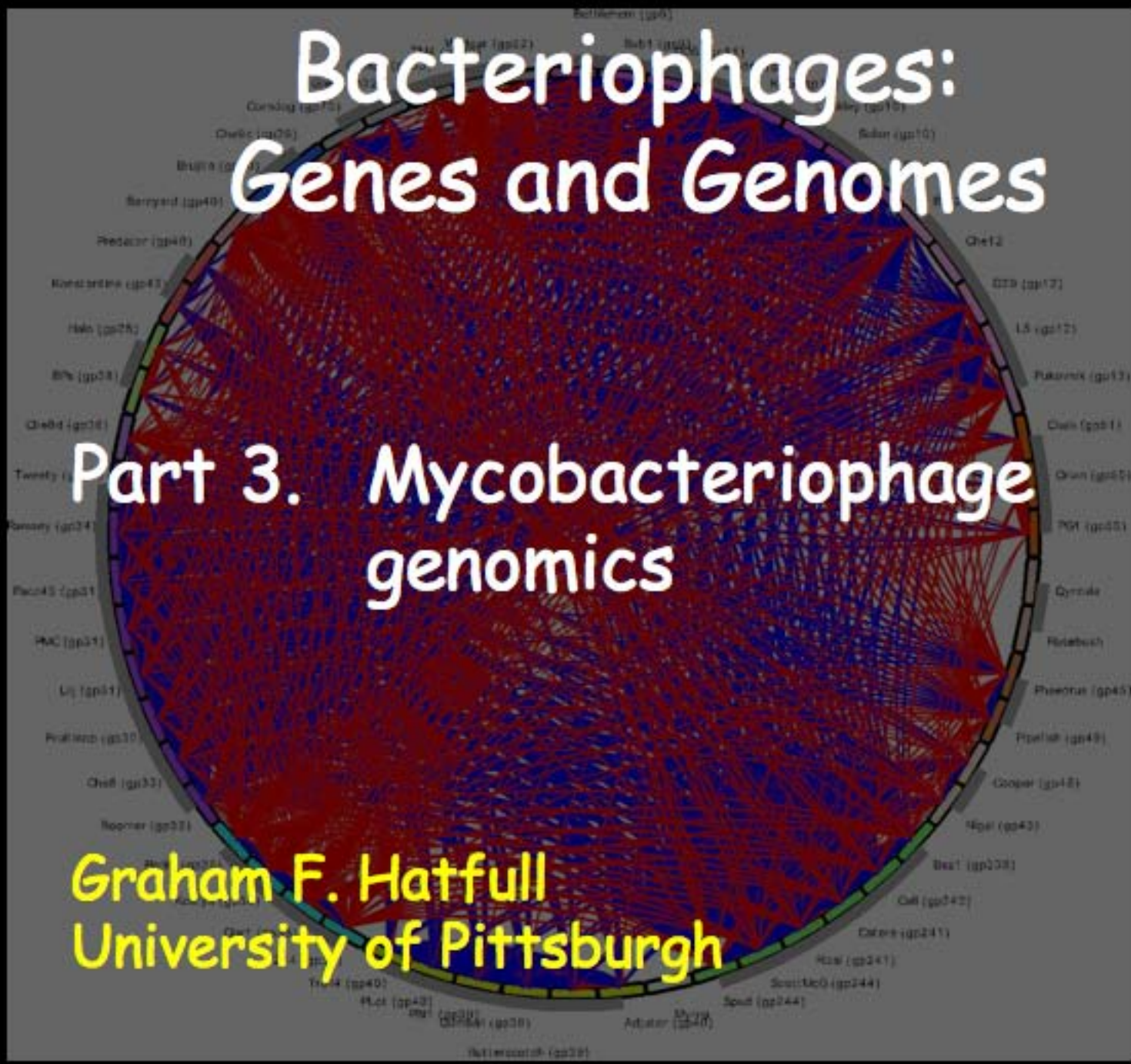# Bacteriophages: Genes and Genomes

Part 3.   Mycobacteriophage genomics

Graham F. Hatfull
University of Pittsburgh

Comparing genomes of a common host

Phages of a common host are expected to be in genetic communication and more closely related

Which hosts to choose?

Bacterial pathogens; environmentally-important bacteria, other bacteria of biological interest

Why mycobacteriophages?

• Mycobacteriophages are viruses that infect mycobacterial hosts

• Hosts include *Mycobacterium tuberculosis* and *Mycobacterium smegmatis.*

• *M. tuberculosis* kills more than people than any other single infectious agent

• About one-third of world's population is infected with *M. tuberculosis.* Mostly latent infections
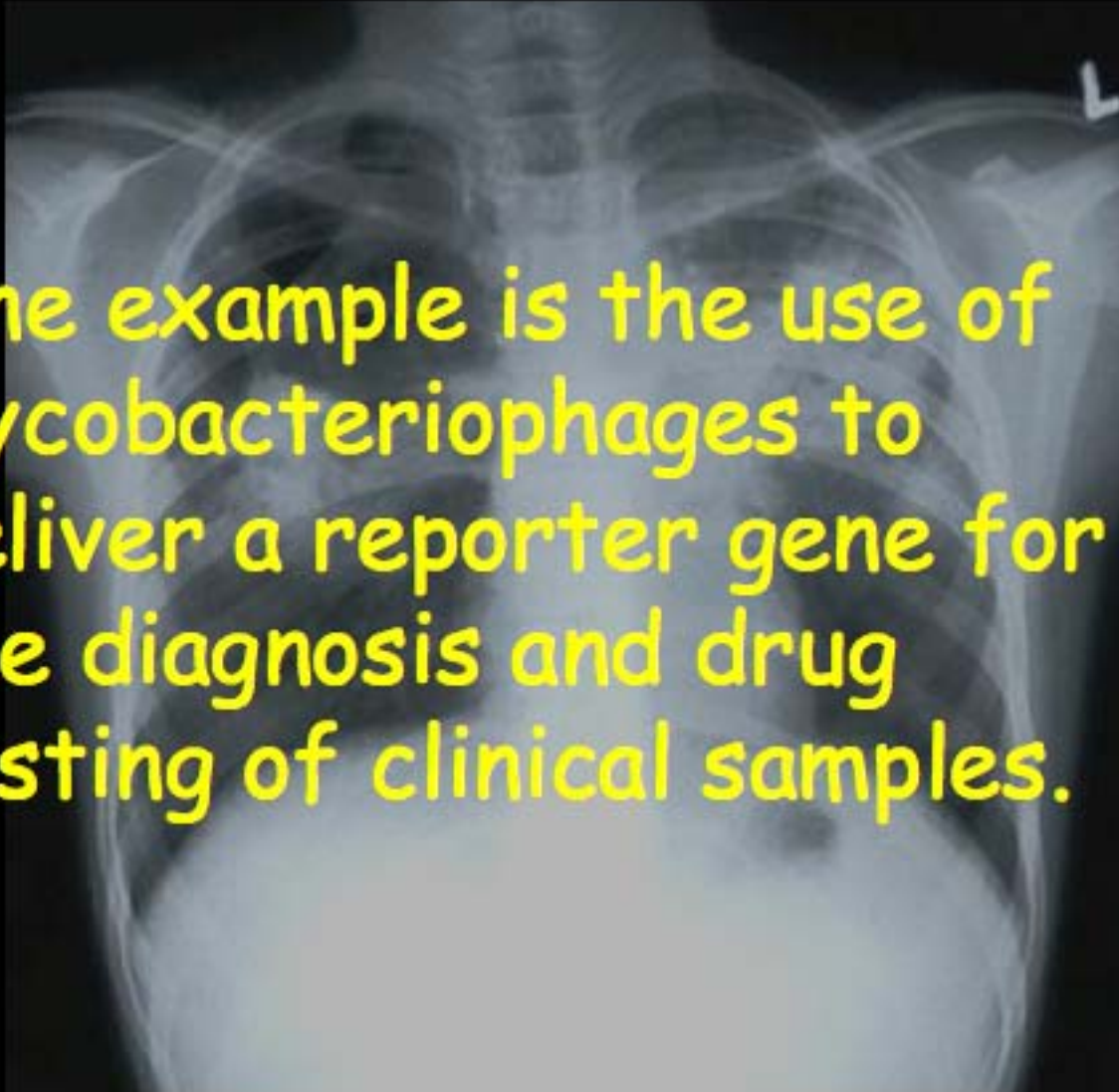
Why mycobacteriophages?

Mycobacteriophages cannot only teach us about viral diversity, but also provide tools for controlling tuberculosis.
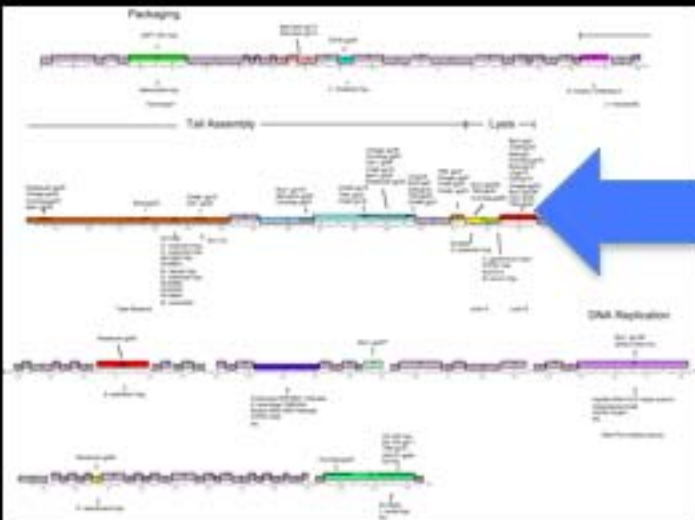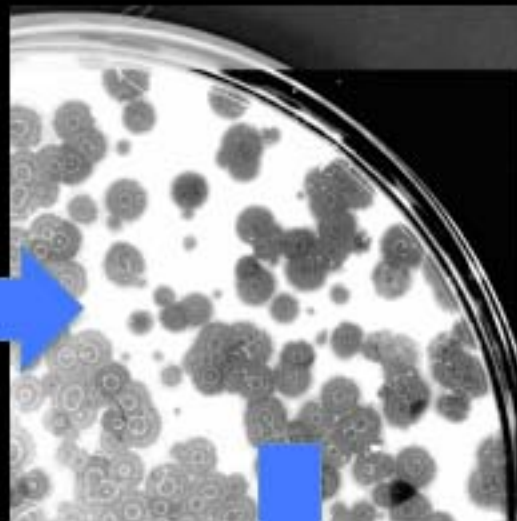
Why mycobacteriophages?

One example is the use of mycobacteriophages to deliver a reporter gene for the diagnosis and drug testing of clinical samples.

Discovering new mycobacteriophages

AGCTGCTGATCG
ATGCGCCGTCGAT
GCACGTCGATGCT
A

(x 10,000's)

Mycobacteriophage isolation

Host: *M. smegmatis*

Fast-growing; non-pathogen

Source: Soil, compost etc.
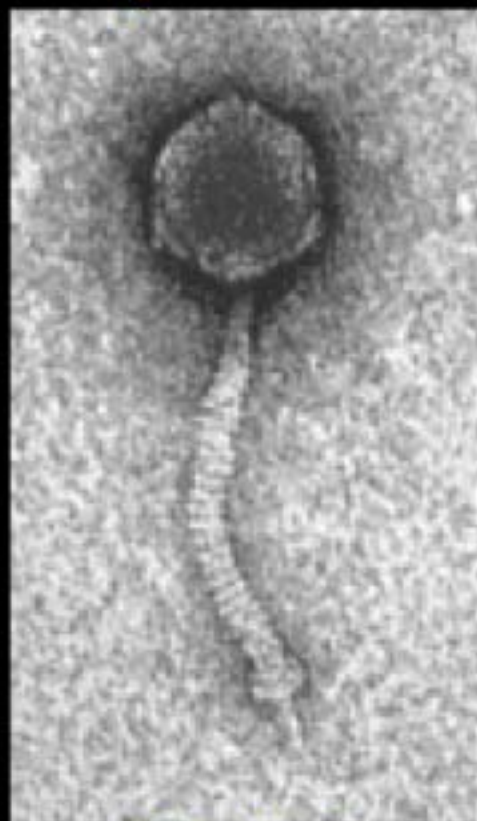
Direct plating without enrichment

Mycobacteriophage isolation

- **Thousands isolated**

- **154 genome sequences (60 published)**
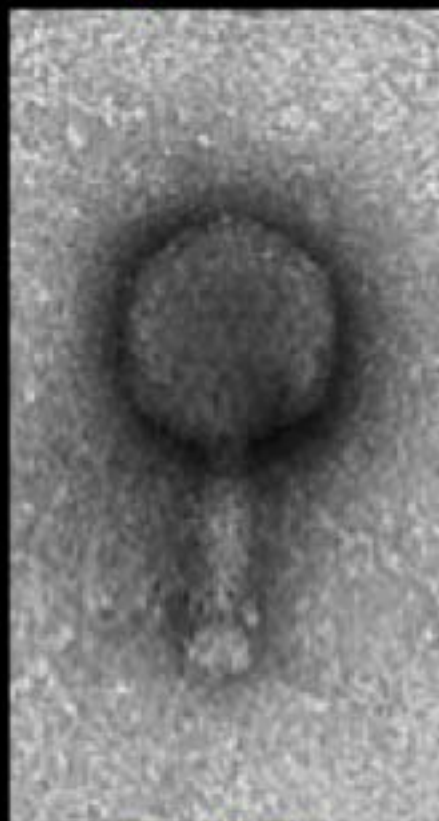
- **All dsDNA tailed phages**

All are either Siphoviridae or Myoviridae (no Podoviridae)
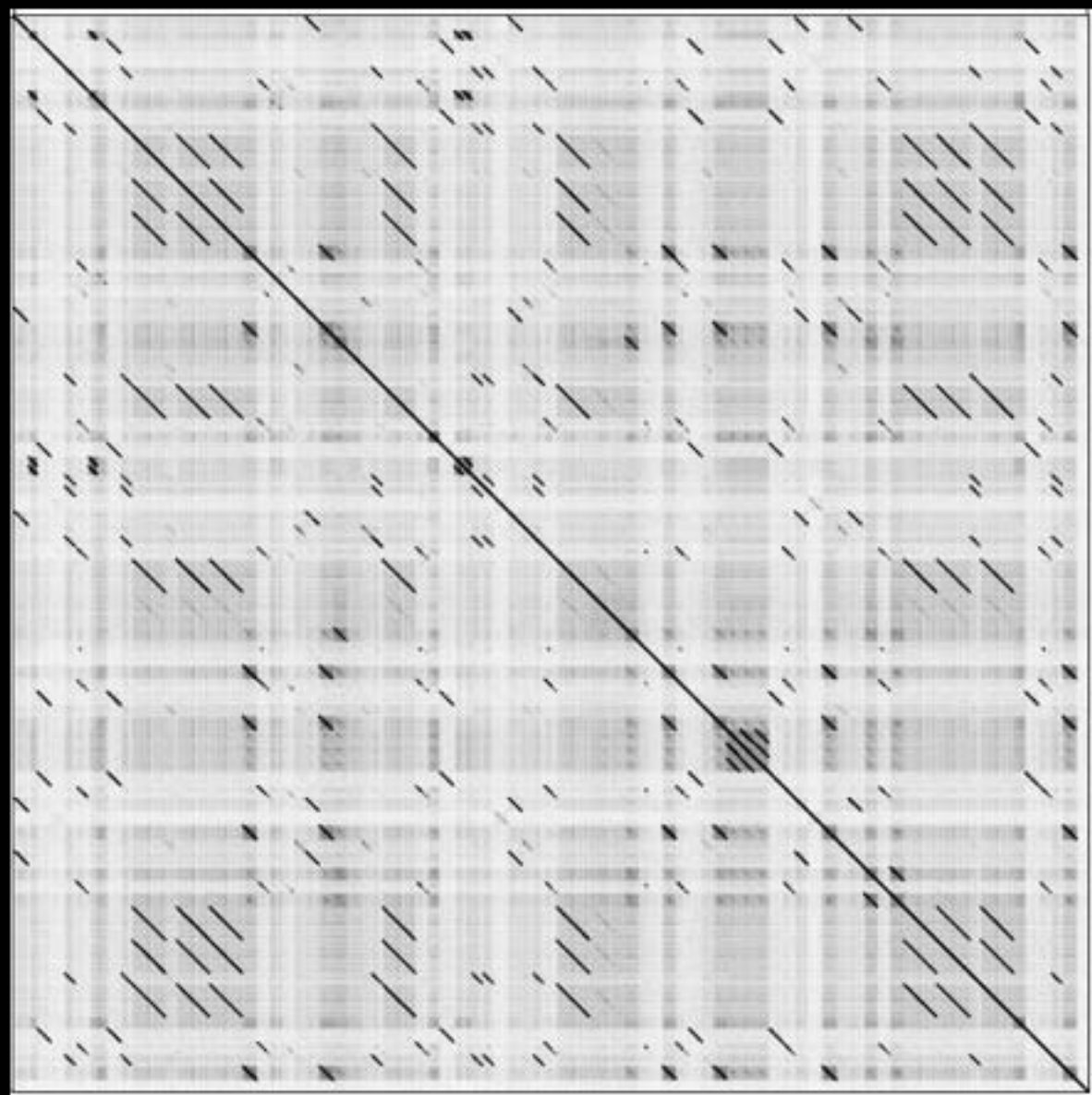

Siphovirus
53/60

Myovirus
7/60

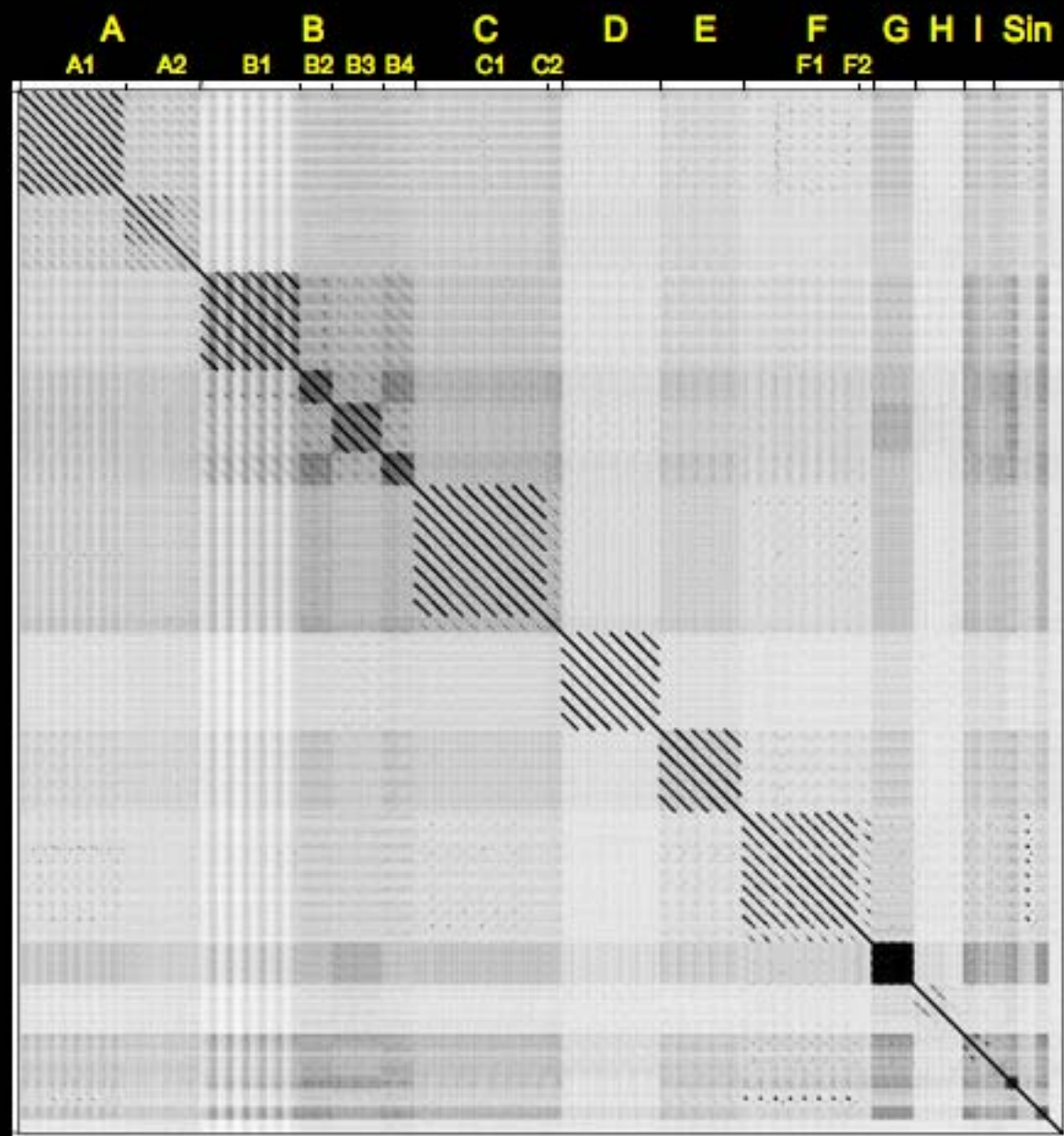# What do their genomes look like?

For 60 genomes:

- Average length: 72,588bp
- Range: 41,901bp – 164,602 bp
- Average GC%: 63.4%
- GC% range: 56.3 – 69%
- Total orf's: 6,858
- Average orf's: 114
- Average orf size: 616 bp

Genome Clustering: Diversity and heterogeneity

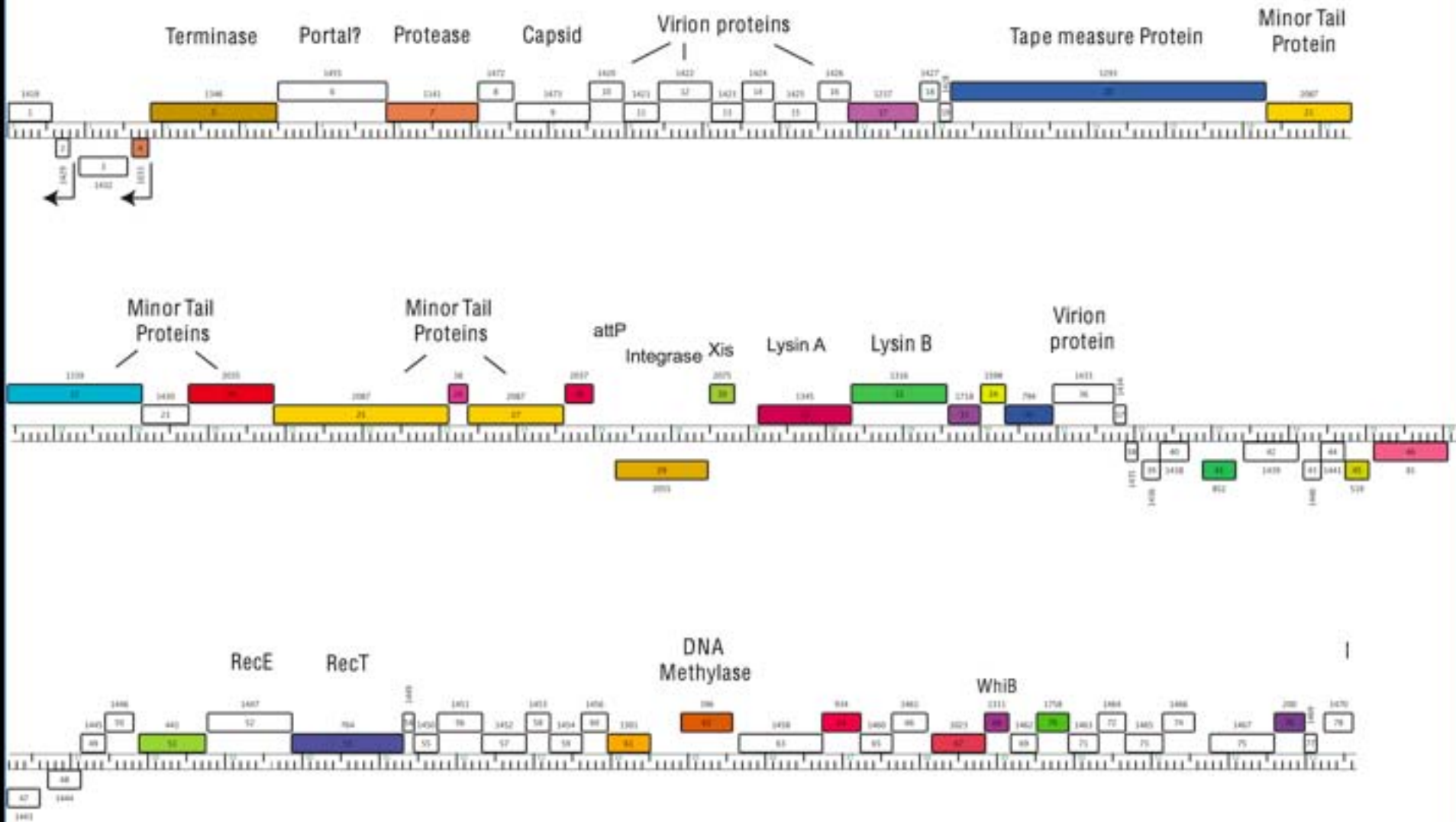# Genome Clustering: Diversity and heterogeneity

Clusters and Subclusters

A	B	C	D	E
A1	**B1**	**C1**		
Bethlehem	Chah	Bxz1	Adjutor	244
Bxb1	Orion	Cali	Butterscotch	Cjw1
Jasper	PG1	Catera	Gumball	Kostya
KBG	**B2**	Rizal	PBI1	Porky
Lockley	Qyrzula	ScottMcG	PLot	
Solon	Rosebush	Spud	Troll4	
U2	**B3**	**C2**		
A2	Phaedrus	Myrna		
Che12	Pipefish			
D29	**B4**			
L5	Cooper			
Pukovnik	Nigel			
Bxz2				

Clusters and Subclusters (continued)

F	G	H	I	Sin
F1		**H1**		
Boomer	BPs	Predator	Che9c	Corndog
PMC	Halo	Konstantine	Brujita	Giles
Llij		**H2**		TM4
Che8		Barnyard		Wildcat
Tweety				Omega
Fruitloop				
Ramsey				
Pacc40				
F2				
Che9d				

Giles: Example of a genome map

Genome general features

- Genes are tightly packed: few non-coding regions

- Many genes, few operons

- Virion genes often grouped

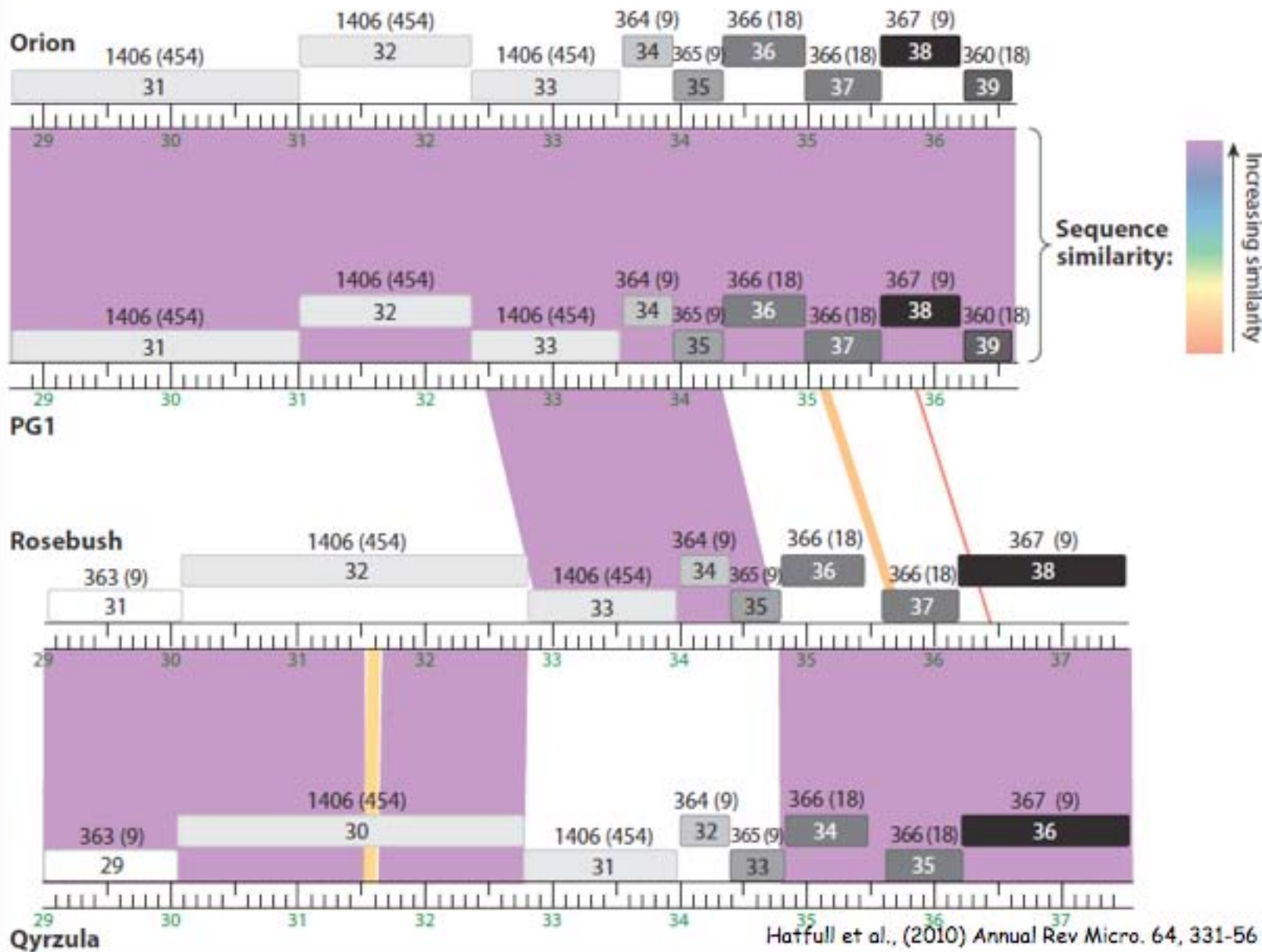- Many genes of unknown function

Sorting genes into phamilies

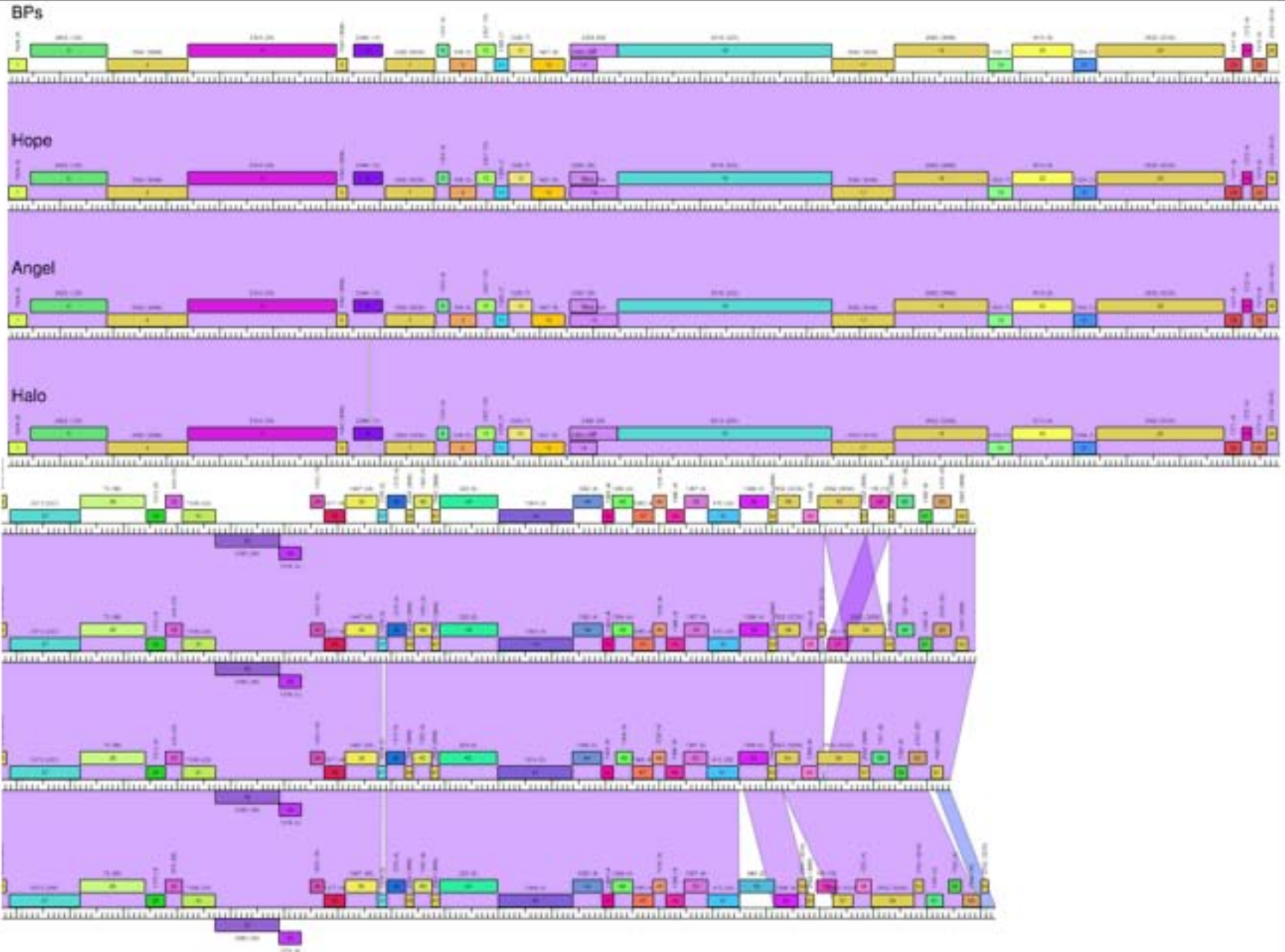Comparing protein sequences reveals distant evolutionary relationships

- Total orfs: 6,858

- Phamilies: 1523
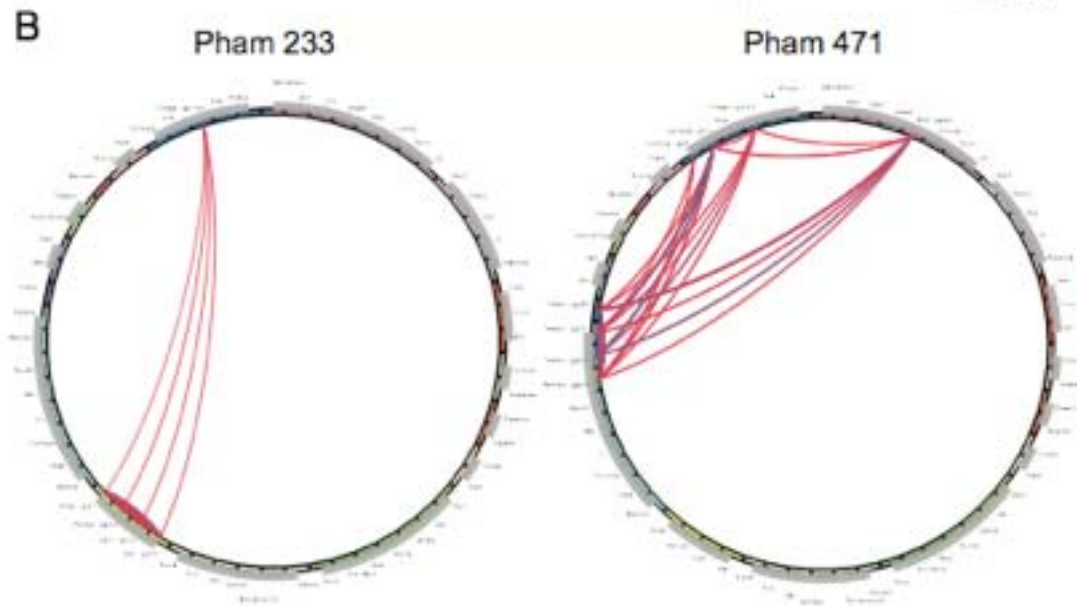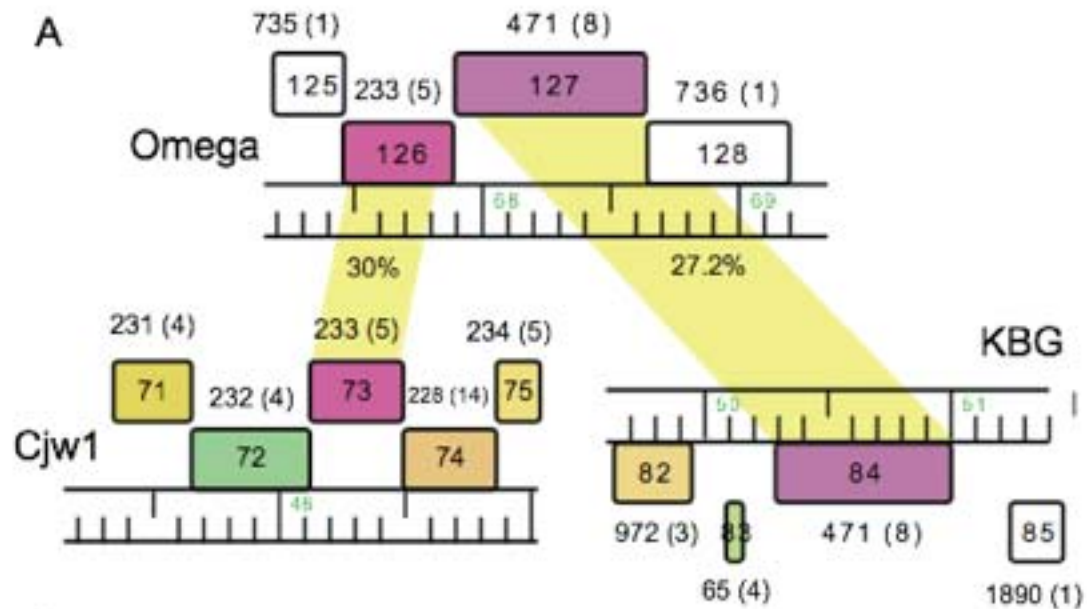
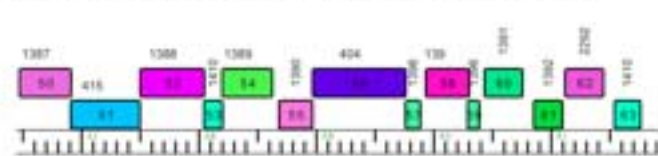- Orphams: 773 (44.8%)

Comparing genomes

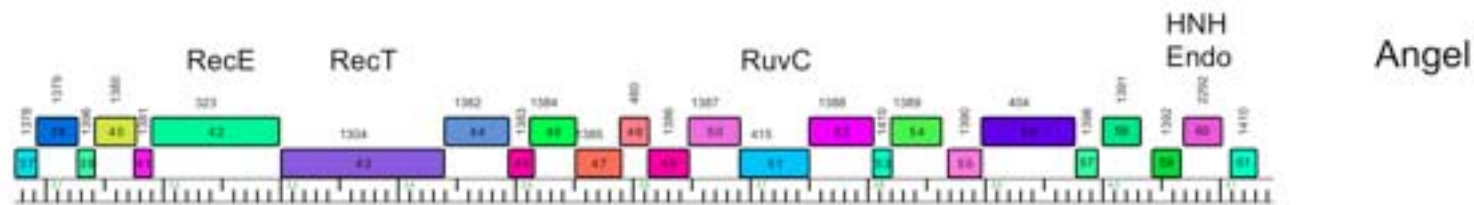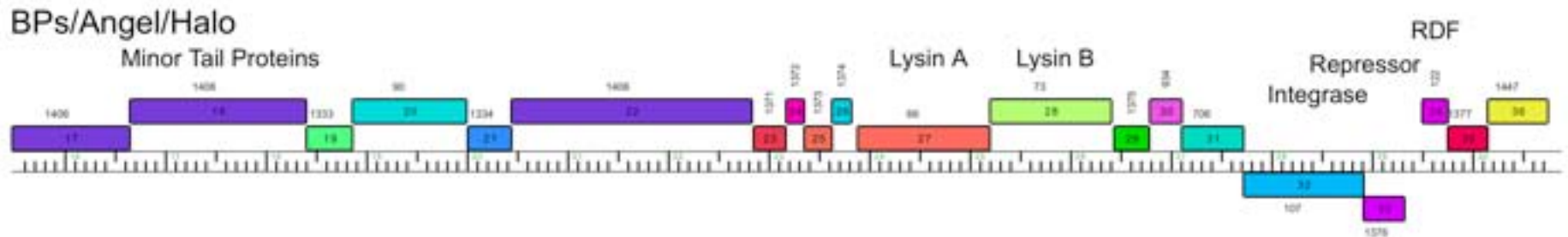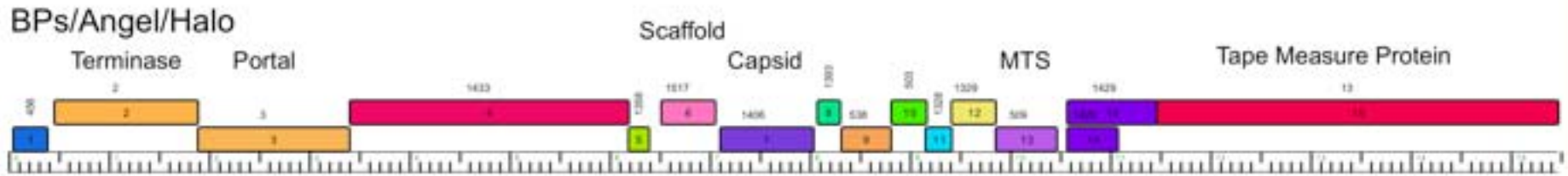Hatfull et al., (2010) Annual Rev Micro. 64, 331-56
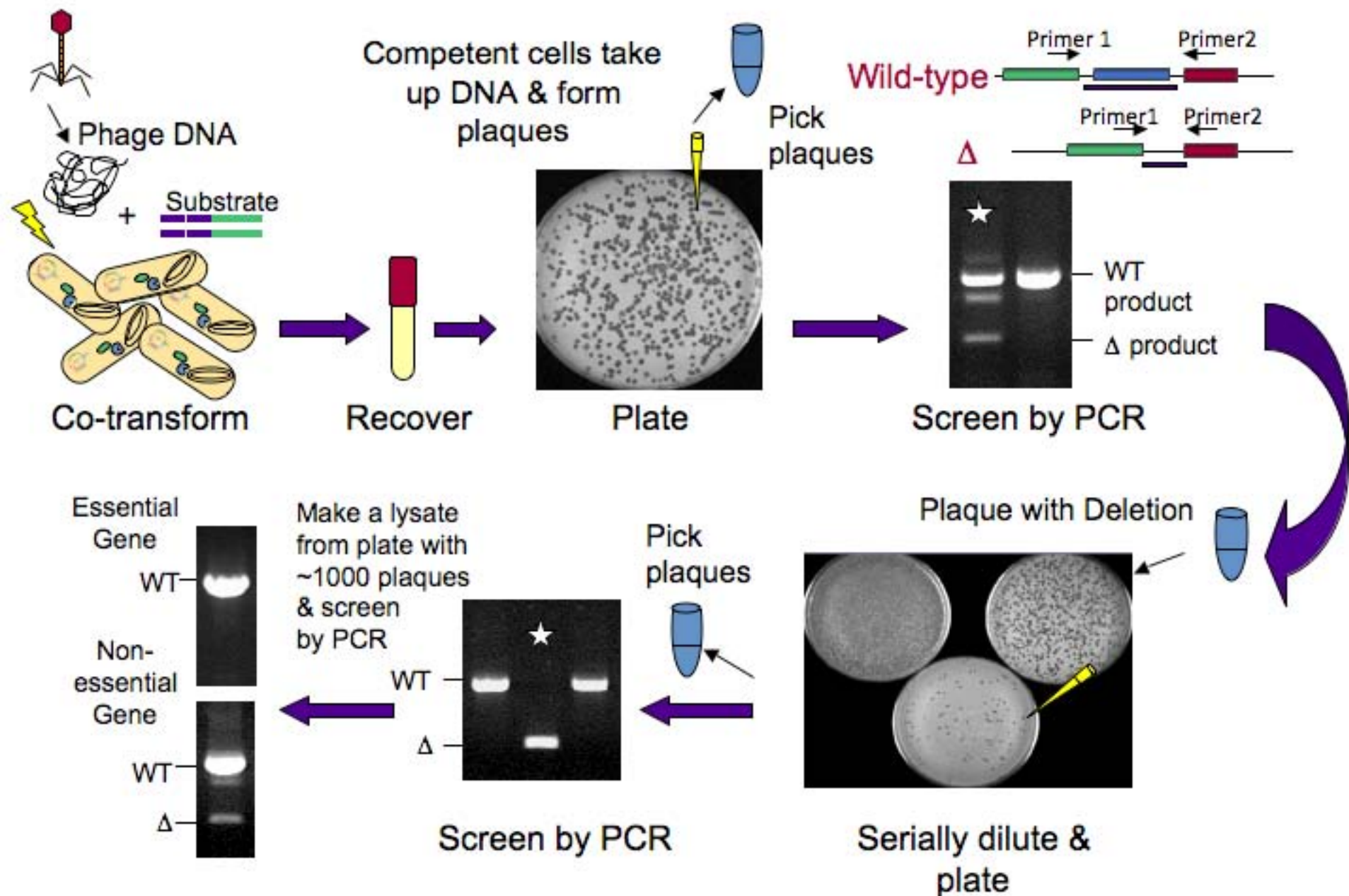
Cluster G phages

Genome mosaicism

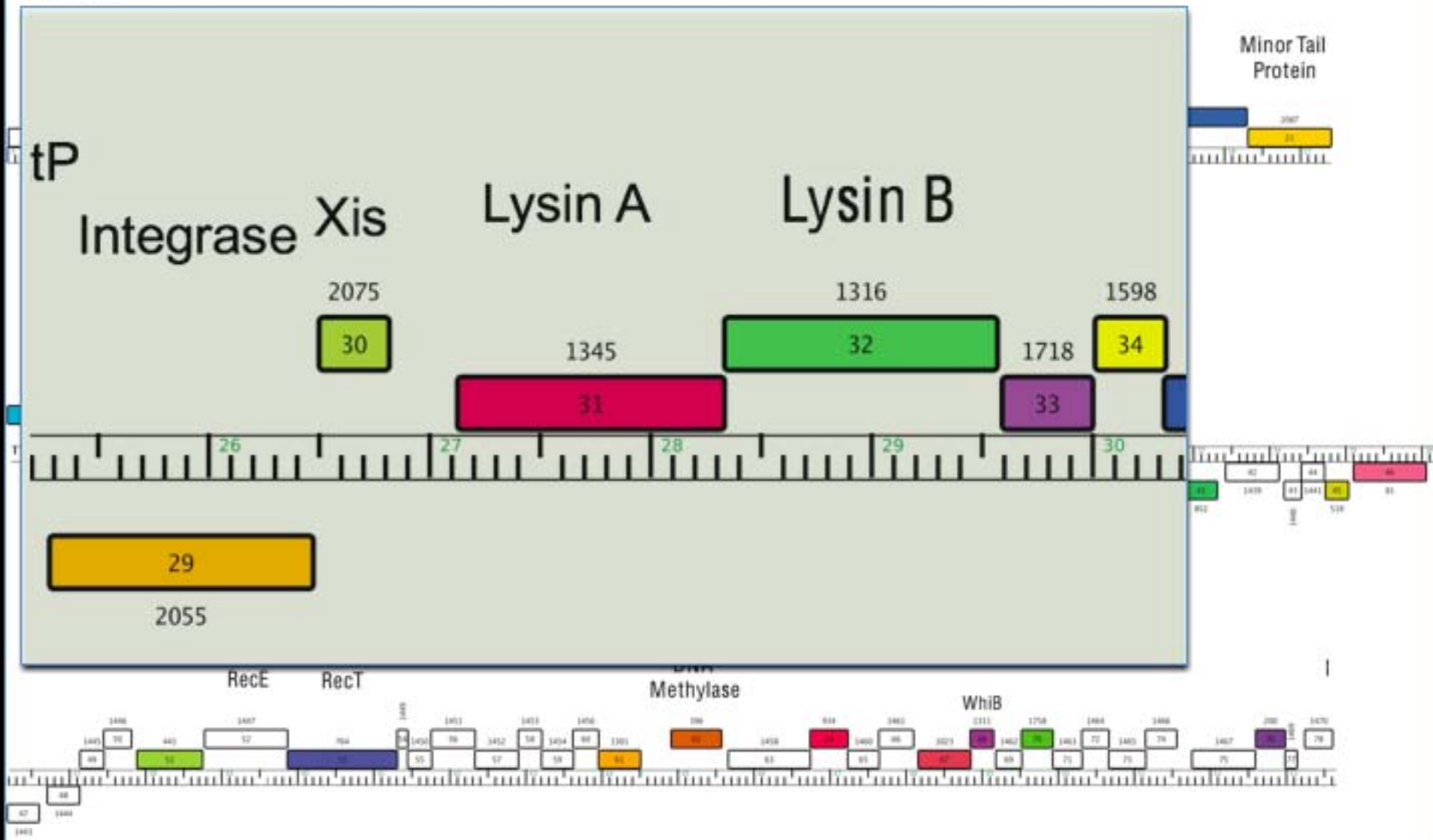# We don't know what most of these genes do!

We don't know what most of these genes do!

- 153 genomes
- 17,087 Orfs
- 2,942 Phamilies
- Mean orfs/Pham = 5.8
- 1212 Orphams (41.2%)
- 80% Phams novel genes

BRED: Bacteriophage recombineering on Electroporated DNA

Phage DNA + **Substrate**

Co-transform → **Recover** → **Plate** → Competent cells take up DNA & form plaques → Pick plaques → **Screen by PCR**

Wild-type — Primer 1 ... Primer2

Δ — Primer1 ... Primer2

WT product
Δ product

Plaque with Deletion

Serially dilute & plate → Pick plaques → **Screen by PCR** → Make a lysate from plate with ~1000 plaques & screen by PCR
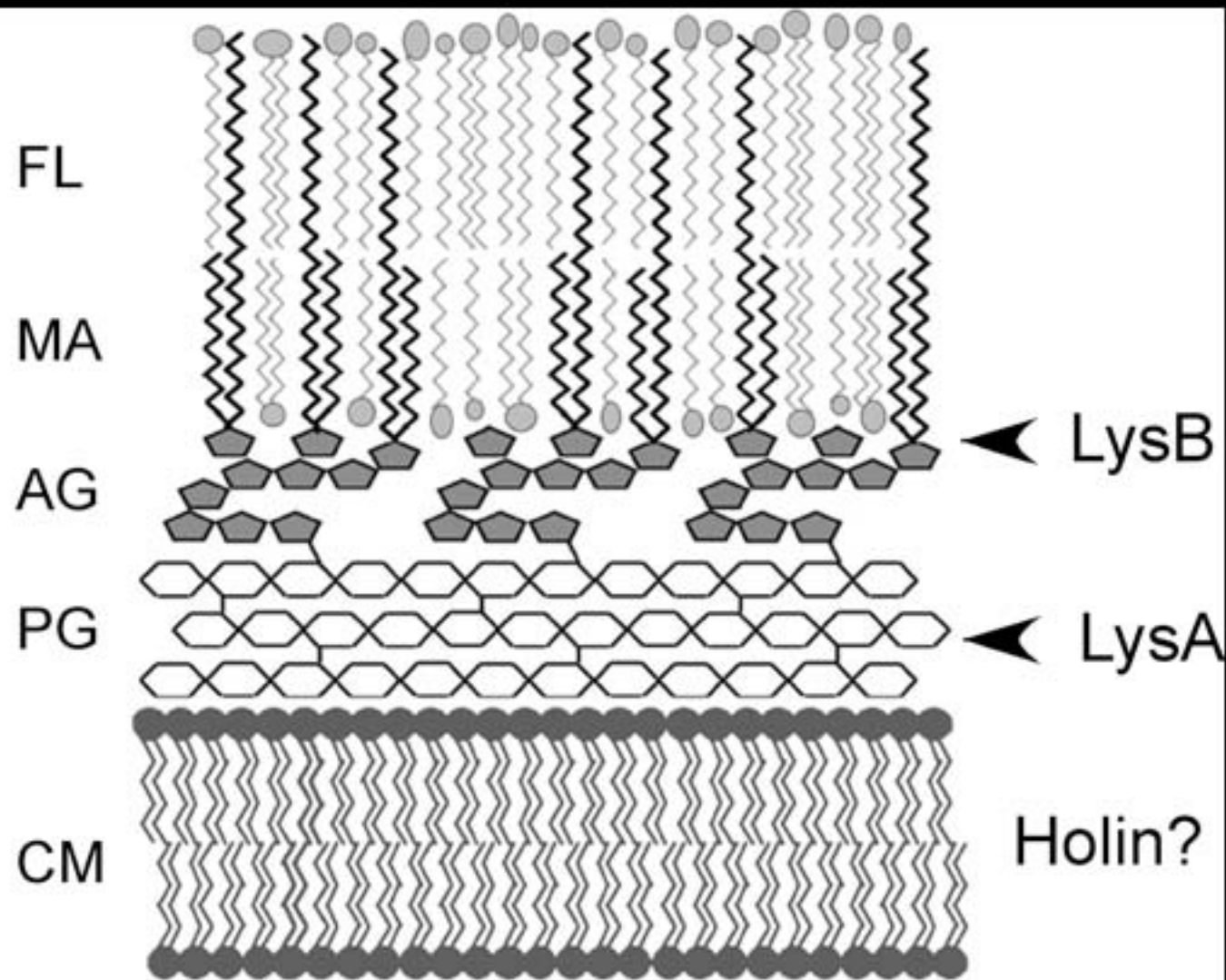
WT
Δ

Essential Gene
WT

Non-essential Gene
WT
Δ

Giles: Example of a genome map

Roles of Giles Lysin A, Holin, and Lysin B in lysis

In conclusion...

• Mycobacteriophages are highly diverse

• They have architecturally mosaic genomes

• 80% of genes are of unknown function

• The genomes can be readily manipulated

• They present a powerful toolbox for understanding and controlling tuberculosis

Acknowledgements

Steve Cresawn
Mike Ford
David Hanauer
Roger Hendrix
Bill Jacobs
Debbie Jacobs-Sera
Ching-Chung Ko
Laura Marinelli
Jeffrey Lawrence
Kim Payne
Marisa Pedulla
Craig Peebles
Dan Russell
Julia van Kessel

National Institute of Health
Howard Hughes Medical Institute