

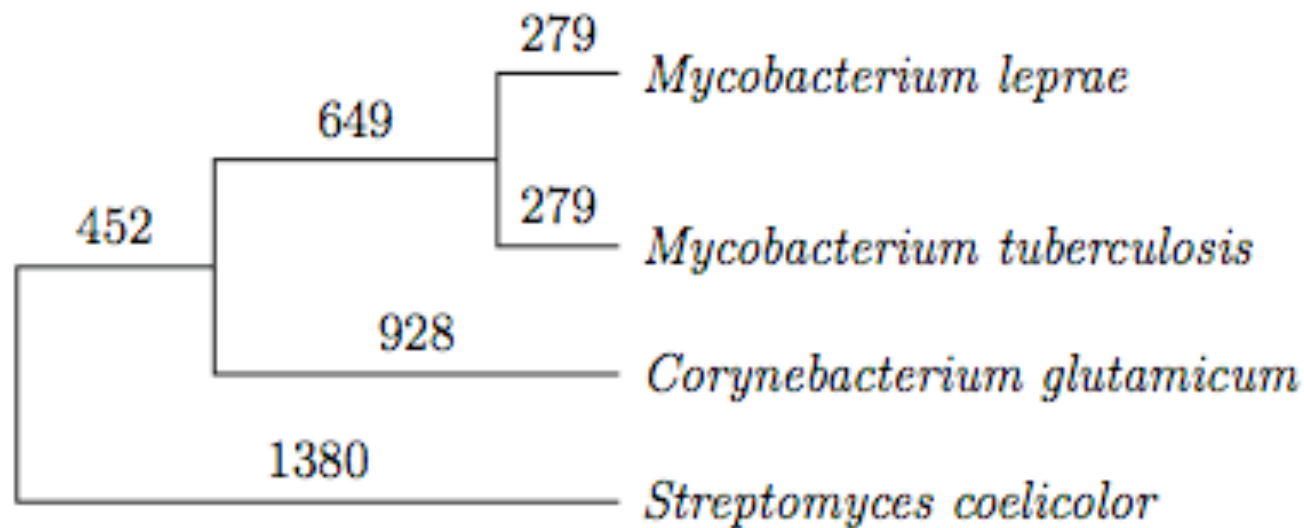
# **A Genomic Approach to Two Related Human Scourges**

John Ferguson  
Scott H. Harrison  
Mary Spratt

*“Bioinformatics” Notes and  
Materials*

- These materials are extracts from:  
Harrison 2006. Open Reading Frame Composition and Organization as Indicators of Phenotypic Diversity in *Bacteria* and *Archaea*. Michigan State University Dissertation.

# What genomes to get data from?



Times of divergence for 4 Actinobacteria. Branch length units are in millions of

[Data adapted from Battistuzzi et al 2004]

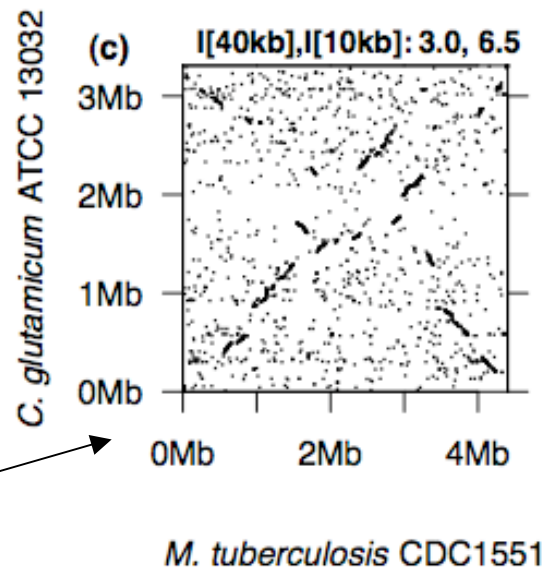
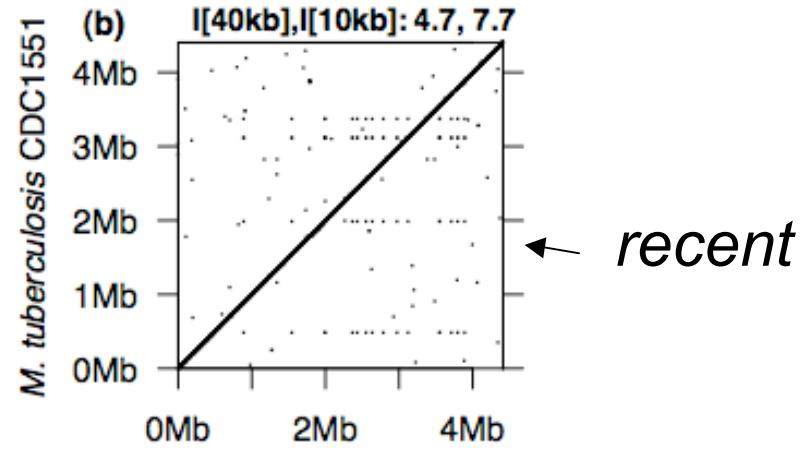
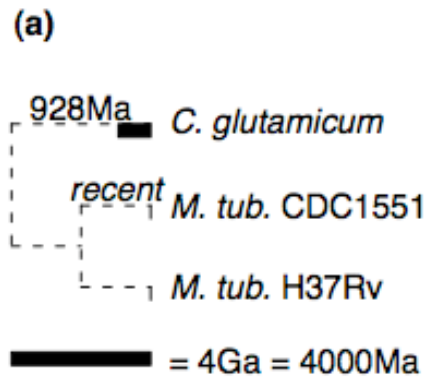
## Laboratory Exercise Module #1:

# How do the genomes differ at a more macroscopic level?

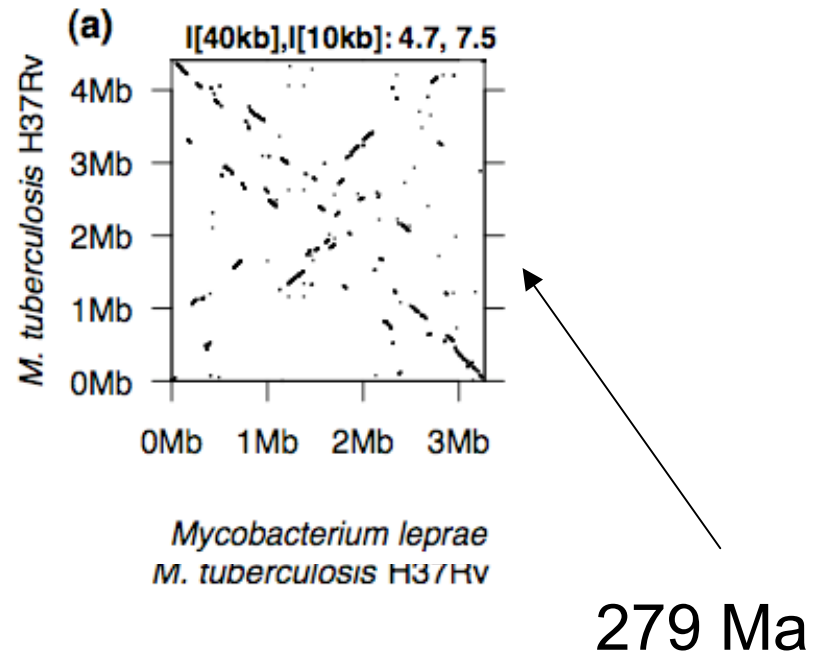
- Go to NCBI web site and find the sizes of the two genomes.
- Go to the NCBI and IMG web sites and find the number of ORFs of the two genomes
- What is the density of ORFs on these genomes?

This ratio of total ORF content to chromosome Size varied from 49.5%(Mycobacterium leprae) to 96.8% (Pirellula sp. 1)

The lowest density is **one ORF for every 2,036 bp (M. leprae)**, and the highest density is one ORF for every 853 bp (Pyrobaculum aerophilum str. IM2).



928 Ma



## **Laboratory Exercise Module #2: Investigating genes and pathways across four Actinobacteria genomes**

- IMG and pathway analysis (KEGG)
- Choose ORF(s) from our “conserved” menu list, other menu list, and choose your own ORF from *M. leprae* or *M. tuberculosis*
- Gather homologs from the four genomes, *M. leprae*, *M. tuberculosis*, *C. glutamicum*, and *S. coelicolor*

# **Laboratory Exercise Module #2:**

## **Investigating genes and pathways across four Actinobacteria genomes**

- Based on BLAST scores for your ORF from the “conserved” menu list as found across the four Actinobacteria genomes (best hits):
  - Which two genomes are most closely related?

# Laboratory Exercise Module #2: Investigating genes and pathways across four Actinobacteria genomes

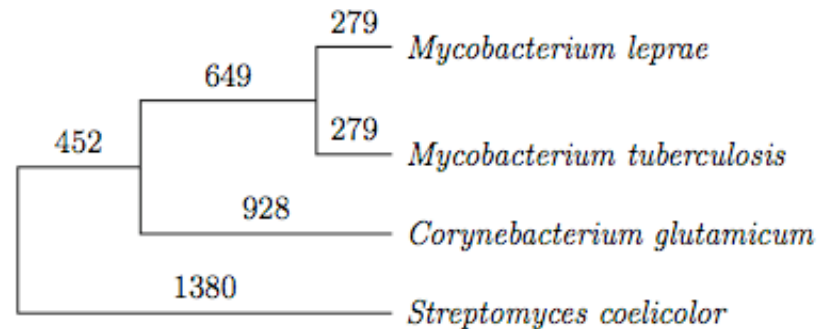


Figure 12: Times of divergence for 4 Actinobacteria. Branch length units are in millions of years (Ma).

- Go to IGM: Do all four genomes have high similarity matches for your ORF from the “other” menu list?
- If ORFs are only found on three of the genomes and not four, may this imply a loss of an ORF?
- Look at the following trees. If ORFs are present in only two genomes, does this mean loss of ORF in the other two genomes or gaining of ORF in the other two genomes?

# **Laboratory Exercise Module #2:**

## **Investigating genes and pathways across four Actinobacteria genomes**

- Is your ORF functional? Is it conserved across two or more genomes? Is its length conserved? If it is on only one genome, does it occur elsewhere in phylogeny (use IGM)?
- Codon usage? Pathway analysis?

- There is extensive gross-level conservation of genome structure compared to mosaic-like differences in smaller-scale structures. An instance of this phenomena can be observed with the 3 species of Mycobacterium : M. leprae, M. tuberculosis, and M. bovis (Philipp et al., 1998).

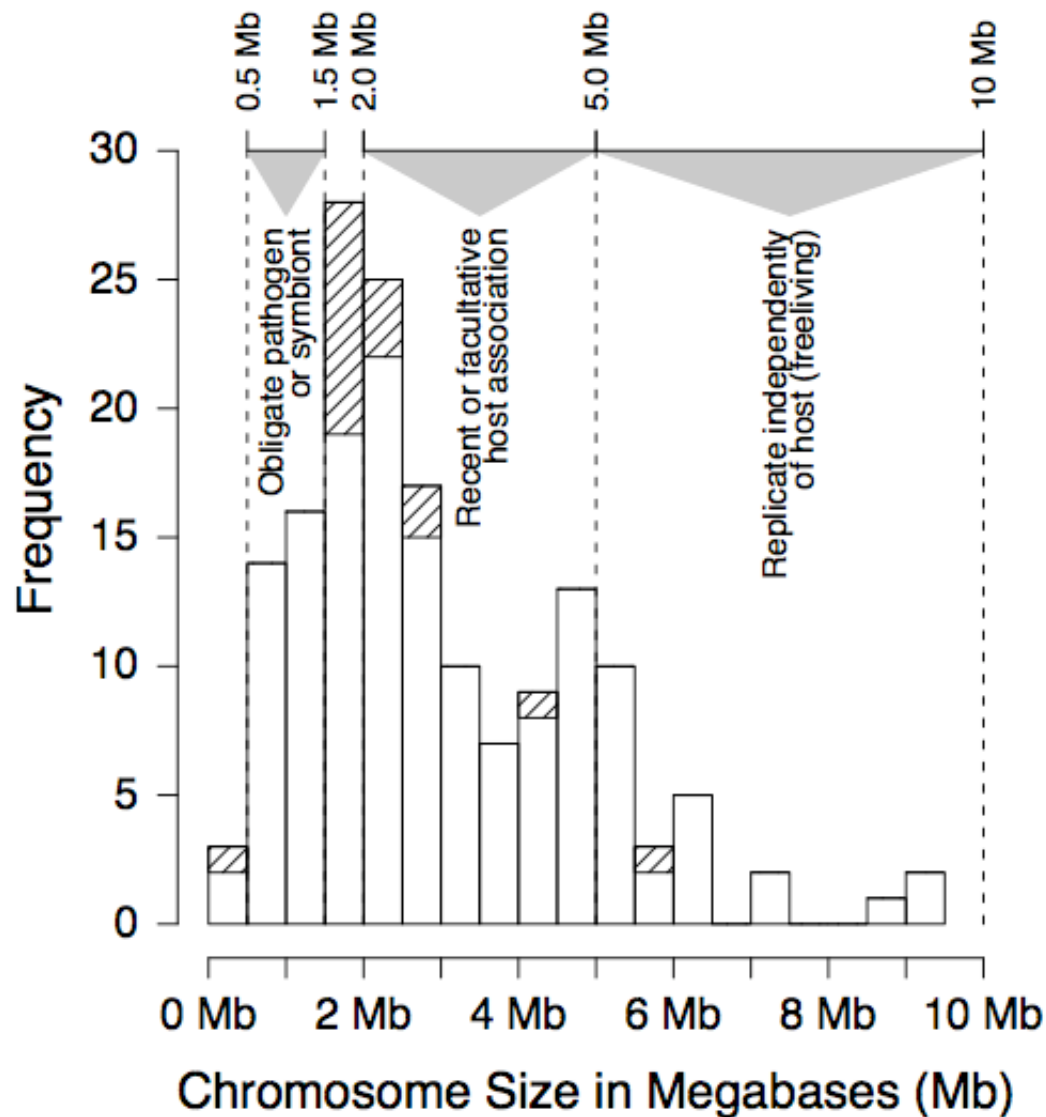


Figure 36: Histogram of 165 chromosome sizes. Bin size is 500,000 base pairs. Chromosome sizes are shown for 165 different chromosomes coming from 155 genome-sequenced strains. Frequencies of archaeal chromosome sizes are indicated by shaded boxes stacked above the frequency counts of bacterial chromosomes shown by unshaded boxes. The approximate boundaries for three lifestyle-based ranges of genome sizes are listed at the top of the figure and are indicated by vertically descending, dashed lines.

Related References:  
 Ochman & Davalos (2006) and Moran & Plague (2004)