SAND2022-13385C

Sandia National Laboratories

Exceptional service in the national interest

Applications of Precise Prophage Mapping

Catherine M. Mageeney, Ph.D.

Sandia National Laboratories

Livermore, CA

Presented at JGI VEGA October 6, 2022

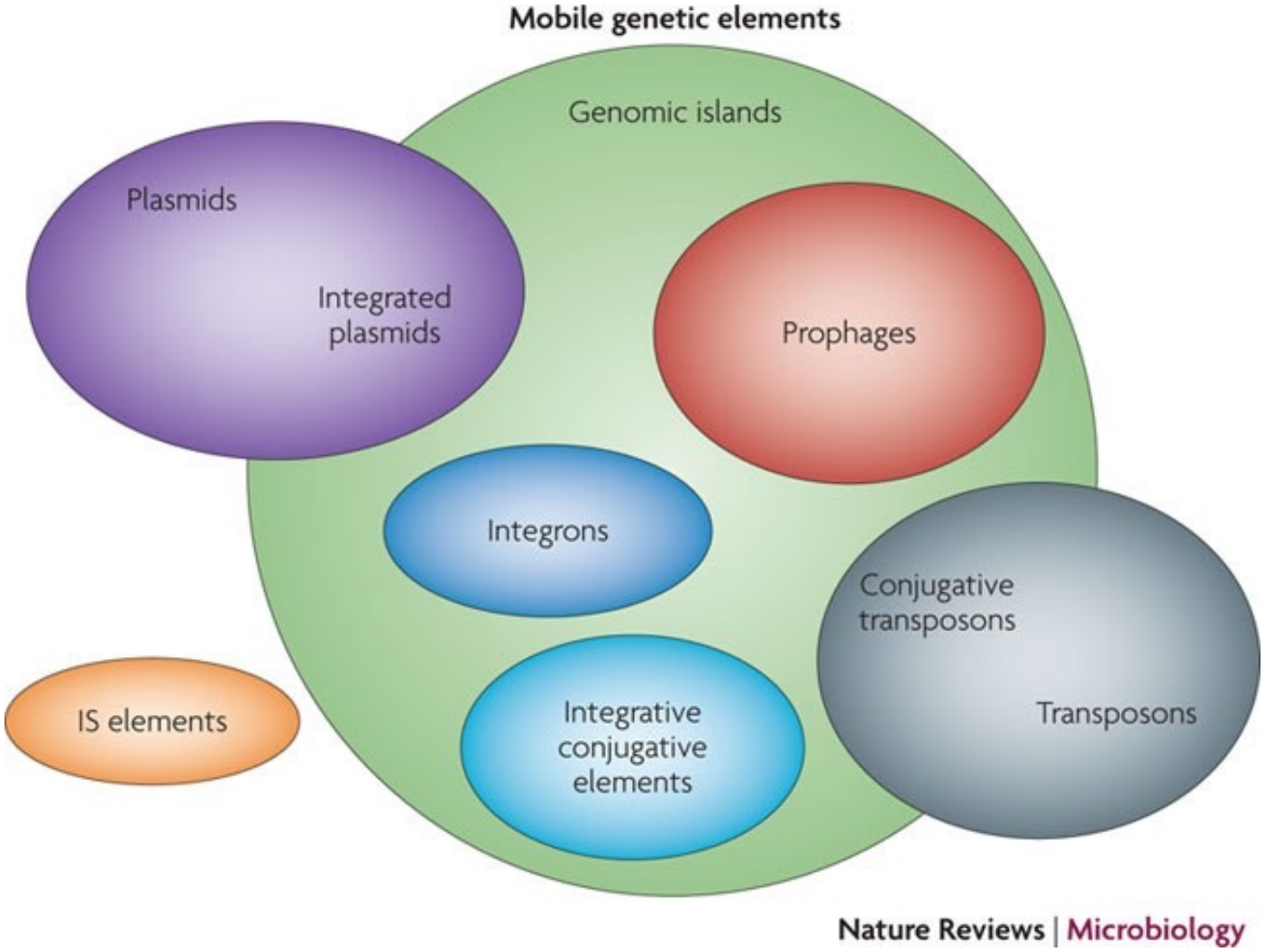
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Overview

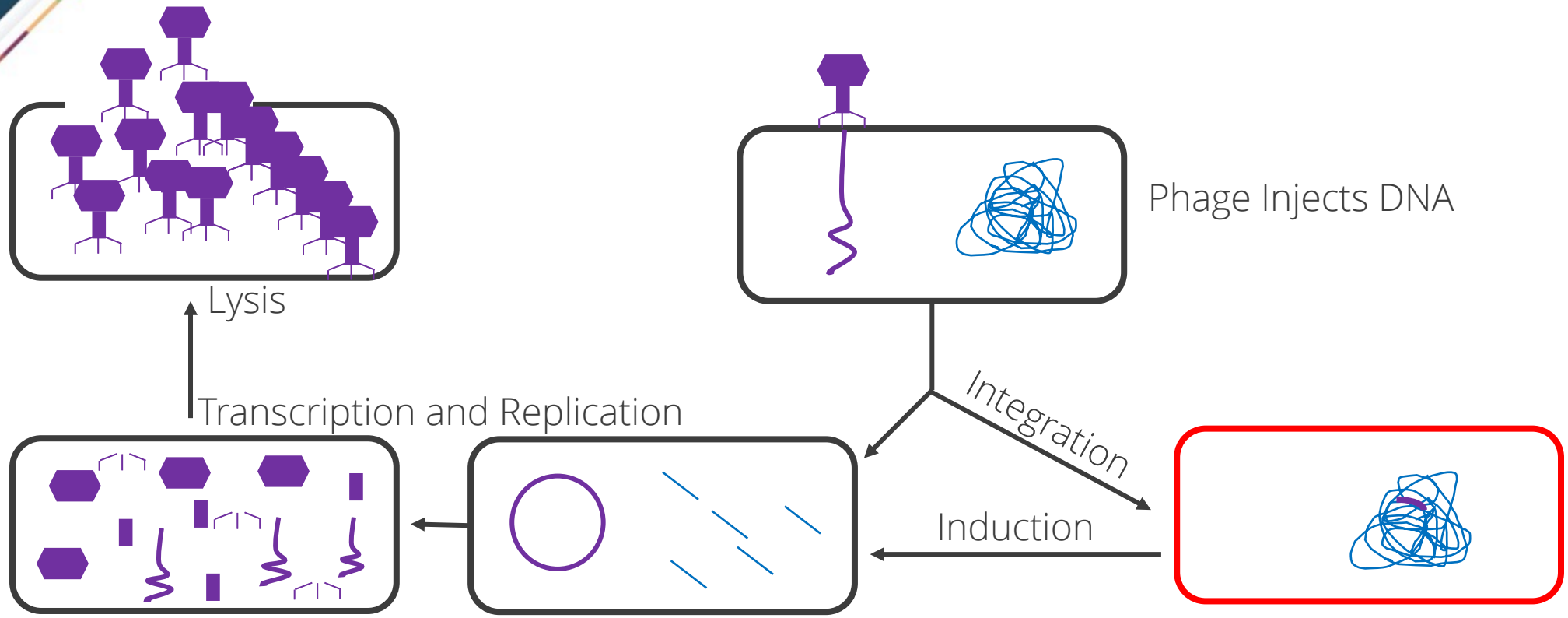
- Prophage Biology
- Precise Prophage Mapping
  - TIGER
  - TIGER database speedup
  - TIGER2
- bigDNA
- Phage Factory
  - For therapeutic applications applied to *Pseudomonas aeruginosa*
  - For energy applications applied to *Burkholderia cepacia* complex
- HES-PICI

Prophages are a subclass of mobile genetic elements



Langille, Hsiao, Brinkman, Nature Reviews Microbiology, 2010

Prophages mined from bacterial genomes yield far more phages



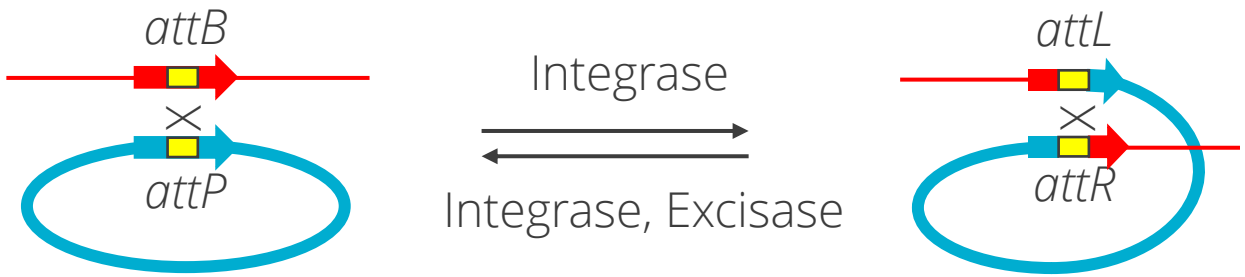
Traditional Way

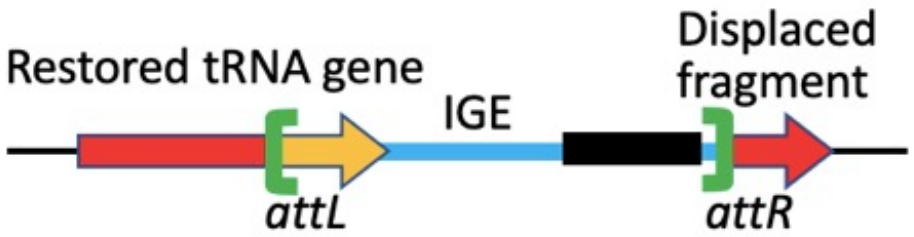
- Fishing approach can have low yield
- Not necessarily host-adapted (may be better adapted to other host bacteria)
- ~17,500 unique genomes in GenBank

Our WAY

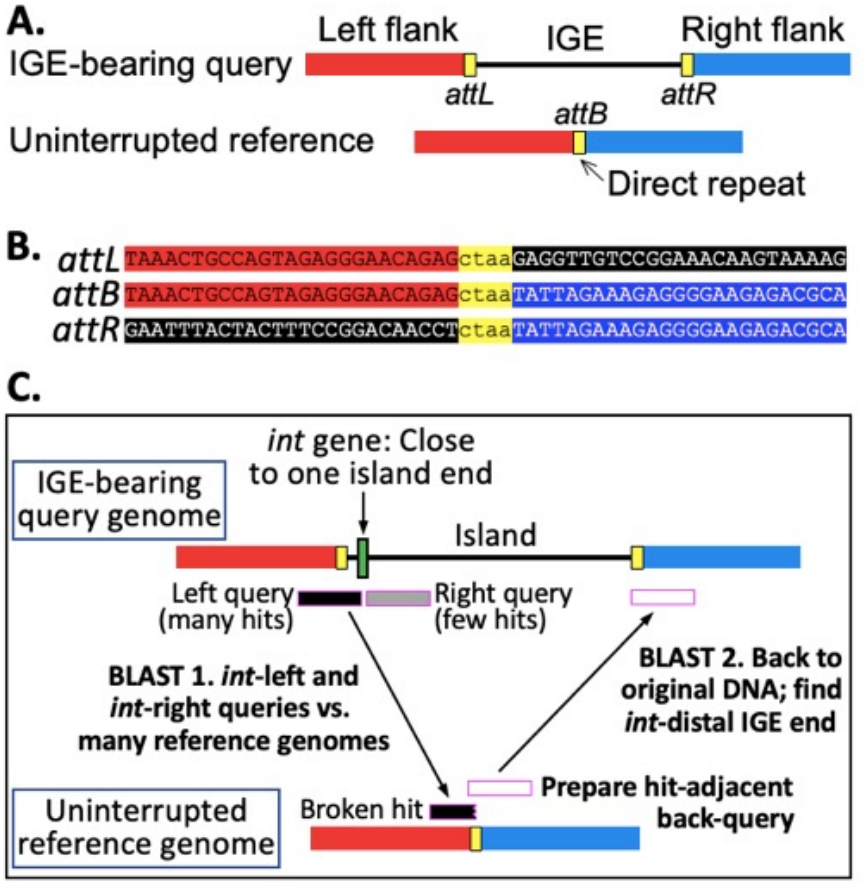
- Bacterial genomes are nets that catch phages
- Phages are host-adapted because we choose them from close relatives

Our software discovers genomic islands precisely



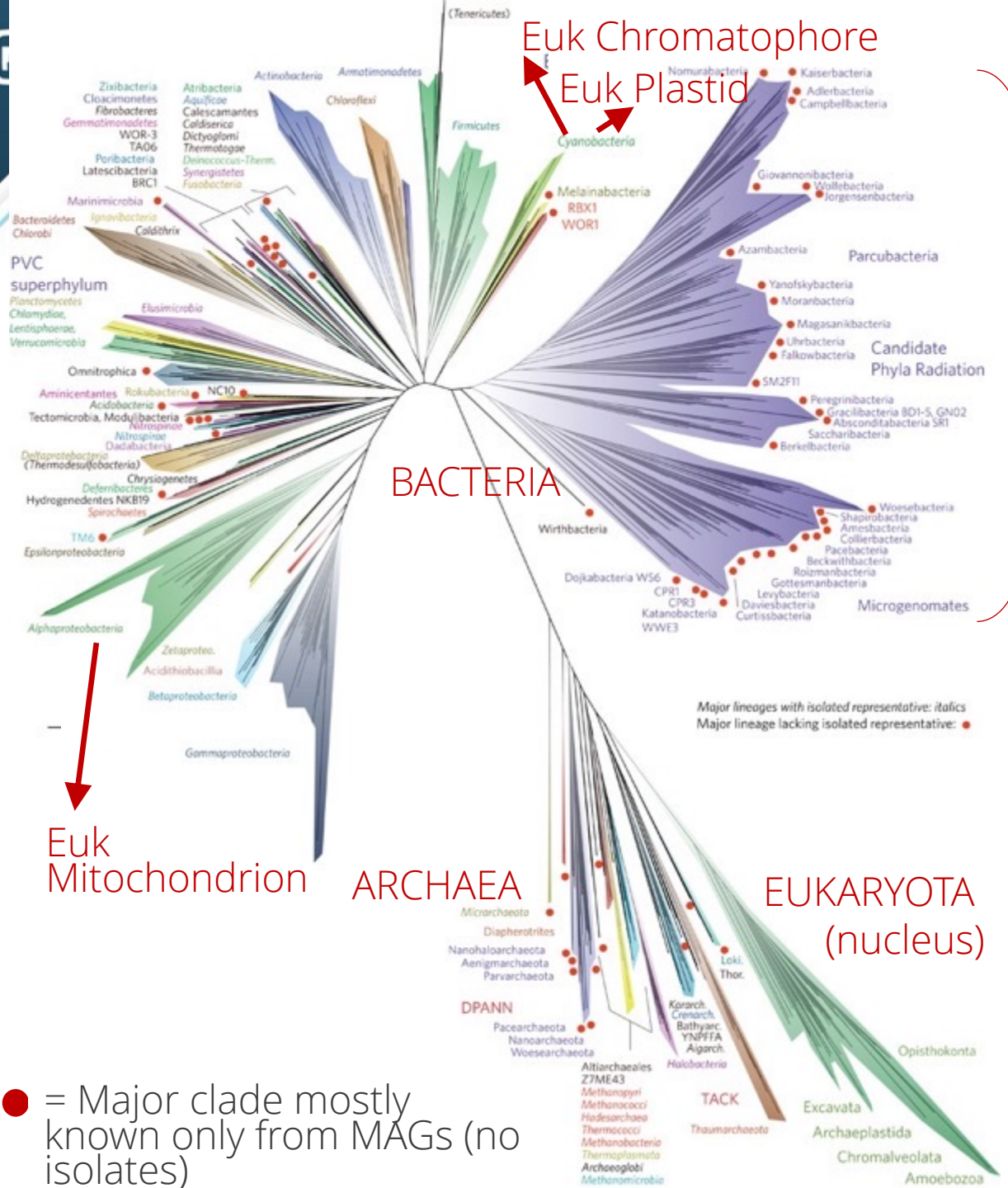


Islander (Hudson et al., 2015, NAR)



TIGER (Mageeney et al., 2020, NAR)
TIGERv2 (Mageeney et al., 2022, Frontiers in Bioinformatics)

Genomic Island Discovery for the tree-of-life

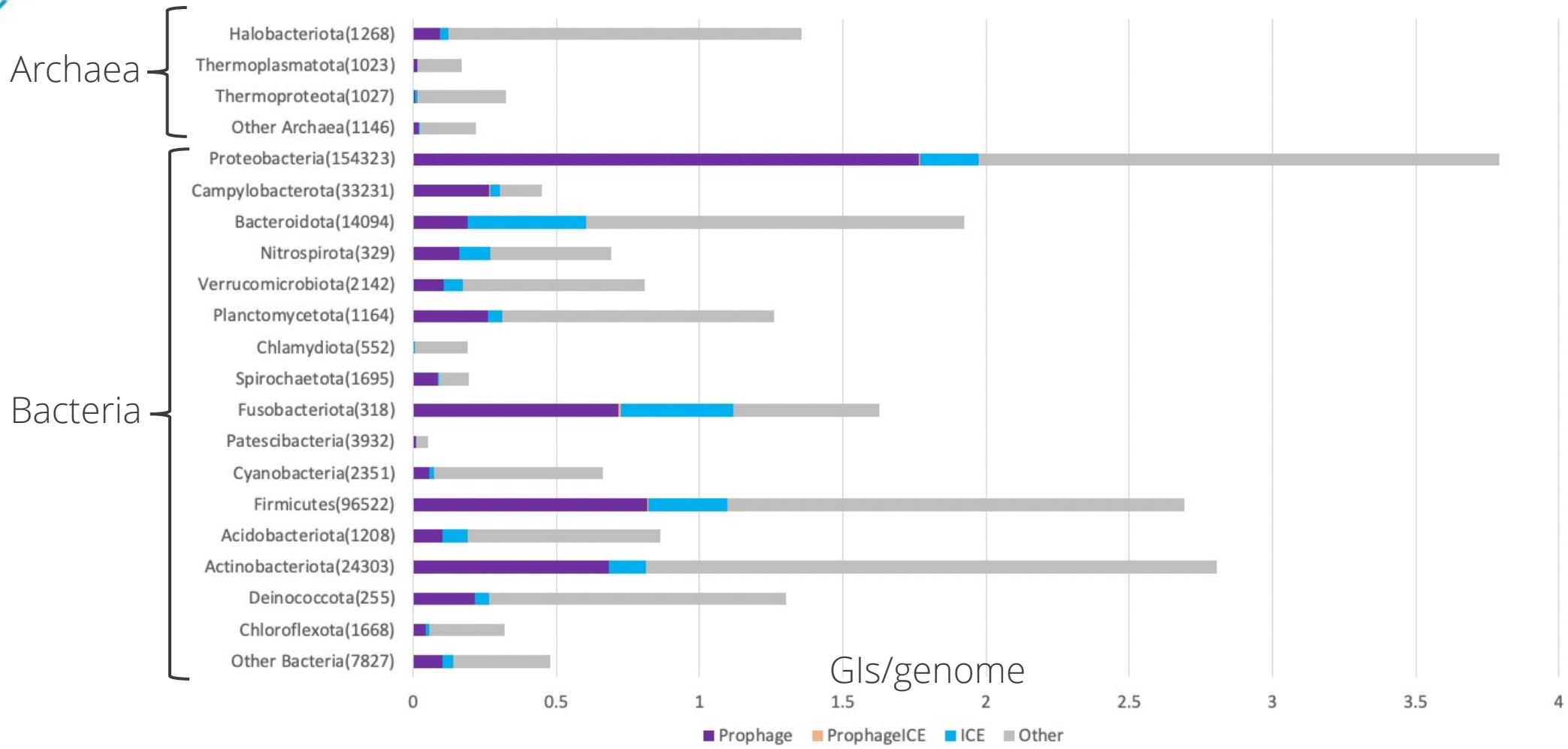


- Treated <u>all</u> ~48000 GTDB species comprised of >350K genomes
- Found 969,929 GIs = 2.8 GIs/genome
  - 382,576 Prophages = 1.1 prophage/genome

● = Major clade mostly known only from MAGs (no isolates)

Unpublished Data

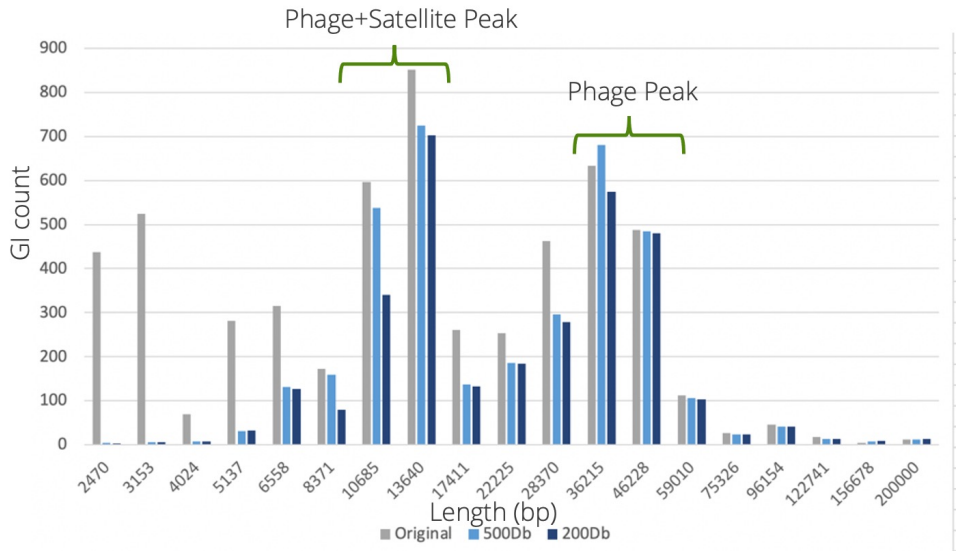
The breakdown of island type across different host taxonomic groups varies greatly



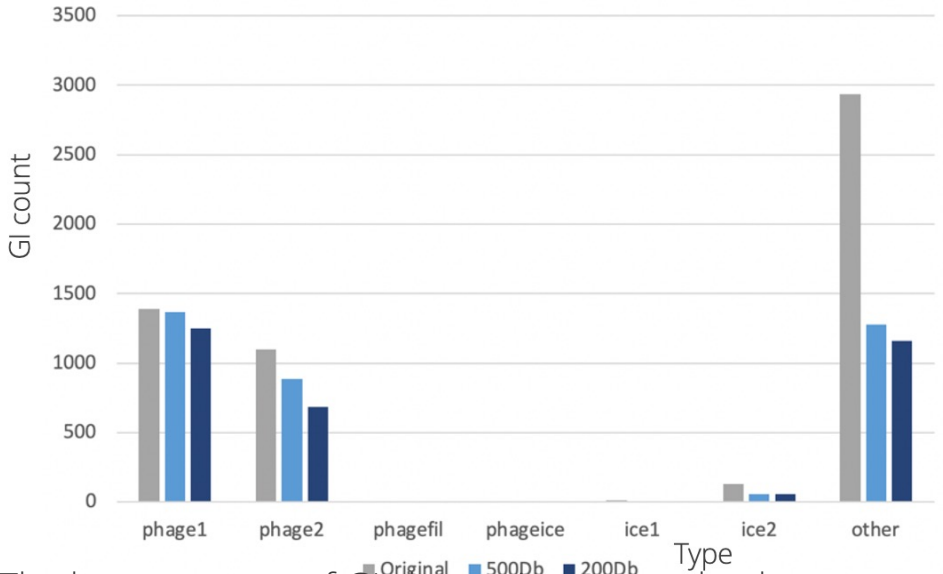
Unpublished Data

Using smaller, carefully-chosen databases speeds up TIGER more than 250x without the loss of prophages or large GIs

- Tested on 58K Salmonella genomes
- Used three databases: 58K, a 200 member and 500 member
  - Maximize diversity
  - 10% out of GTDB species tested
- Speed up was >250x faster
  - Original – 706.4 hours
  - 500DB – 2.78 hours
  - 200DB - 1.17 hours
- Lost small GIs which contained non-canonical integrases



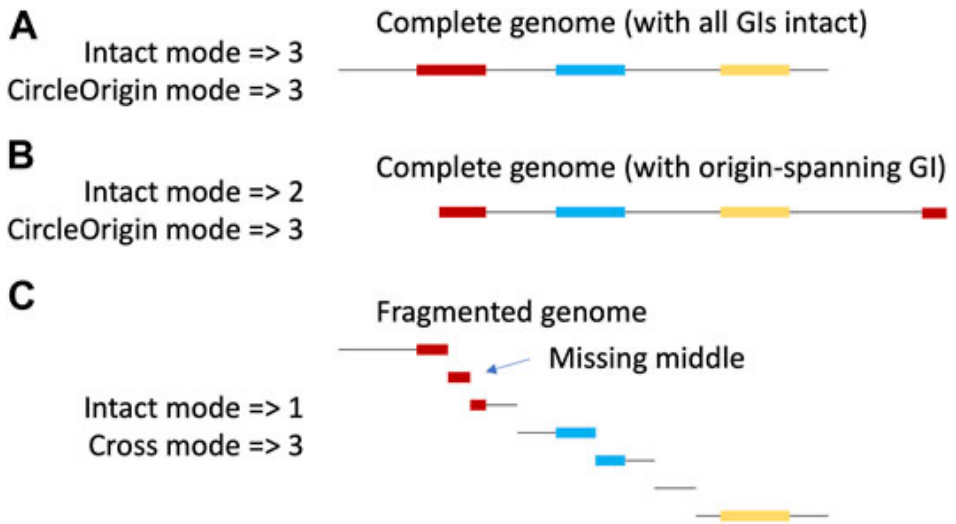
Size Distribution of Salmonella GIs with each database type. Smaller GIs (<8kbp) were predominantly lost

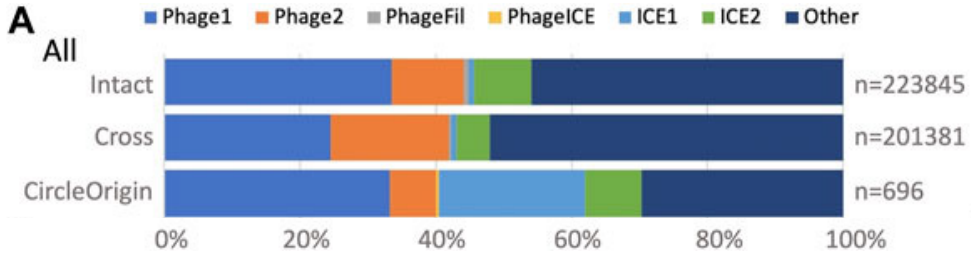


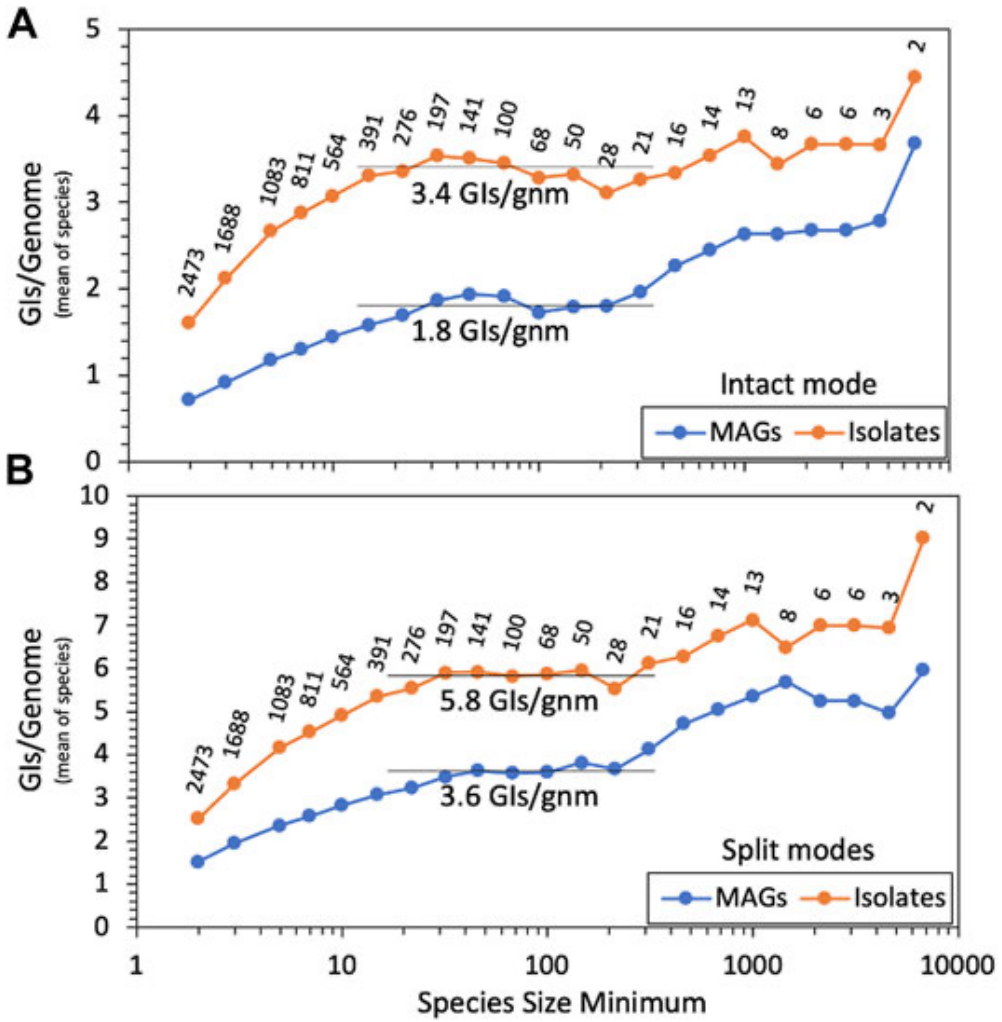
The largest group of GIs lost using smart databases were those which could not be classified as ICE or Prophages.

TIGERv2 allows GI searches across scaffolds and expands precise GI detection into metagenomes

- Needed to detect GIs across scaffolds
- Allowed for expansion to MAGs
- Increased GI count 2x





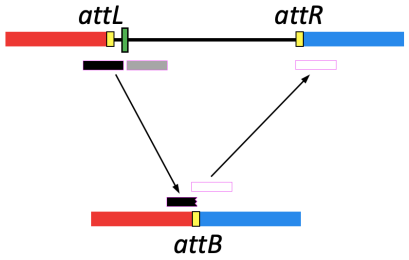


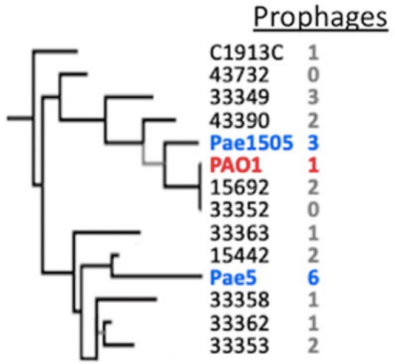
TIGERv2 (Mageeney et al., 2022, Frontiers in Bioinformatics)

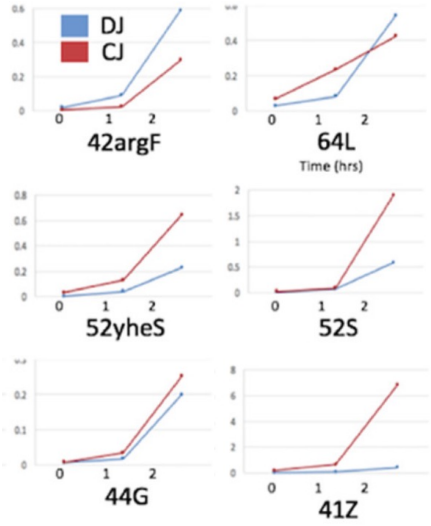
A genomic island database allows new understanding and asks further questions

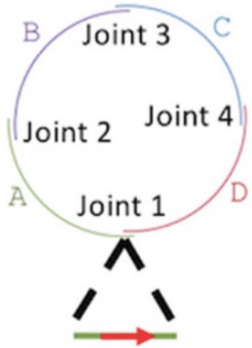
- Mapping all genomic islands allows
  - Better understanding of horizontal gene transfer
    - Delimits natural mechanisms for loss of biocontainment
  - Better understanding of prophage host ranges
  - Understand cross talk between prophages/MEs and prophage genome mosaicism
    - Gene flows between host and ME, between MEs
  - Mapped att sites for 1M+ integrases
  - Diversity of prophages across the tree of life
    - Currently only mapped well in Mycobacterium
- We can begin to ask questions such as
  - Why some bacterial species have large numbers of prophages and others do not?
    - Annotation problem
    - Biological mechanism
    - Poor sequence quality
  - What mechanisms regulate HGT in different species?
    - Do integrases have the same host ranges as the islands that contain them?

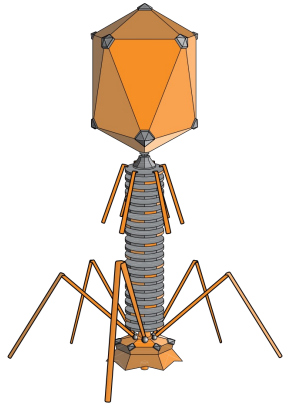
Phage Factory: Bacteria-Agnostic Phage Discovery and Engineering Platform







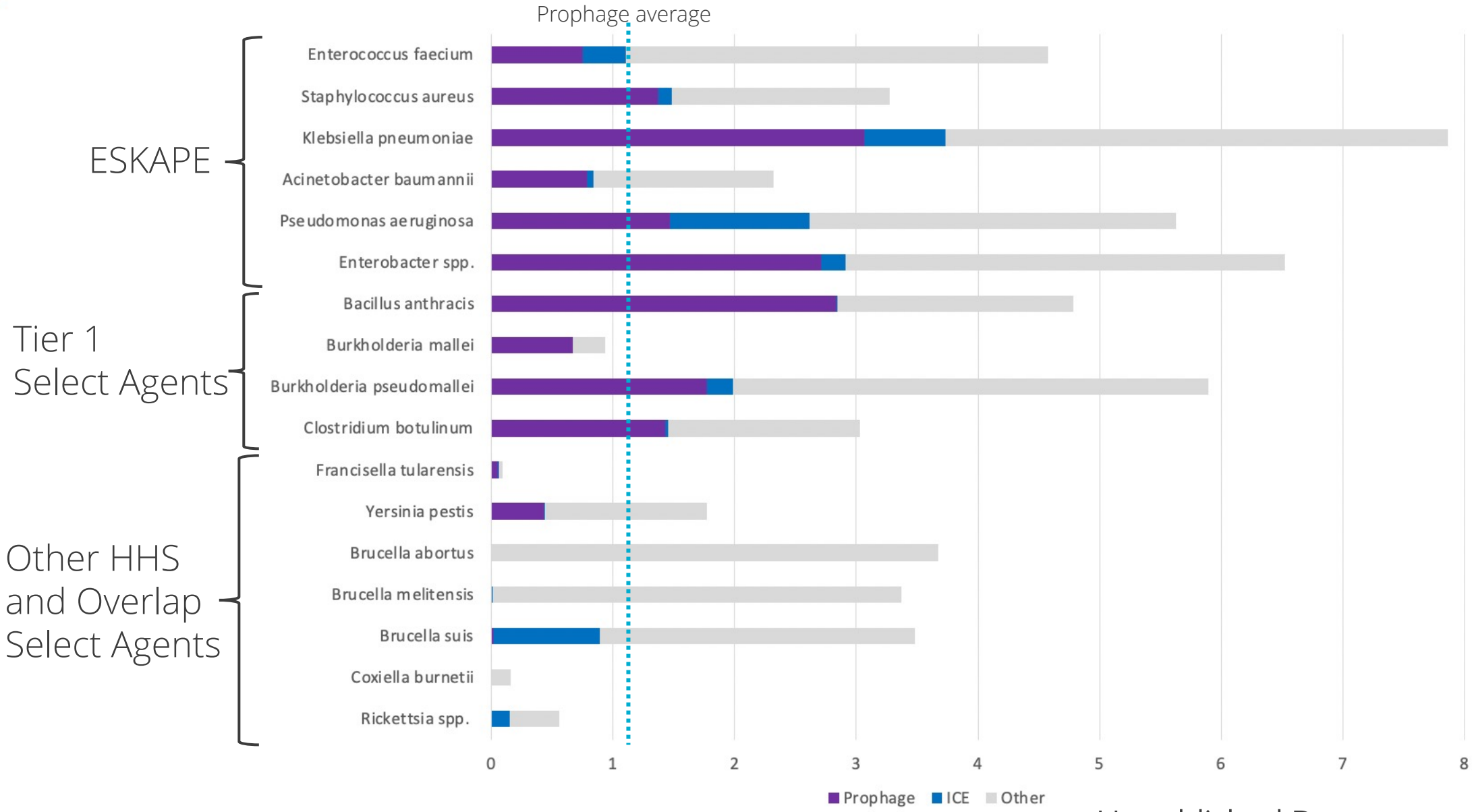




Prophage database → Identify prophage-rich close relatives of target → Validate phages → Engineer phages → Reboot

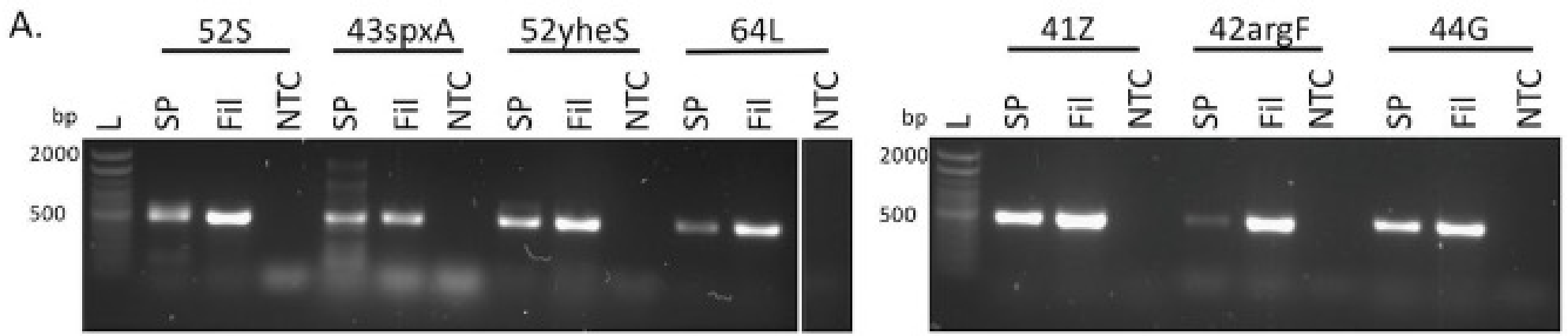
Mageeney *et al.*, 2020, mSystems

Prophage yields are higher than average for pathogens



Unpublished Data

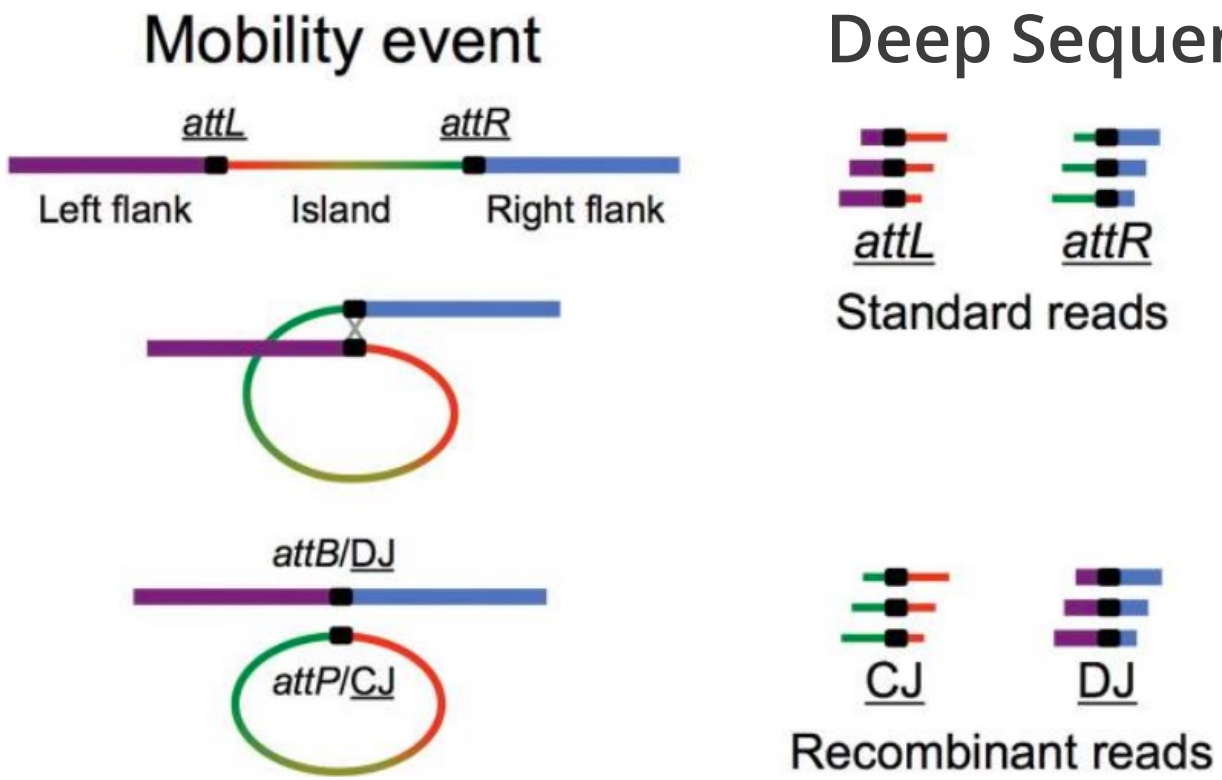
Active prophages can be identified by PCR …



SP = Soaked plaque
Fil = Filtrate
NTC = blank

Mageeney et al., 2020 mSystems

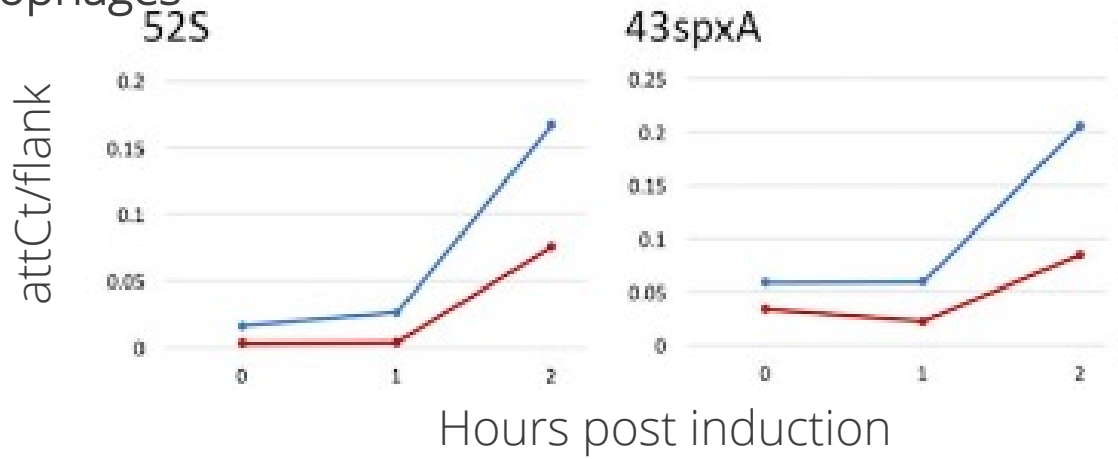
Deep sequencing can detect active prophages



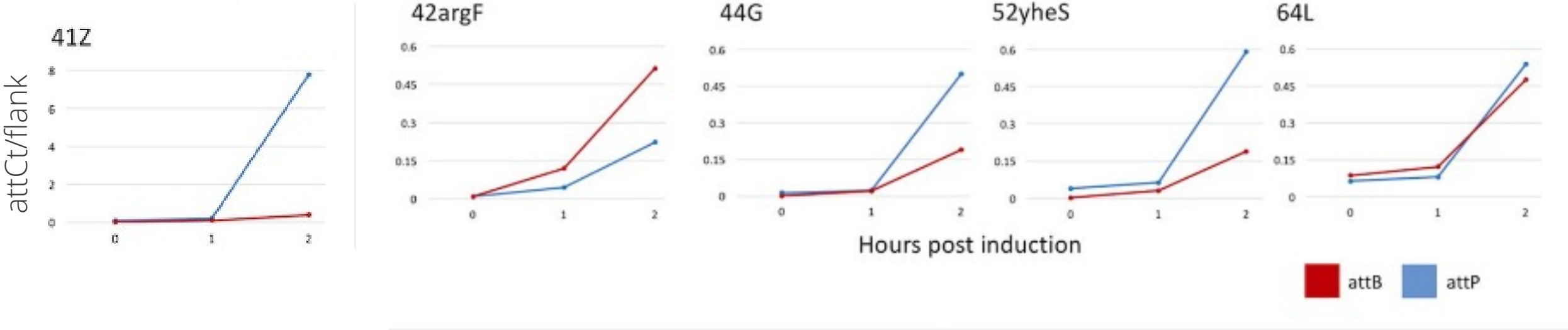
Juxtaposer and attCt - Schoeniger et al., 2016 NAR

Deep sequencing reveals differential island induction behaviors

Pae1505 Prophages

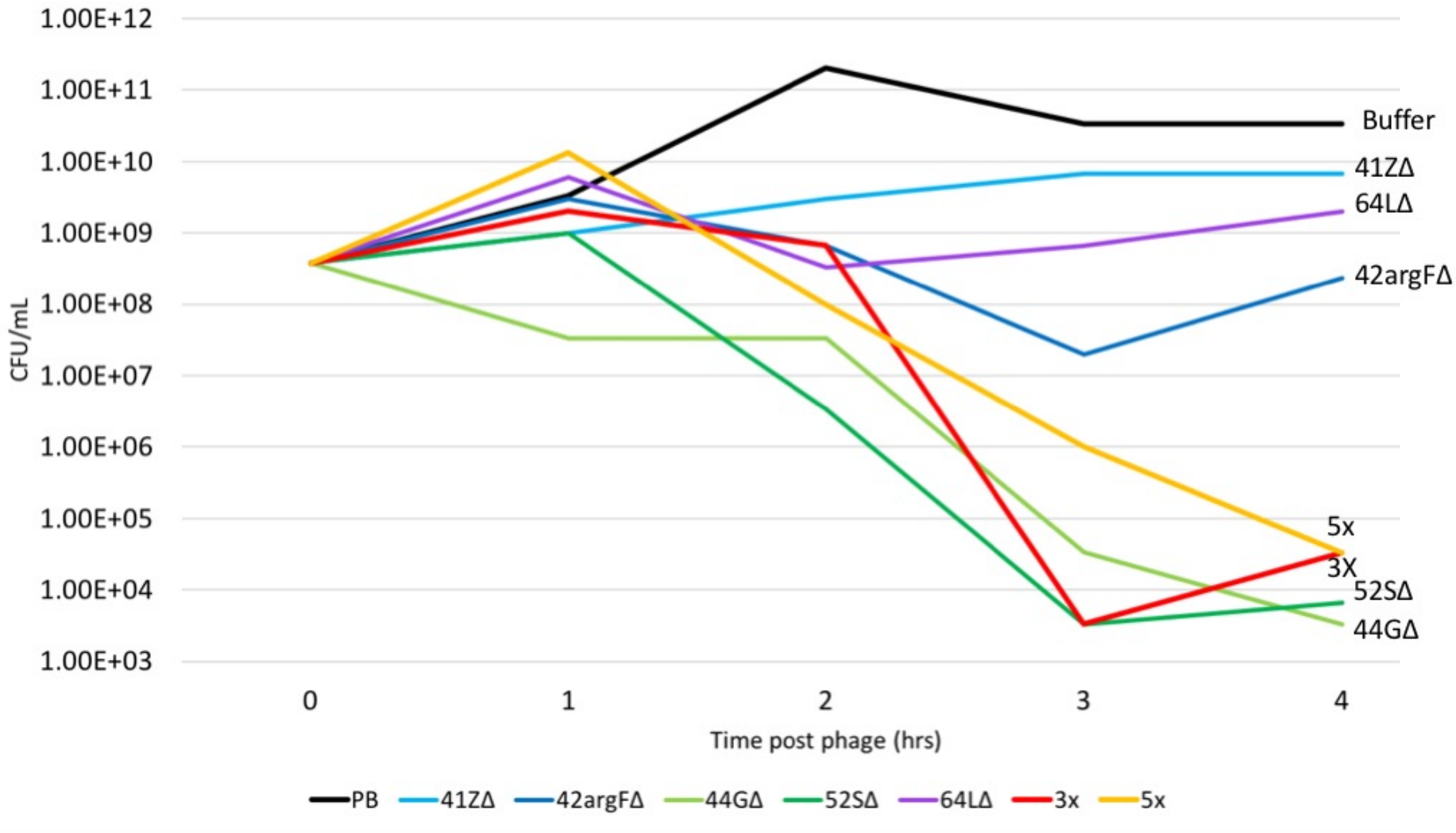


Pae5 Prophages



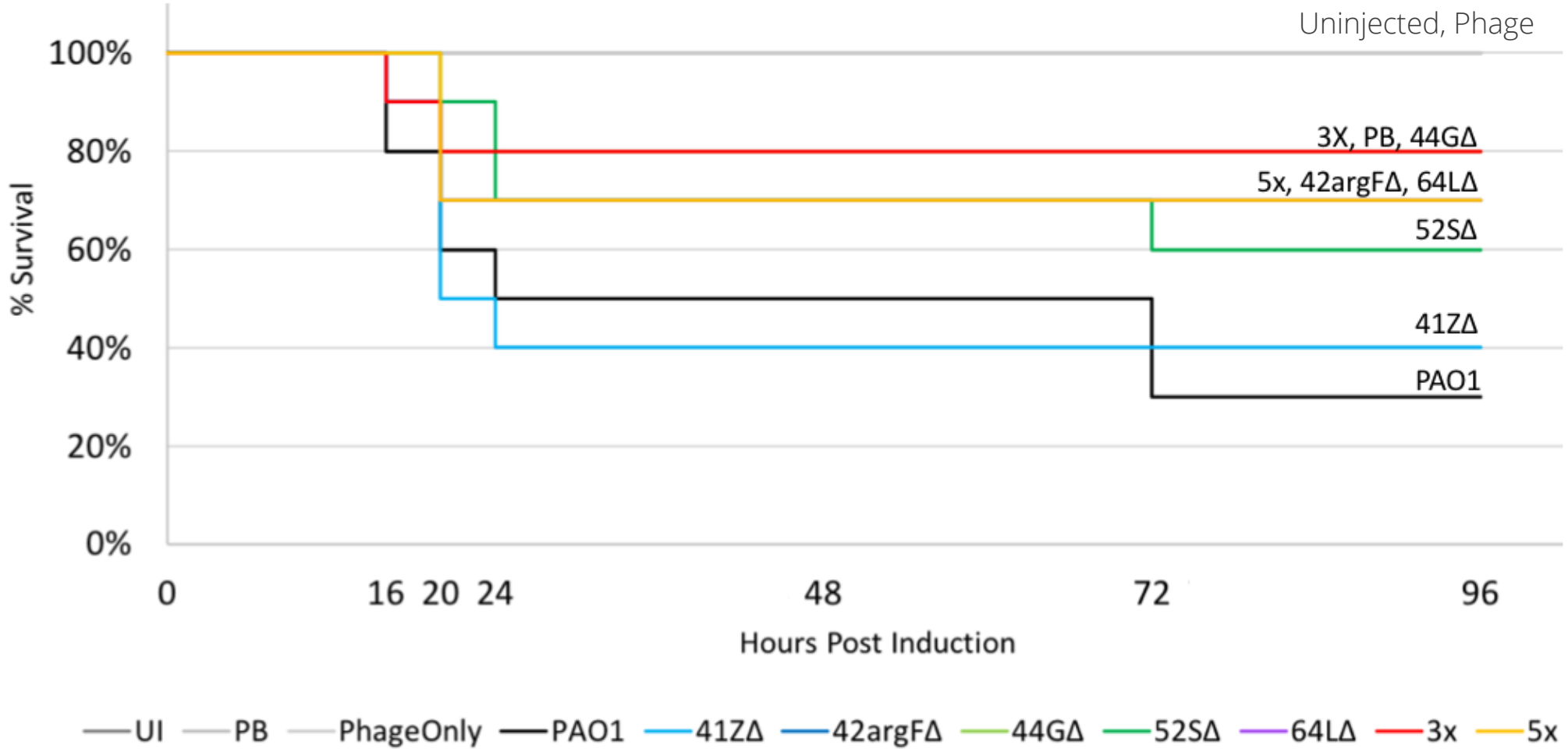
Mageeney et al., 2020 mSystems

Engineered phages kill ***Pseudomonas aeruginosa*** in liquid culture



Mageeney et al., 2020 mSystems

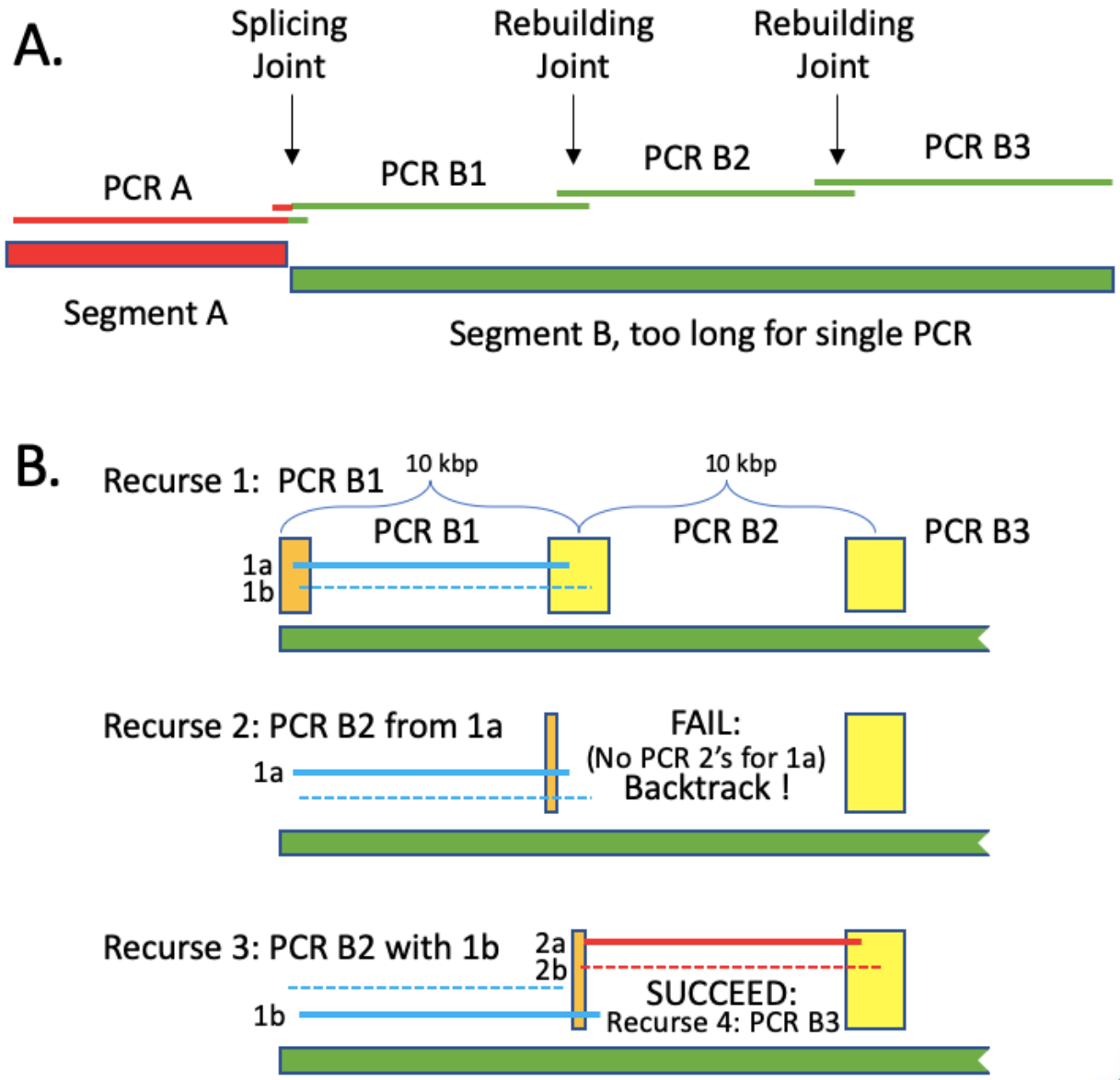
Engineered phages save waxworms in phage therapy trials



Mageeney et al., 2020 mSystems

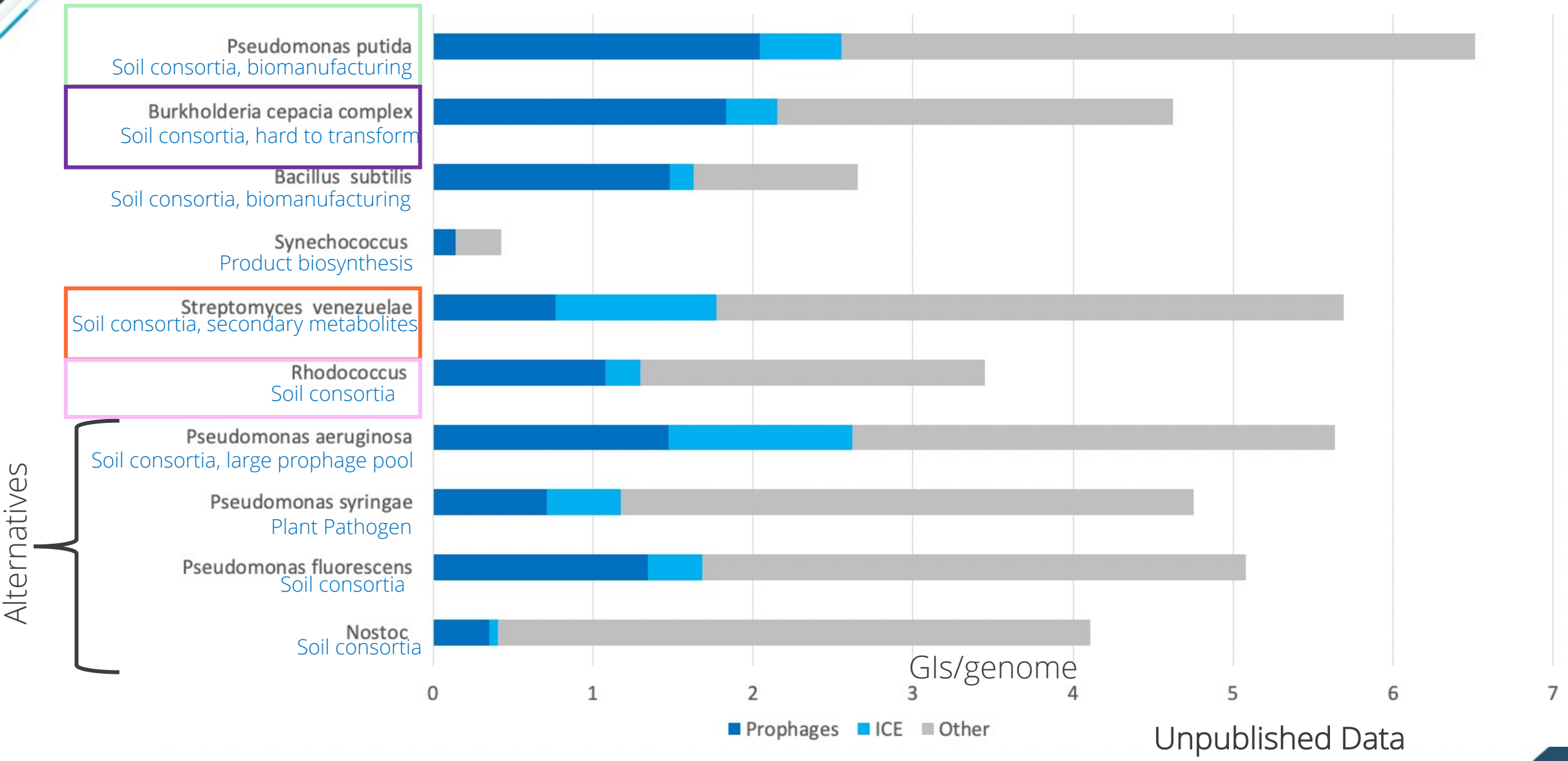
Precise prophage end mapping allows for better phage engineering design

- We typically build large Gibson Assemblies for our Phage Engineering
- Needed a high throughput way to accomplish this
  - Designed bigDNA software to use recursion/backtracking to design long PCR primers that will be capable of synthetic phage Gibson Assembly
- bigDNA can design WT phages, phages with gene deletions or gene additions



Vuong et al., Submitted to Phage

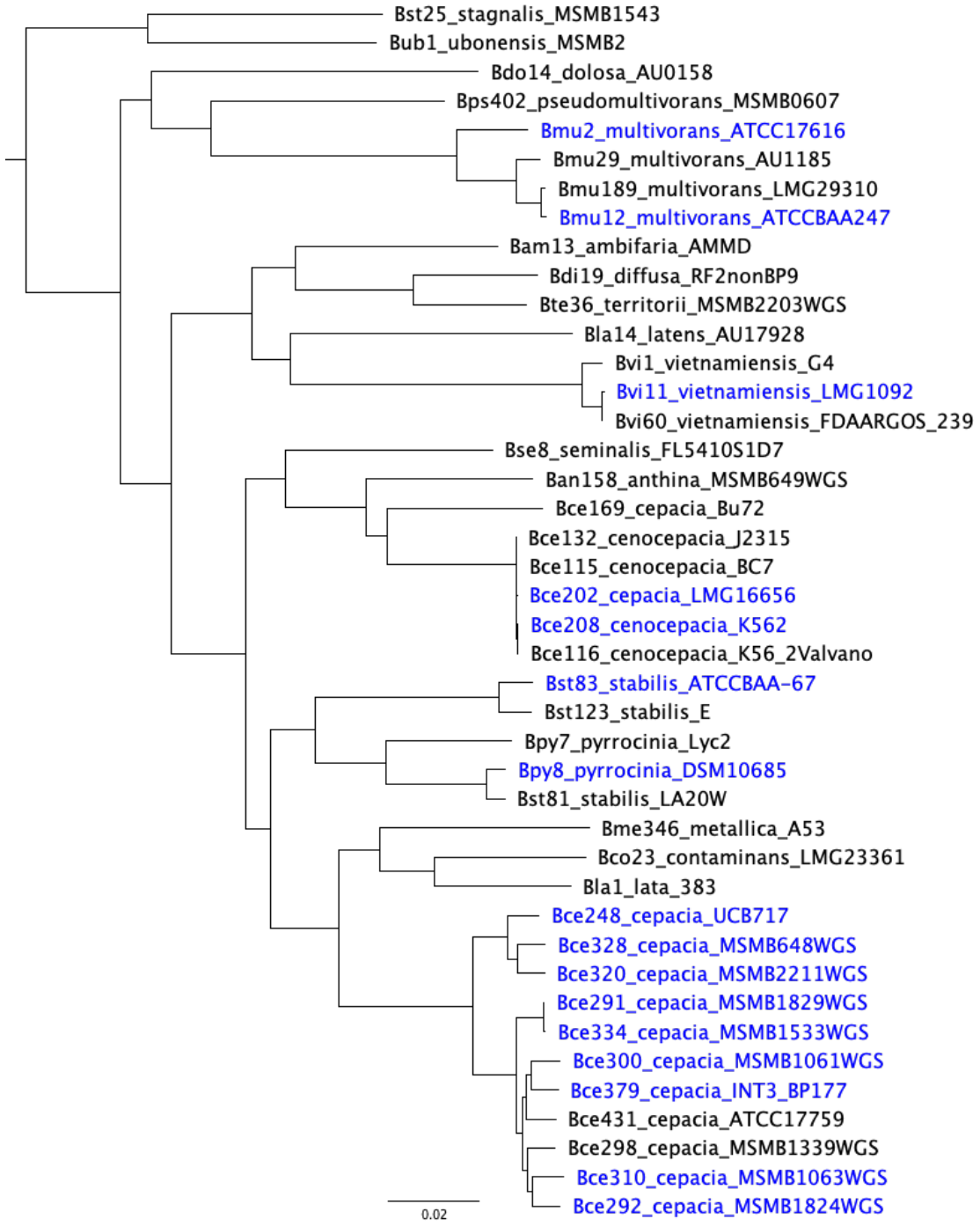
We have treated many bacteria that are related to energy applications



Unpublished Data

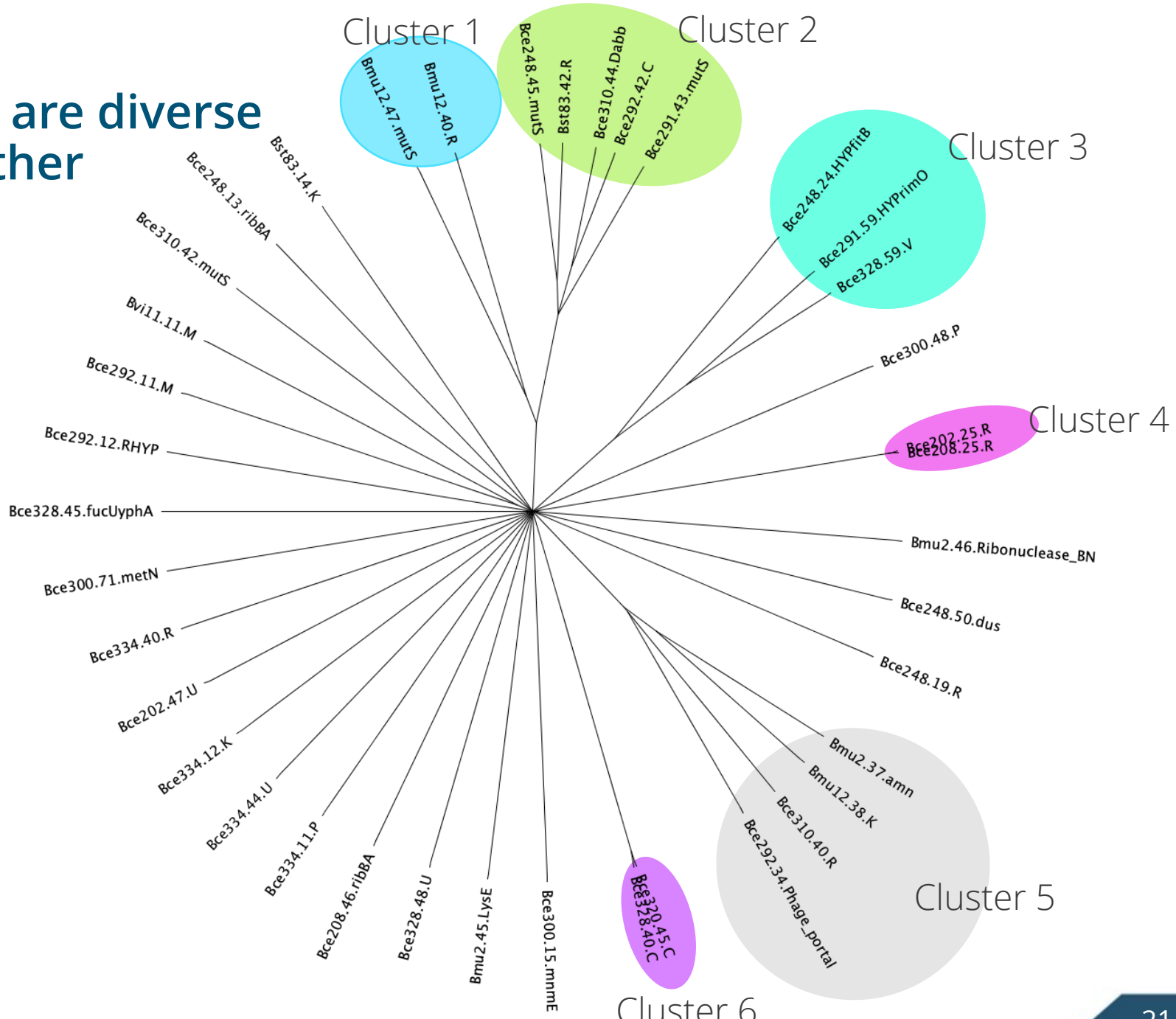
B. cepacia complex strains harbor many prophages

- 16 strains treated
  - 9 cepacia, 2 cenocepacia, 2 multivornas, 1 vietnamiensis, 1 stabilis, 1 pyrrocinia
- 123 Genomic Islands Discovered
- 39 Prophages
  - 30 Active Full Length
    - 20 Induced by MMC
    - 10 Not tested
  - 6 Decayed
  - 3 Filamentous



Burkholderia prophages are diverse with few clustering together

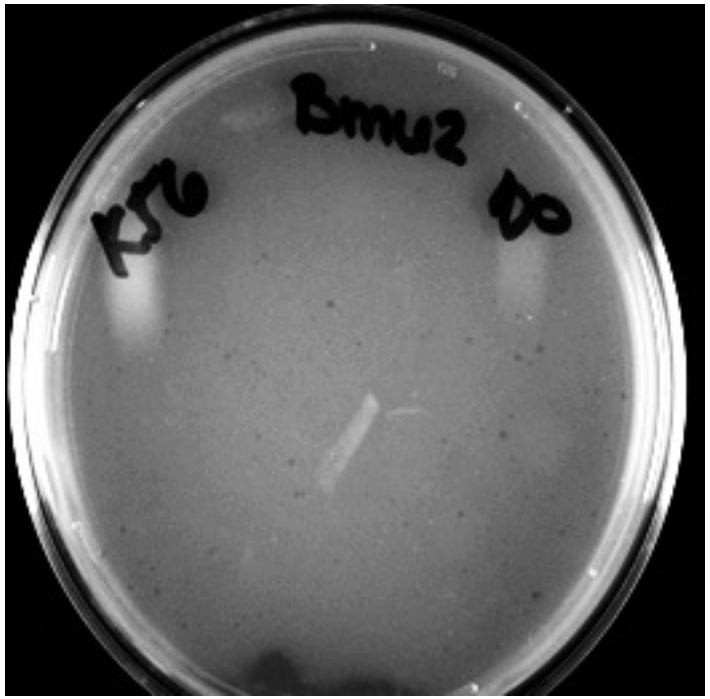
- Prophages form 6 cluster with 90% ANI spanning >50% of the genome
- There are 21 Singletons not clustered with any other prophages
- Most of these do not cluster with environmentally isolated phages



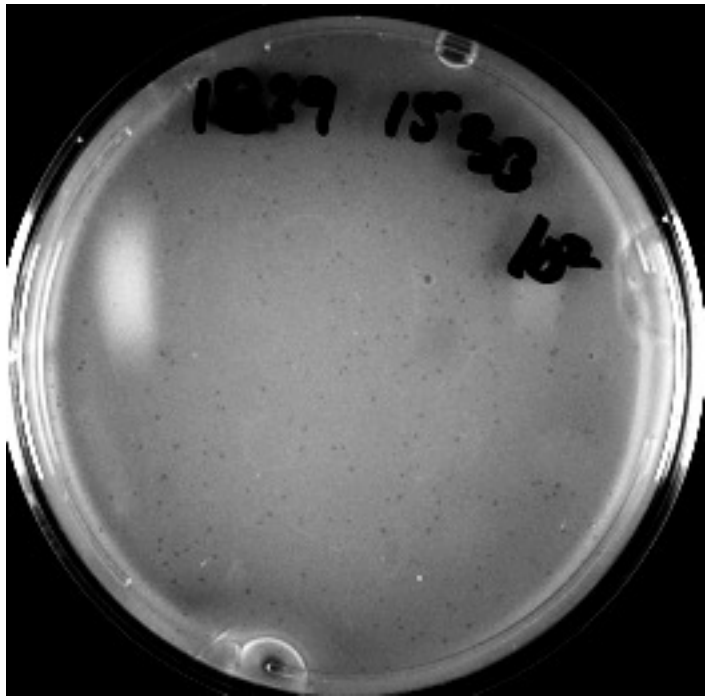
Prophage isolation suggests ***Burkholderia*** prophages have a large host range

We have isolated 3 prophages from various *Burkholderia* strains on

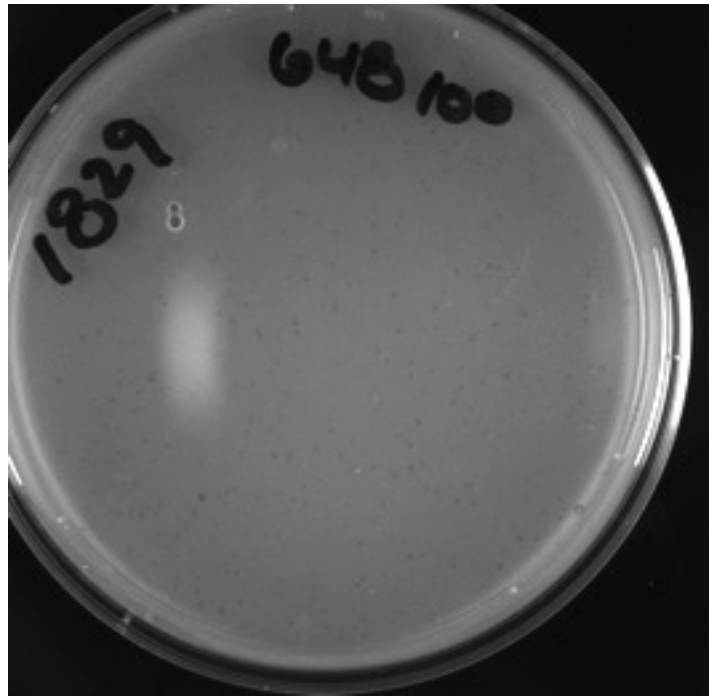
- *B. cepacia* 1829: Bce334.40.R and Bce328.59.V
- *B. cenocepacia* K56-2: Bmu2.37amn



Bmu2.37.amn



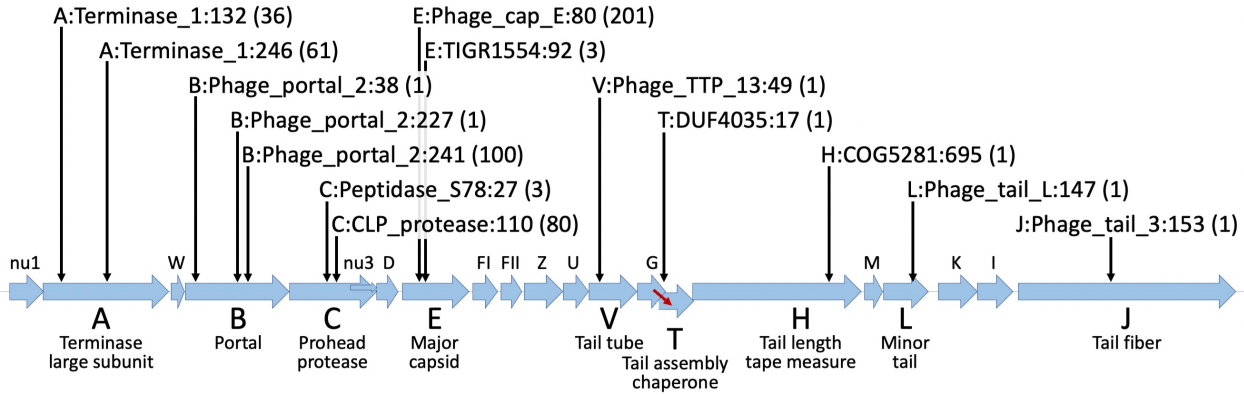
Bce334.40.R

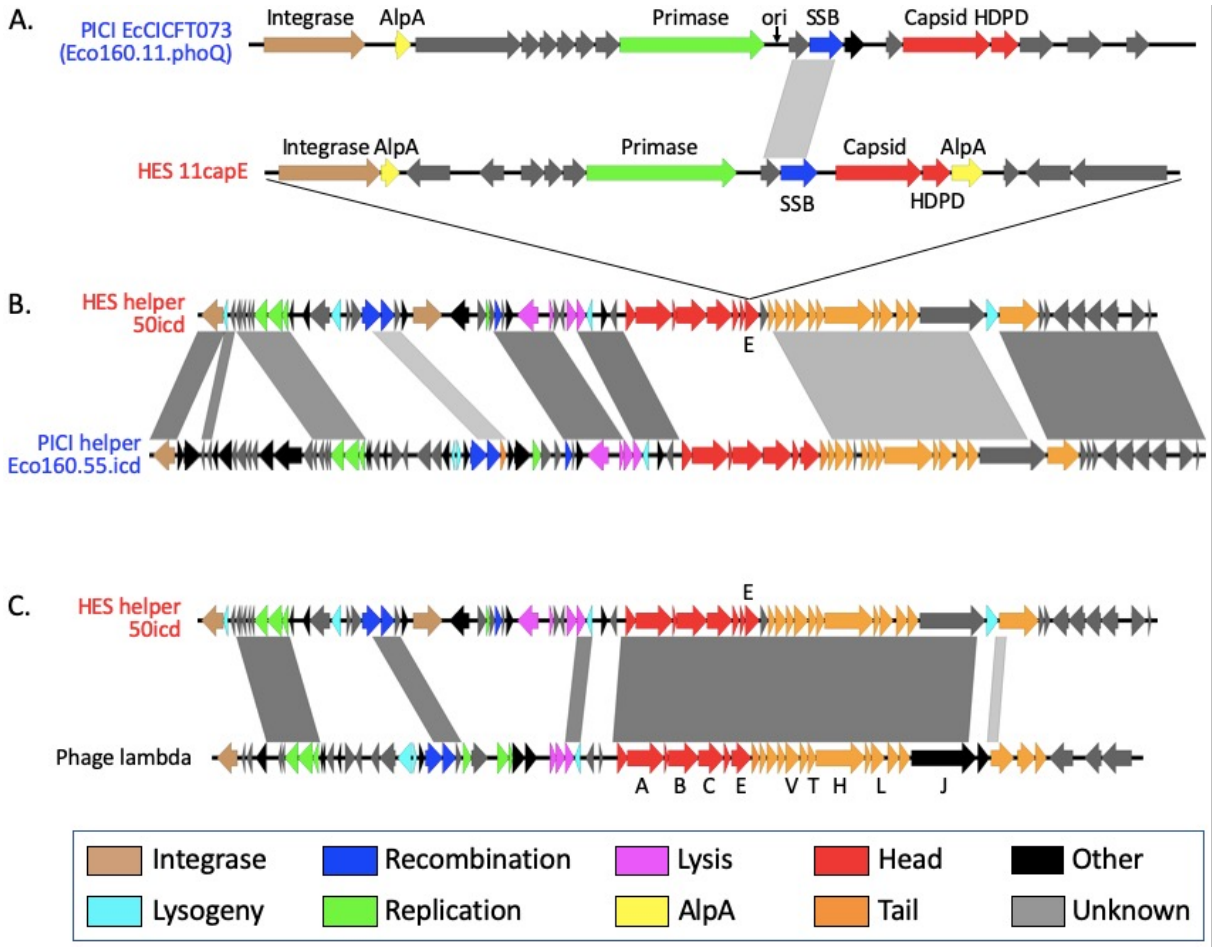


Bce328.59.V

Precise prophage mapping allows for discovery of new satellite phages

- Discovered small GI in our database (11capE) that appeared to be integrated into the capsid protein of another prophage (55icd) in a *E. coli* NRG857c
- This GI resembled the previously described PICIs
- New class of satellite discovered helper-embedded satellite phage-induced–chromosomal islands (HES-PICI)
- Found 491HES-PICI in *Enterobacterales* sp.





Tommasini et al., Submitted to NAR Genomics and Bioinformatics

Conclusions

- Prophages can be mined for every known bacterial species
- MAGs contain less GIs than isolates
- Engineered prophages can treat *P. aeruginosa* infections both *in vitro* and *in vivo*
- Burkholderia prophage have large host ranges
- HES-PICI can be found in numerous lambda cognate locations and are widespread in *Entrobacteriales*

Acknowledgements

Sandia National Laboratories

- Kelly P. Williams
- Joseph Schoeniger
- Steve Branda
- James Jaryenneh (Poster #)
- Anupama Sinha
- Raga Krishnakumar
- Jesse Cahill
- Year-Round Interns: Dario Tommasini, Lara Wiedmeier, Rohan Krishan
- DOE SULI Interns: Ivan Vuong, Catherine Ly, Madelyn Peck
- Summer Interns: Alicia Rokes, Britney Lau, Shawn Barman

Collaborators

Farren Isaac, Yale University
Michael Jewett, Northwestern University
Jennifer Doudna, UC Berkeley
Linda DeVeaux, New Mexico Tech

Funding: This material is based upon work supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under the Secure Biosystems Design Initiative project Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC);, and by the Sandia Laboratory Directed Research and Development (LDRD) program. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories

National Nuclear Security Administration

U.S. DEPARTMENT OF ENERGY