

# **The Evolution of Gene Regulation in Bacteria**

Paramvir Dehal

Lawrence Berkeley National Lab

Bioinformatics 2008  
Warsaw, Poland  
April 22

# Questions

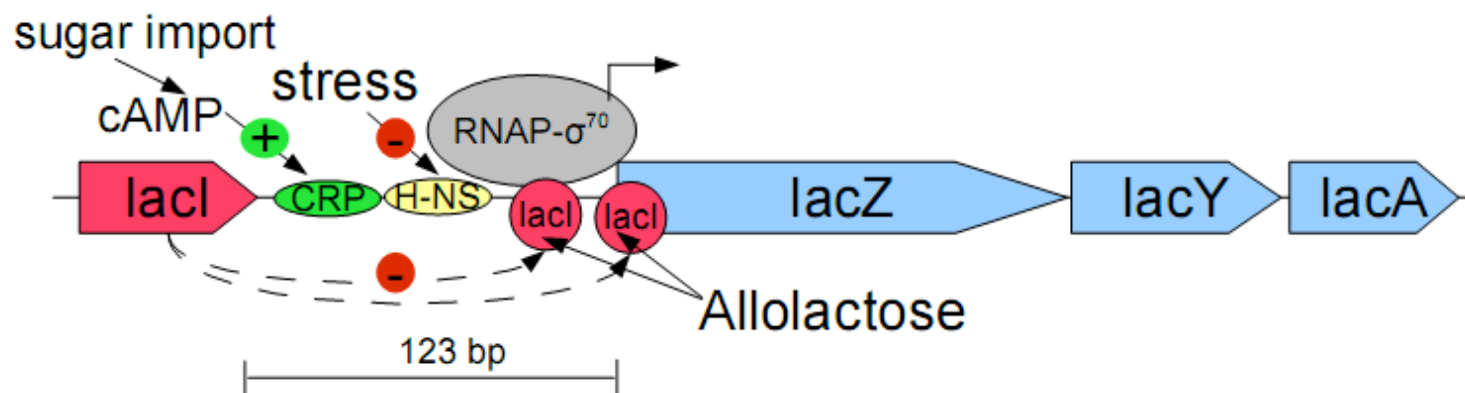
- How does gene regulation evolve?
  - Source of new transcription factors
    - Horizontal Gene Transfer (HGT)
    - Gene duplication
  - Regulation of new gene targets
    - HGT
    - Gene duplication
- How conserved is gene regulation?
- How conserved are regulatory interactions?

# Our Approach

- **GOAL:** Compare *E. coli* to its sequenced relatives
  - Where do TFs come from?
  - Where do regulatory sites come from?
  - How much regulation is conserved?
- **Solution:** Evolutionary histories of TFs
  - Vertical inheritance
  - Horizontal gene transfer (HGT)
  - Duplication

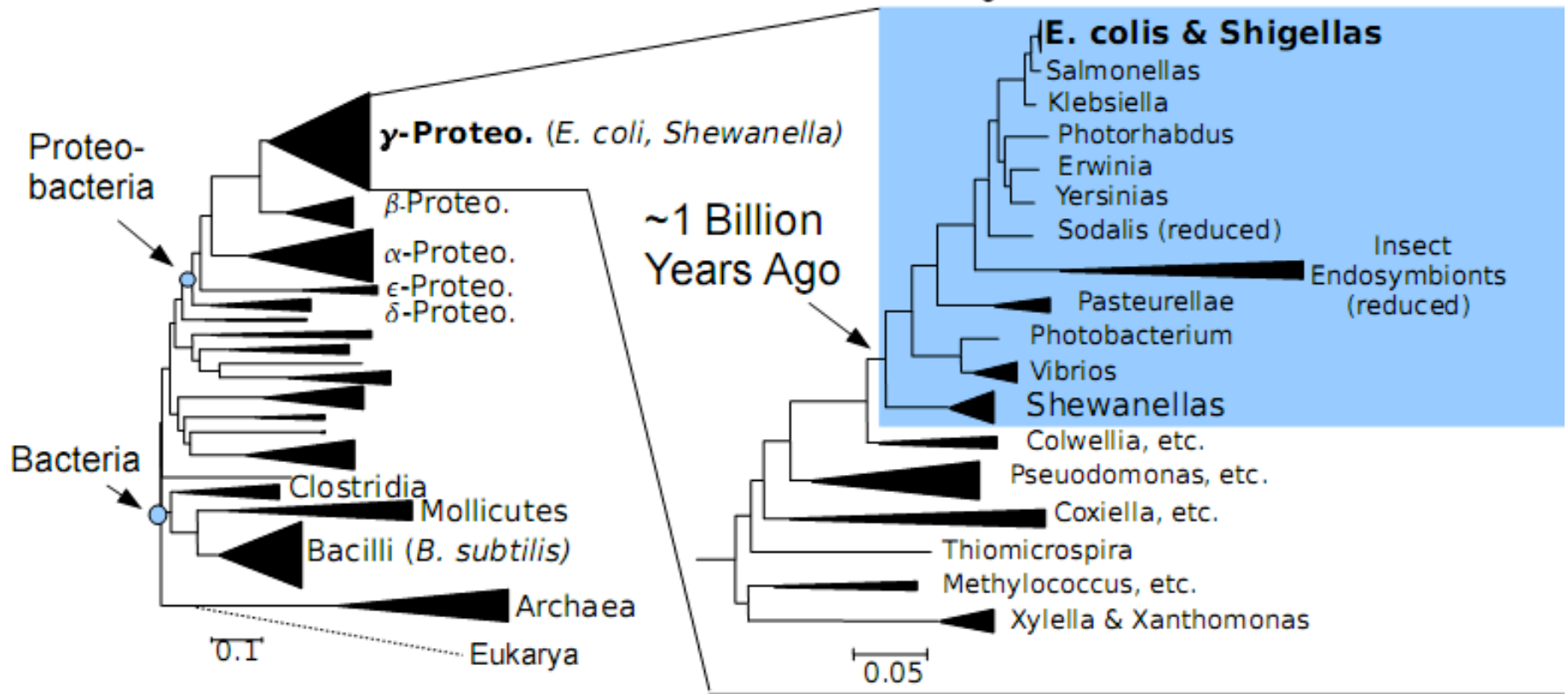
# Transcription Factors in *E. coli*

- ~250 TFs, >150 characterized (RegulonDB)
  - DNA binding & sensing domains, homodimers
- Large families
  - E.g. 45 LysR-like TFs with diverse functions
- “Neighbor regulators” vs. global TFs



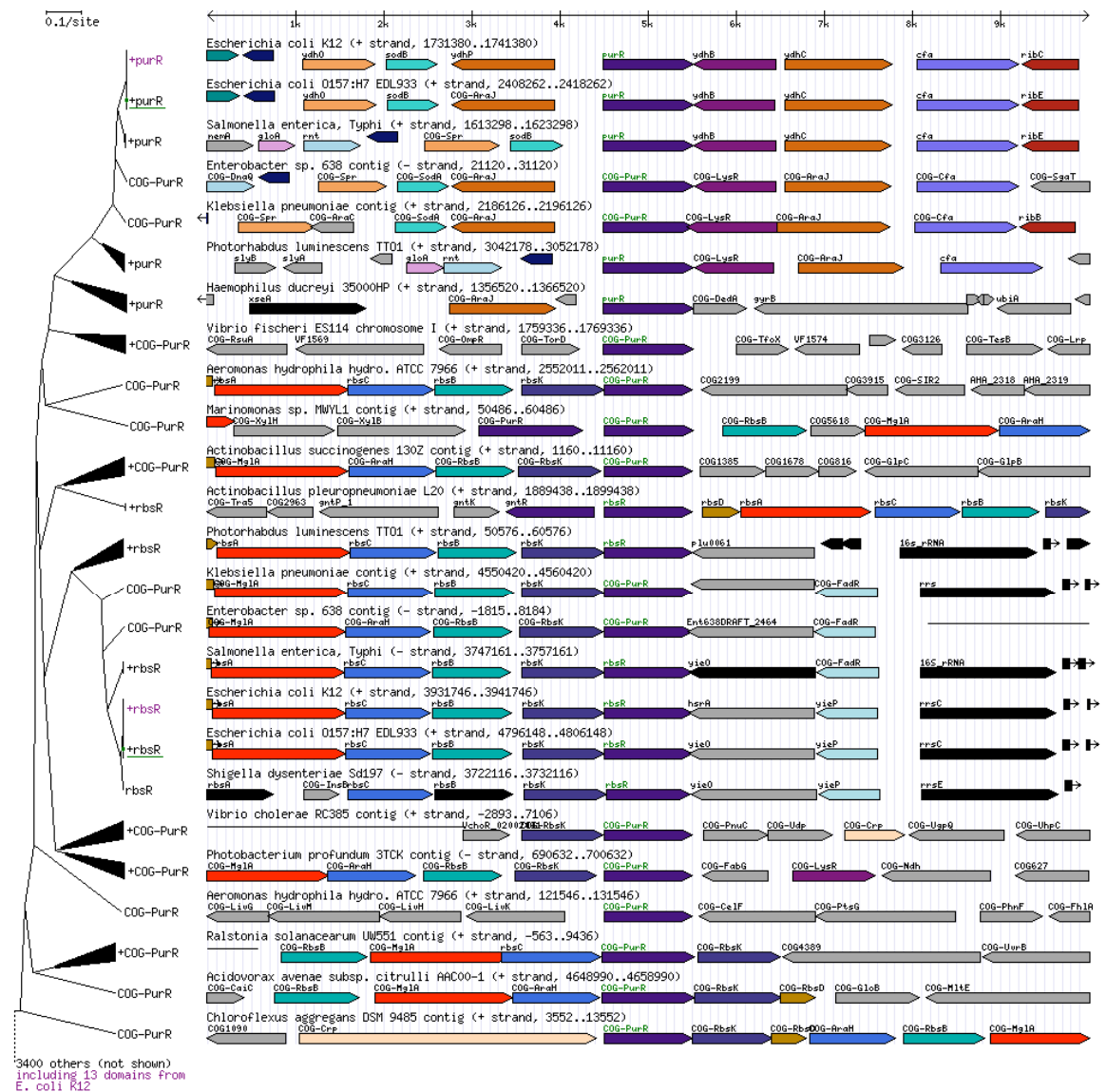
# The Tree of Life

## The $\gamma$ -Proteobacteria



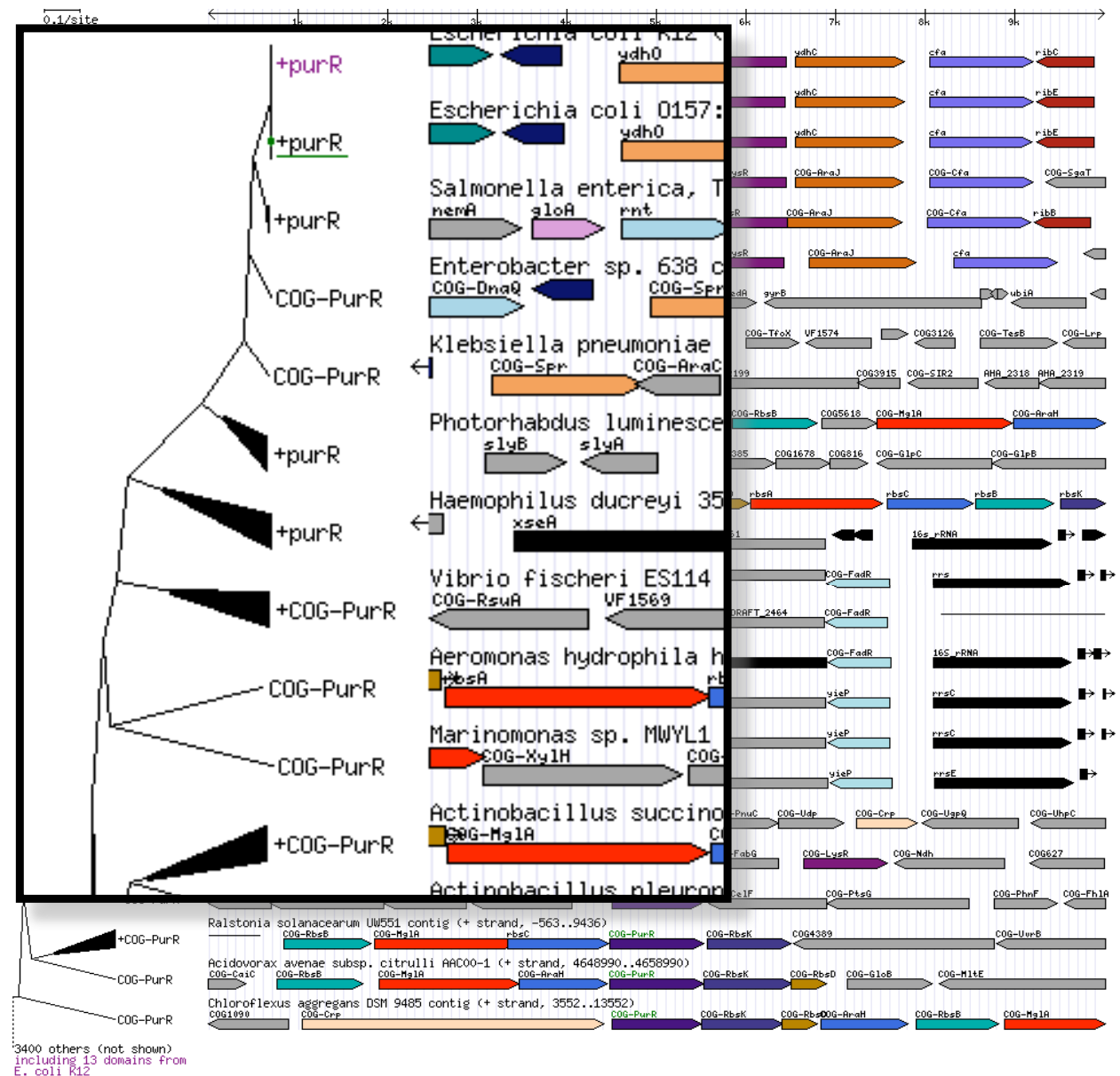
# MicrobesOnline TreeBrowser

- Phylogenetic trees for all gene families
- Quickly evaluate HGT and duplication events



# MicrobesOnline TreeBrowser

- Neighbor Joining Tree
- Branch lengths proportionate to rate of amino acid substitutions
- Bootstrap values on mouse over
- Expand and collapse interior nodes



# MicrobesOnline TreeBrowser

## Gene Tree

## Species Tree

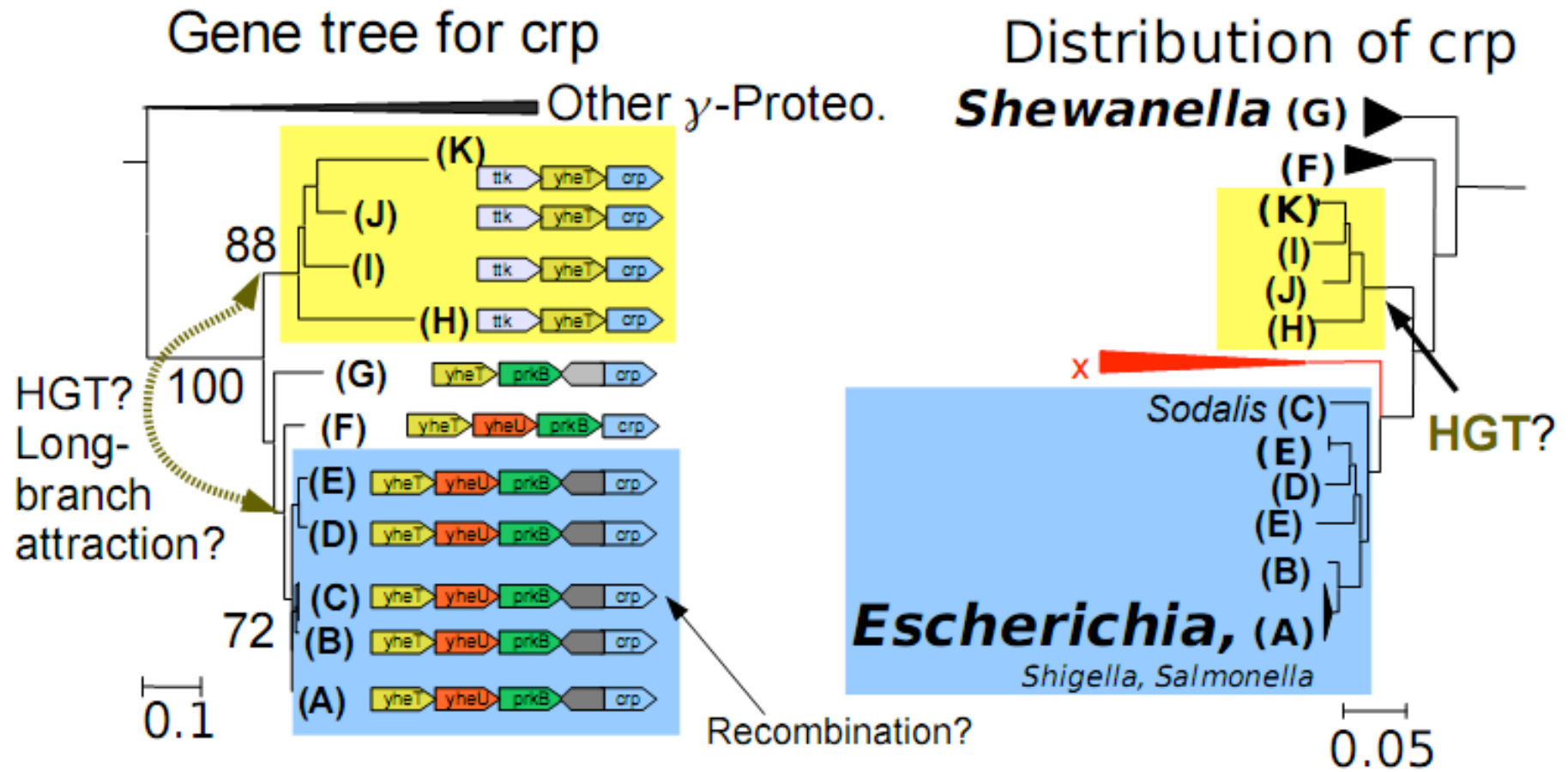




# Analyzing Gene Histories

- Tree-browser gives an overview
  - Hmalign & quicktree (not that accurate)
- More accurate trees are built
  - Realign with MUSCLE
  - Maximum likelihood trees (phym1)
  - Tools from [microbesonline.org](http://microbesonline.org) “gene carts”

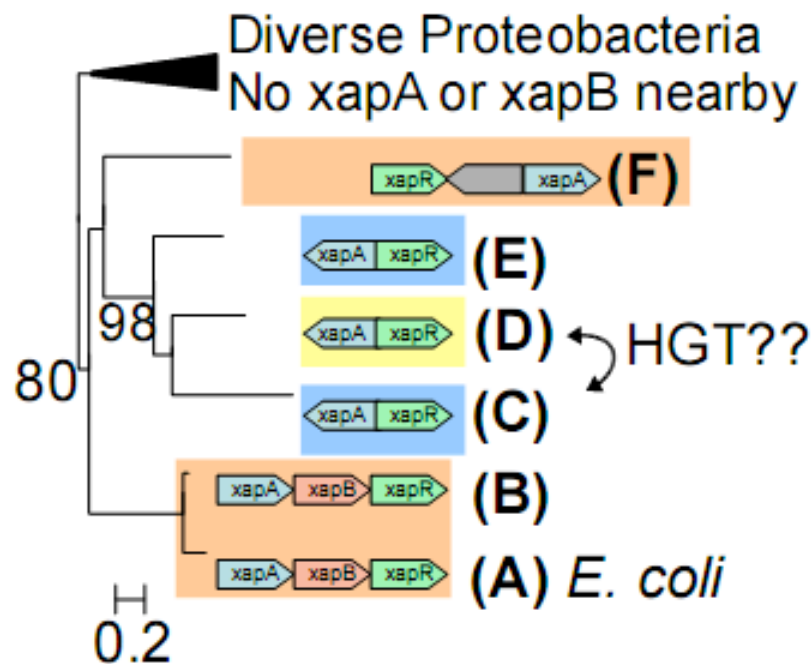
# Vertical Inheritance of Global Regulators



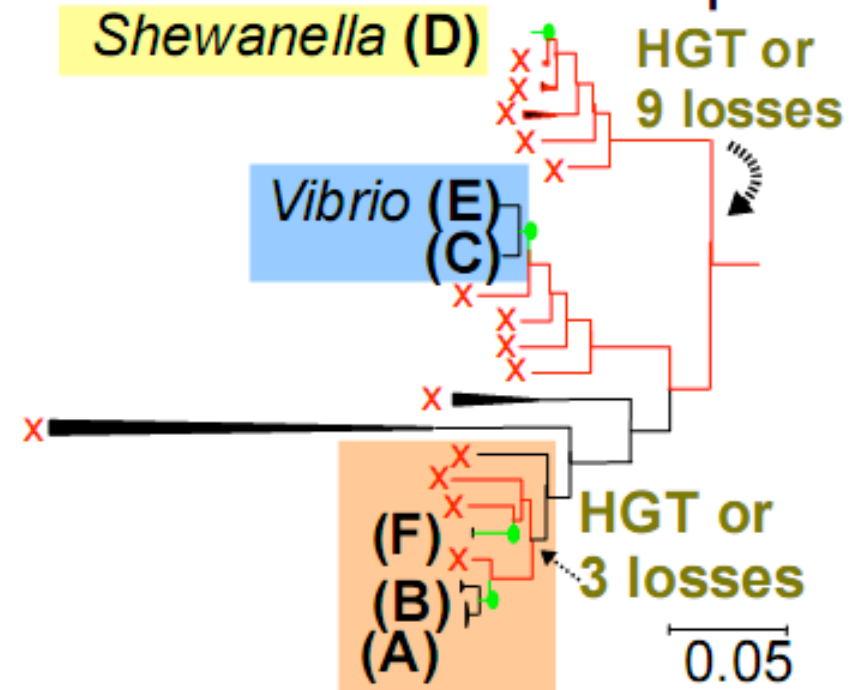
17/20 top global regulators are native

# Complex Histories of Neighbor Regulators

## Gene tree for xapR



## Distribution of xapR

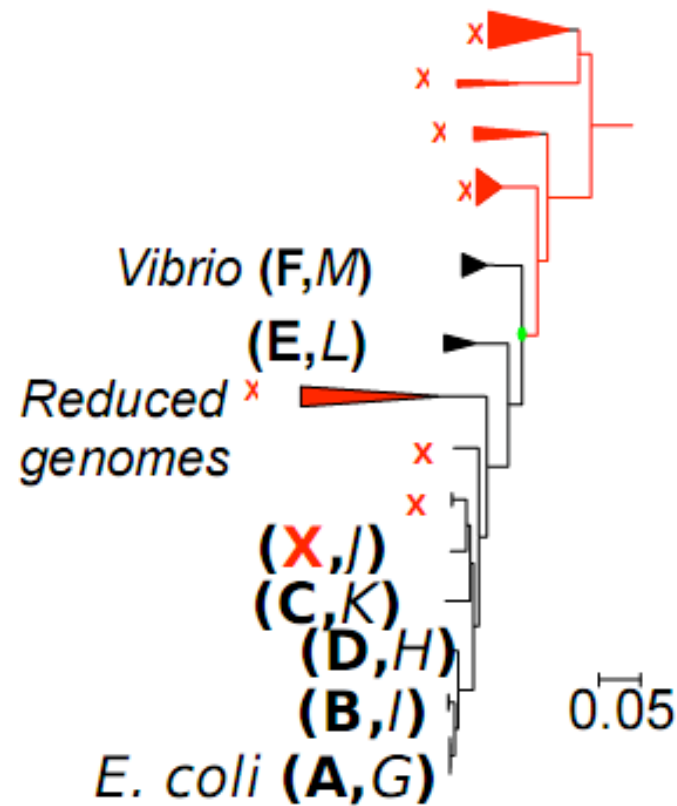
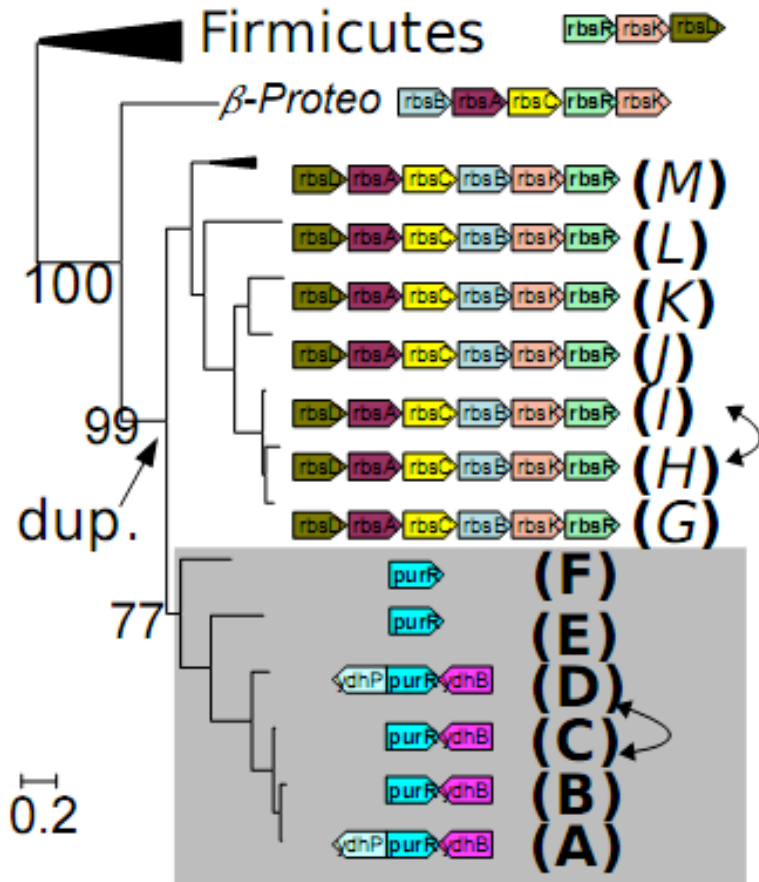


- Co-transfer: ~60% of neighbor regulators
- Repeated HGT within  $\gamma$ -Proteobacteria: ~40% of neighbor reg.

# Duplication of Transcription Factors

Gene tree for rbsR, purR

Distribution of rbsR, purR



- Duplications are rare (13% of TFs)
- Non-overlapping set of regulated genes (~half of dups)

# Putative Regulators

- 9/20 show co-transfer with adjacent genes
  - Most of the ~100 uncharacterized regulators are probable neighbor regulators
- Co-transfer predicts targets
  - 27/30 correct on characterized regulators

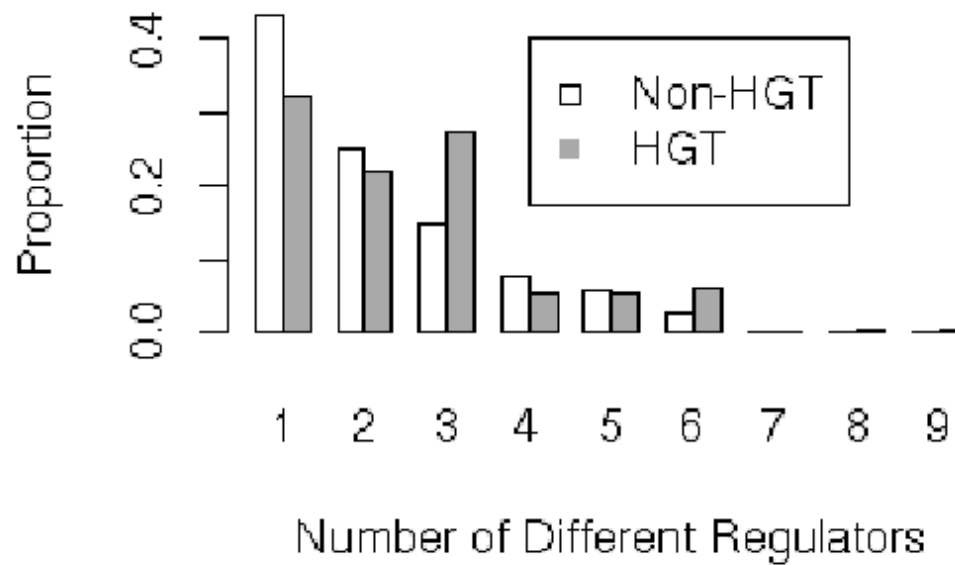
# Evolution of *E. coli* TFs (The Last Billion Years)

- Little duplication (13%)
  - Duplicates often have different functions
- Lots of HGT (63%)
  - Most “paralogs” diverged in different genomes
    - “allopatric gene divergence” to different functions
  - Co-transfer is common
    - Uncharacterized TFs are probably neighbor regulators
    - Neighbor regulation selected for by HGT?
  - Multiple transfers within  $\gamma$ -Proteobacteria
- But global regulators are native & conserved!

# Histories of Regulatory Interactions

- TFs often gained/lost, but 2/3rds of known regulation in *E. coli* is by top-20 global regulators
  - Do their regulons evolve quickly?
  - How are HGT genes regulated?
- Evolution by duplication?
  - Not common for TFs, but what about regulatory interactions?

# Complex Regulation of HGT Genes



E.g., CRP regulates half of HGT genes but only a quarter of other genes.

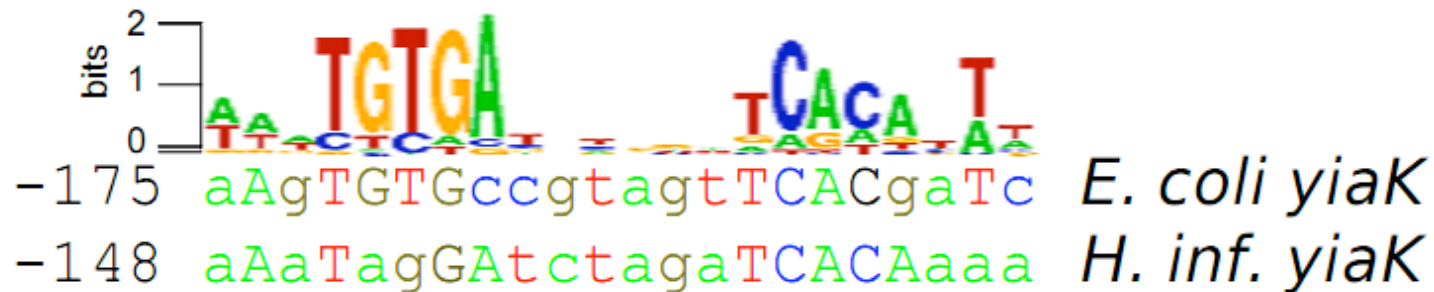


# Sources of the CRP Regulon

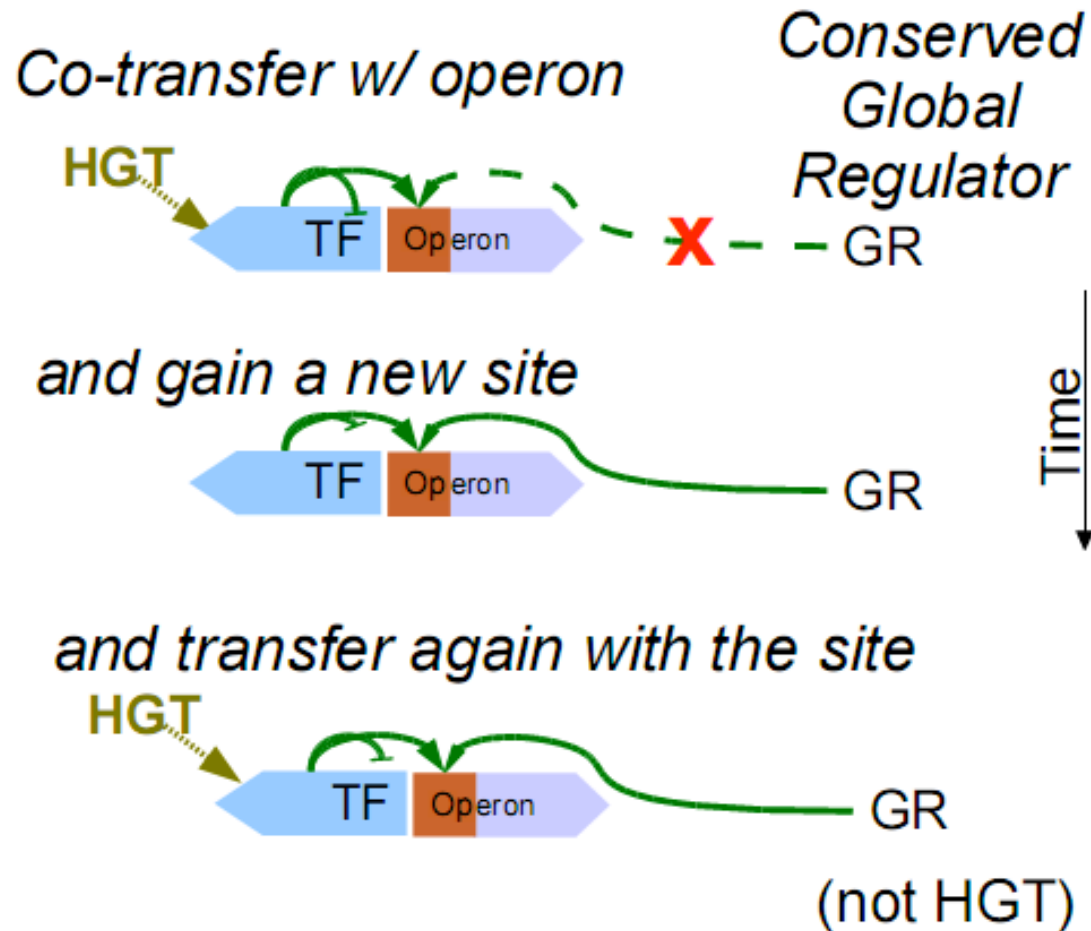
- CRP has orthologs only in  $\beta, \gamma$ -Proteobacteria
  - No functional orthologs known in distant bacteria
  - Other functions for this family: fnr, dnr, etc.
- ~80% of CRP-regulated genes acquired from distant bacteria w/o CRP
- ~20% from  $\beta, \gamma$ -Proteobacteria

# Regulation is Not Conserved Across HGT Events

- CRP site is conserved across HGT in only 4/12 cases for HGT within  $\gamma$ Proteobacteria
- For other global regulators, 6/20 sites conserved across HGT events

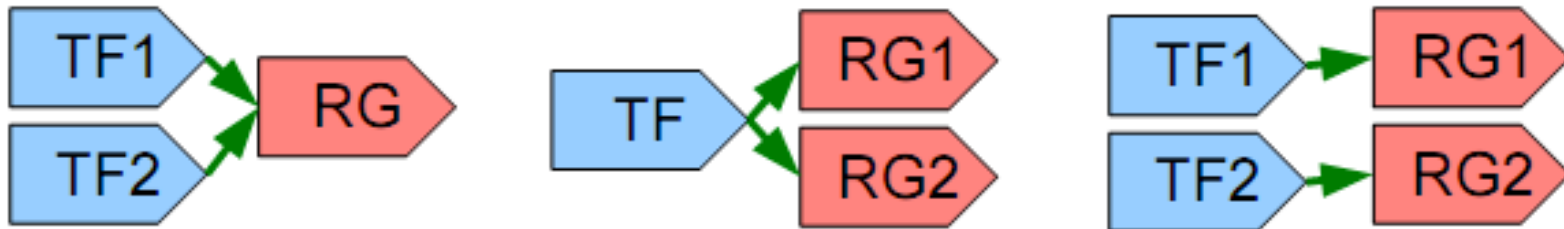


# Niche-Specific Neighbor Regulators vs. Conserved Global Regulators

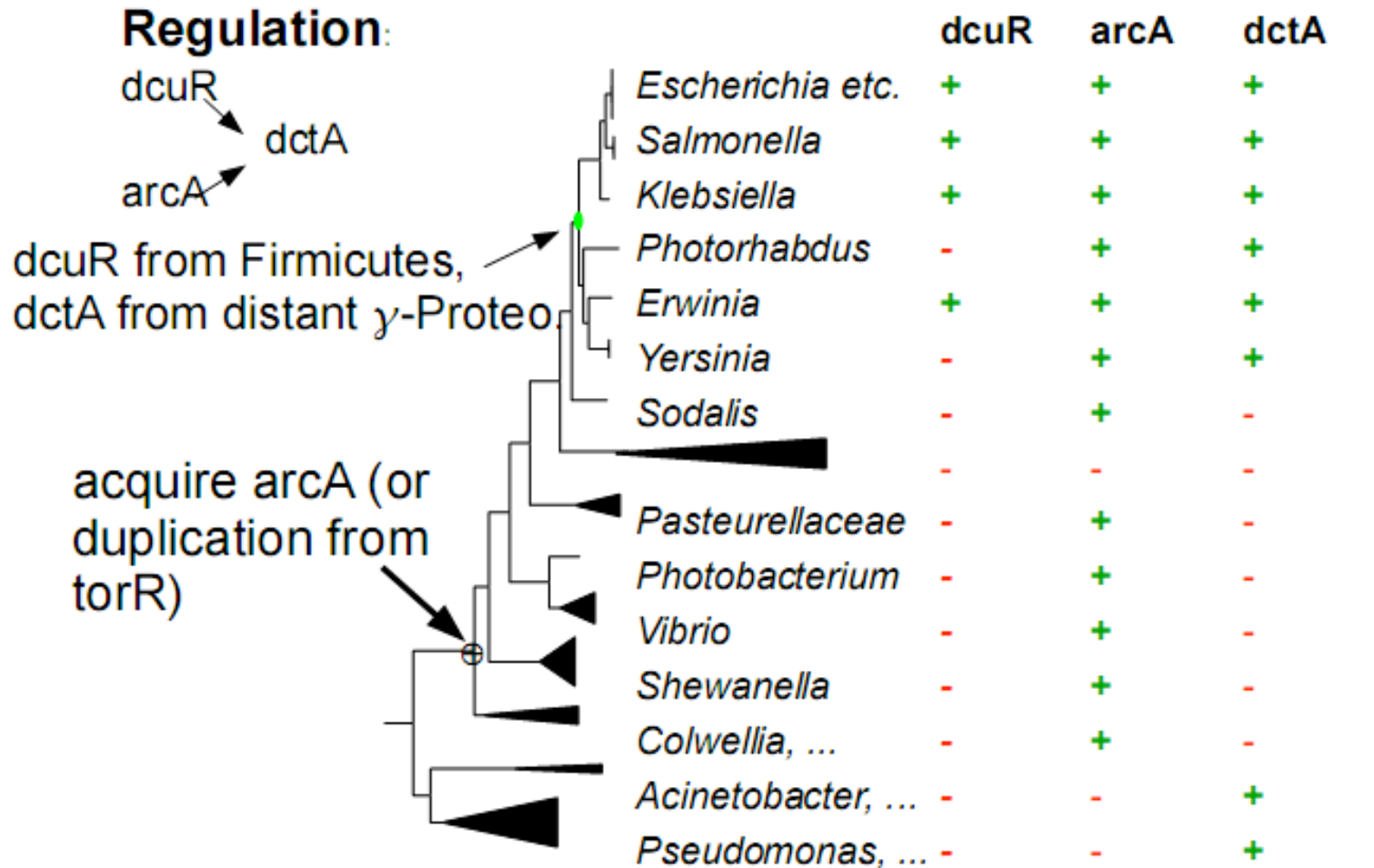


# Evolution of Regulatory Interactions by Duplication?

Three Scenarios:



# Age of Regulation vs. Age of Duplication

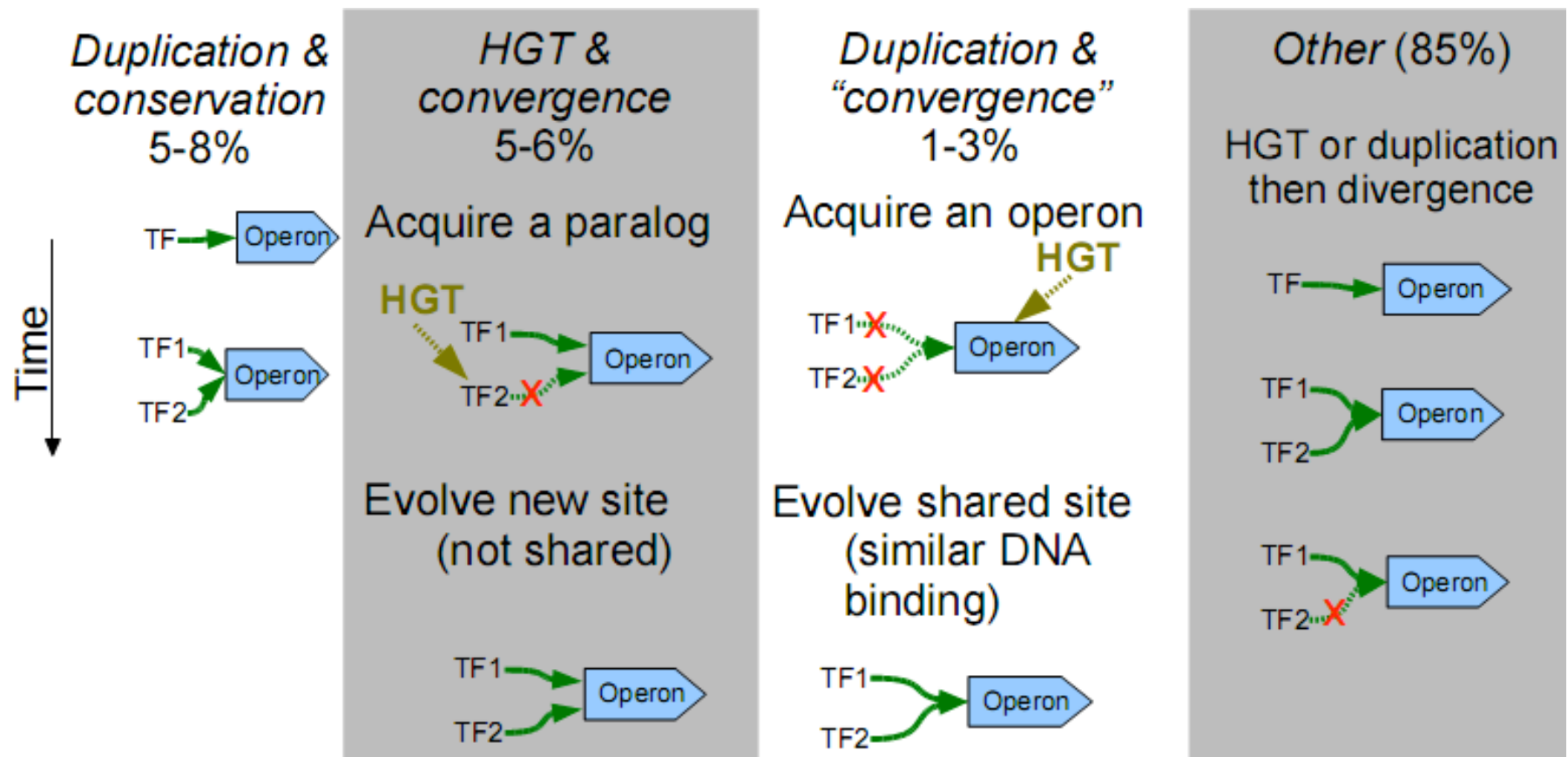


dcuR and dctA were not in the same organism until long after the divergence of dcuR from arcA

# Convergent Evolution of Regulation

- Not conserved from common ancestor
  - Regulation cannot be older than presence of RG & TF in the lineage
  - Regulation of *dctA* evolved after transfer
- Shared interactions are 14% of RegulonDB
  - 5% duplication
  - 6% convergent evolution (>chance,  $P < 0.001$ )
  - 3% unclear

# Summary: Origin of Regulatory Interactions



# Conclusion

- Rapid Evolution of Gene Regulation
- Rampant transfer, not duplication, of TFs
  - “Orthologs” are misleading
  - Annotations are usually incorrect
- “Neighbor regulation” evolves by HGT
  - Co-transfer predicts target of TFs
- Global regulators are much more conserved



# More Information

Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*.

Genome Biology 2008, 9:R4

Orthologous transcription factors in bacteria have different functions and regulate different genes.

PLoS Comput. Biol. 2007 3:1739-50.

Morgan Price, Paramvir Dehal, and Adam Arkin

# Acknowledgements

Morgan Price

Adam Arkin

Eric Alm

Inna Dubchak

Michael Cipriano

MicrobesOnline.org

Dylan Chivian

Marcin Joachimiak

Wayne Huang

Keith Keller

Katherine Huang

Funded by DOE's

Genomics:GTL

<http://MicrobesOnline.org>

# FastTree

- Sub  $N^2$  implementation of Neighbor Joining using Profiles
- ~10,000x faster than traditional approaches with similar accuracy
- 160k 16S tree in 17hrs with support values

<http://MicrobesOnline.org/fasttree>