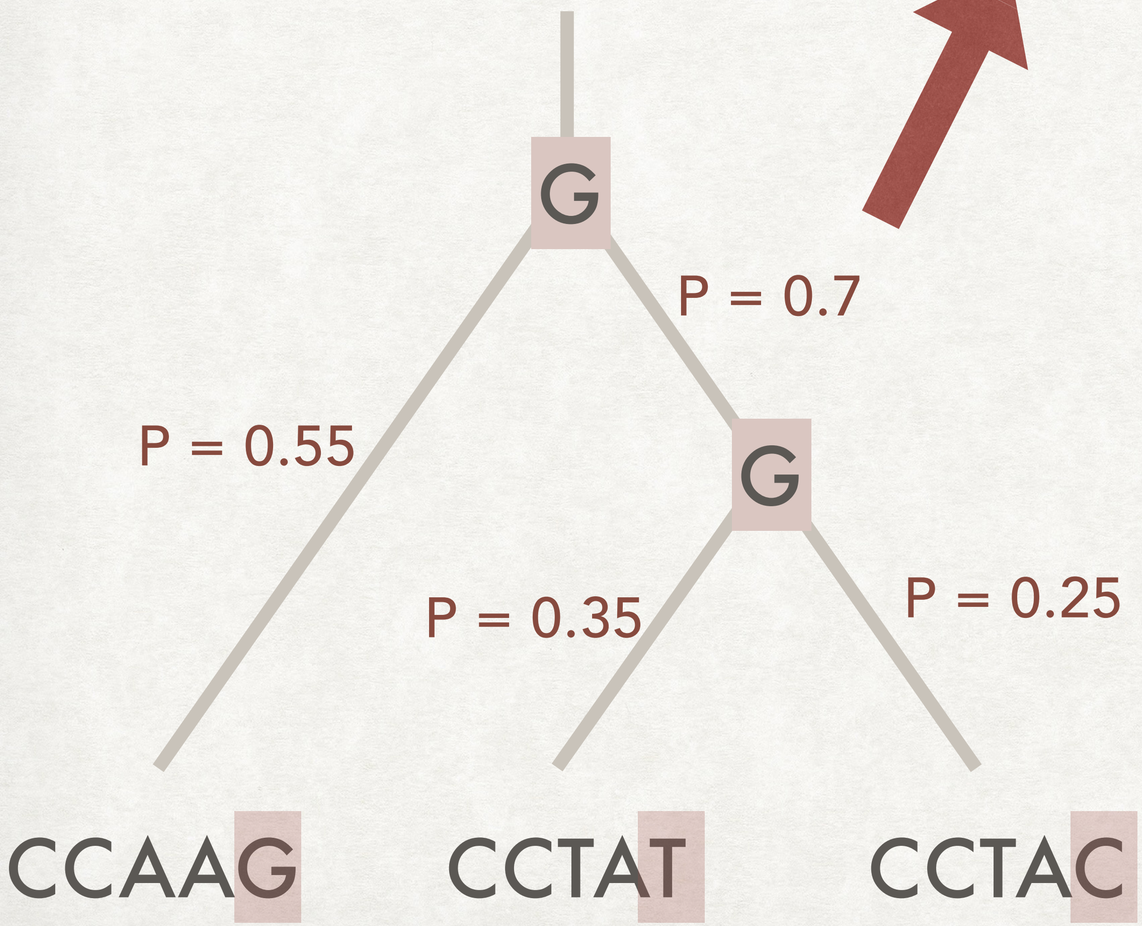
μ : Rate of change (μ x branch length = expected changes in one site)

r(A ←→ C), r(A ←→ G), …, r(G ←→ T): Proportion of change between different nucleotides

πA, πC, πG, πT: “Equilibrium" frequencies of each nucleotide.

TREE ESTIMATION

Total Prob = 0.7 x 0.25 x 0.35 x 0.55 = 0.0337…



Θ ☂

μ : Rate of change (μ x branch length = expected changes in one site)

r(A ←→ C), r(A ←→ G), …, r(G ←→ T): Proportion of change between different nucleotides

πA, πC, πG, πT: “Equilibrium" frequencies of each nucleotide.

ASSUMPTIONS MADE BY THE MODEL

SO… SO… MANY
BUT LET'S LIST SOME

- Each site (nucleotide) changes independently (questionable?)
- The proportion of transitions from A to T is the same as from T to A (fair as far as I understand)
- The rate of change/frequencies/proportions of transitions are the same across sites (questionable again) and branches of the tree (super questionable).
- When applied to aligned genes across the genome, **ALL SHARE THE SAME EVOLUTIONARY HISTORY**

CONCLUSIONS

- Estimation of phylogenetic trees: truly complex problem to attack.
- The current methods used are strict **parametric** models, which we know oversimplify reality a lot.
- When talking about phylogenomics for microbes, maybe we should talk about phylogenetics instead.

P.D. There are methods not discussed here that take gene trees and return an estimate for the species trees (ASTRAL, BEAST,…). But it is still not entirely clear to me what we mean by “species tree” when talking on this scale of life.