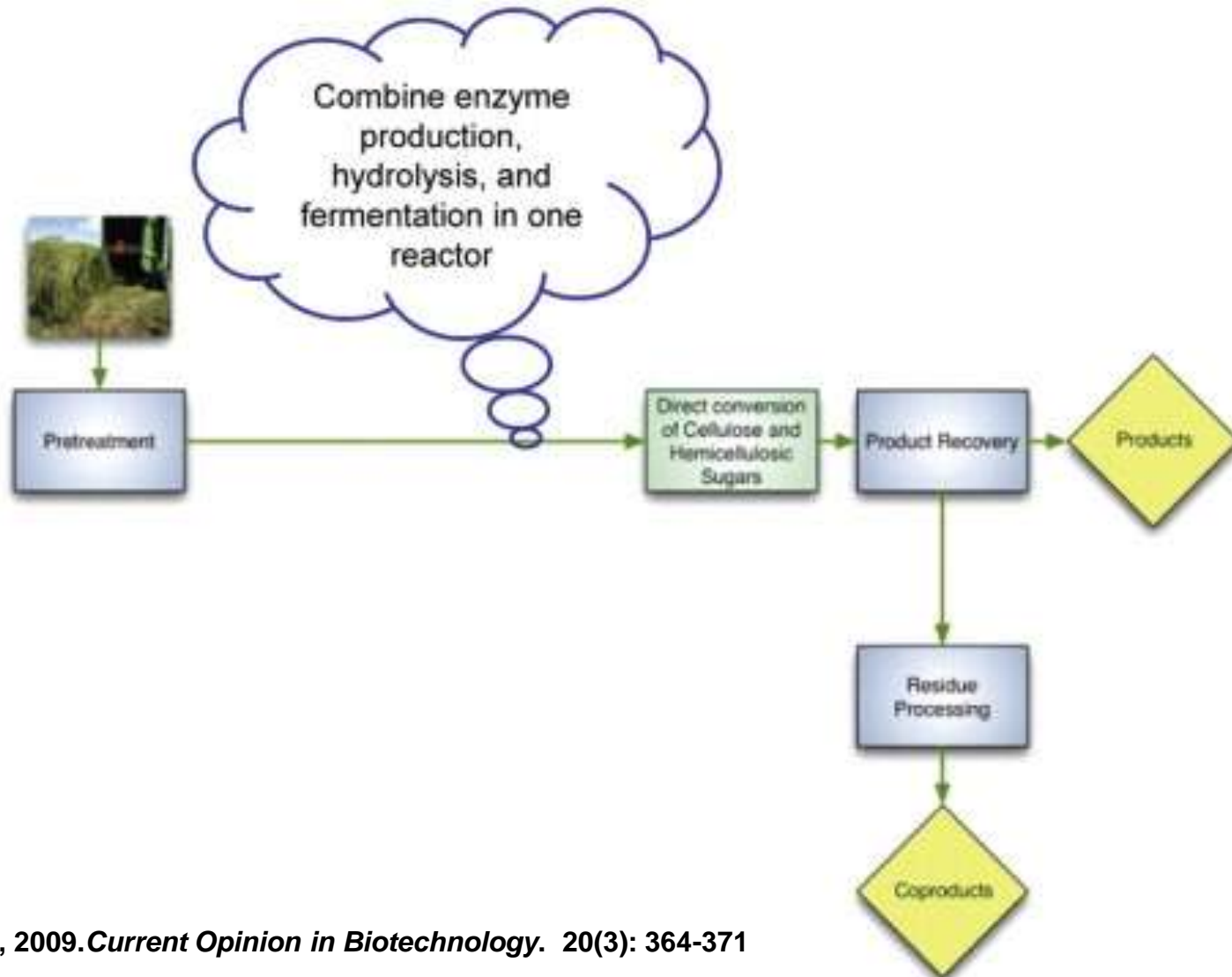# The Structure and Evolution of the Cellulosome

Chris Hemme, Huang Yue,
Yong-Jin Lee and Jizhong Zhou

# Consolidated bioprocessing



Combine enzyme production, hydrolysis, and fermentation in one reactor

Pretreatment → Direct conversion of Cellulose and Hemicellulosic Sugars → Product Recovery → Products

Product Recovery → Residue Processing → Coproducts

Xu, *et al.*, 2009.*Current Opinion in Biotechnology.* 20(3): 364-371

# Microbial strategies for enzymatic biomass degradation

- ## Free Enzymes
  - Secreted into environment
- ## Cellulosomes
  - Multienzyme outer membrane complex
  - Direct binding to biomass
  - Modular
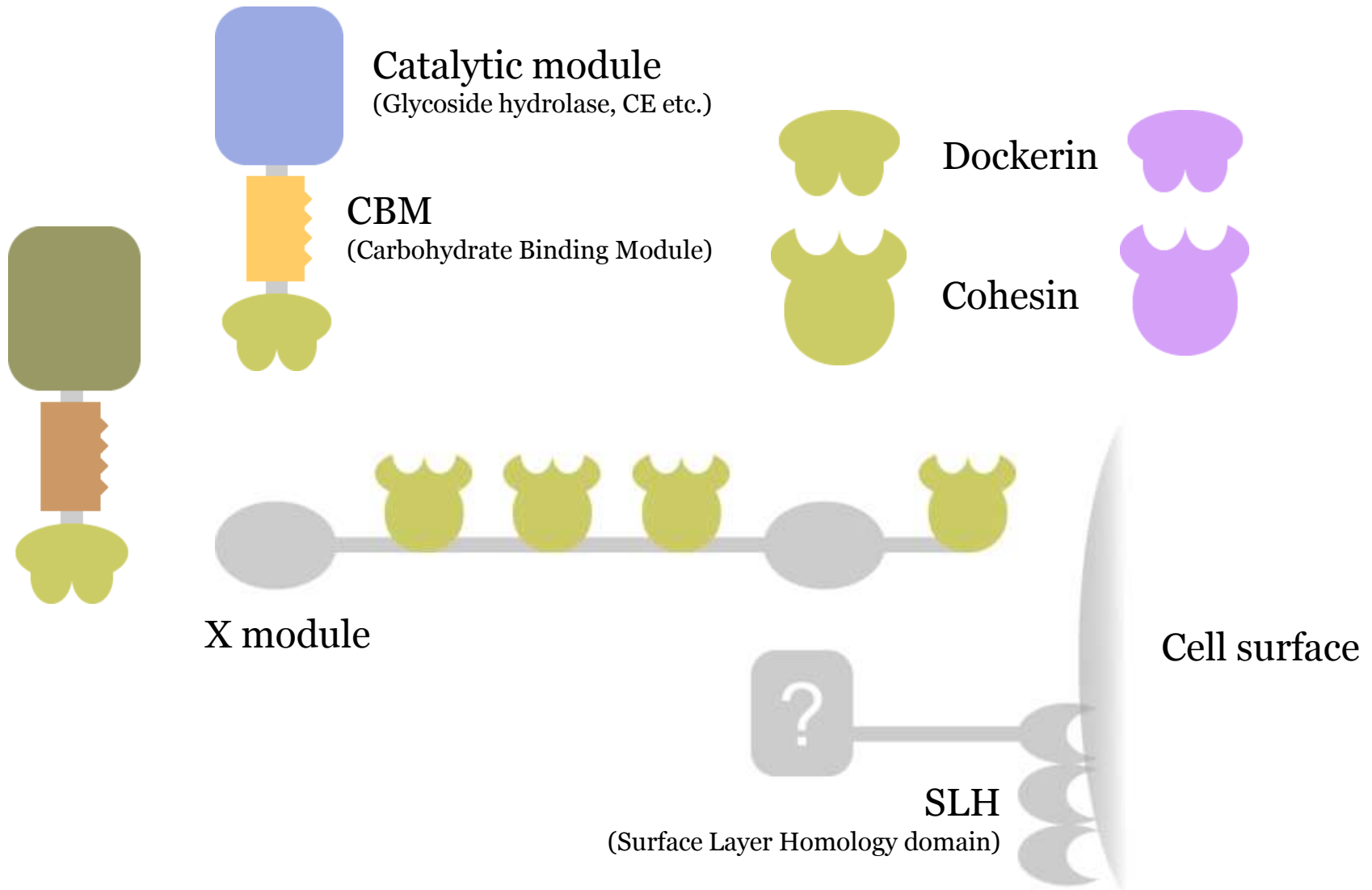
# Cellulosome-encoding Clostridia

1. Clostridium acetobutylicum ATCC 824
2. Clostridium cellulovorans 743B

3. Clostridium papyrosolvens DSM 2782
4. Clostridium cellulolyticum H10
5. Clostridium thermocellum ATCC 27405
6. Clostridium thermocellum JW20
7. Clostridium thermocellum LQR1
8. Acetivibrio cellulolyticus CD2

9. Ruminococcus albus 7
10. Ruminococcus albus 8
11. Ruminococcus sp. 18P13
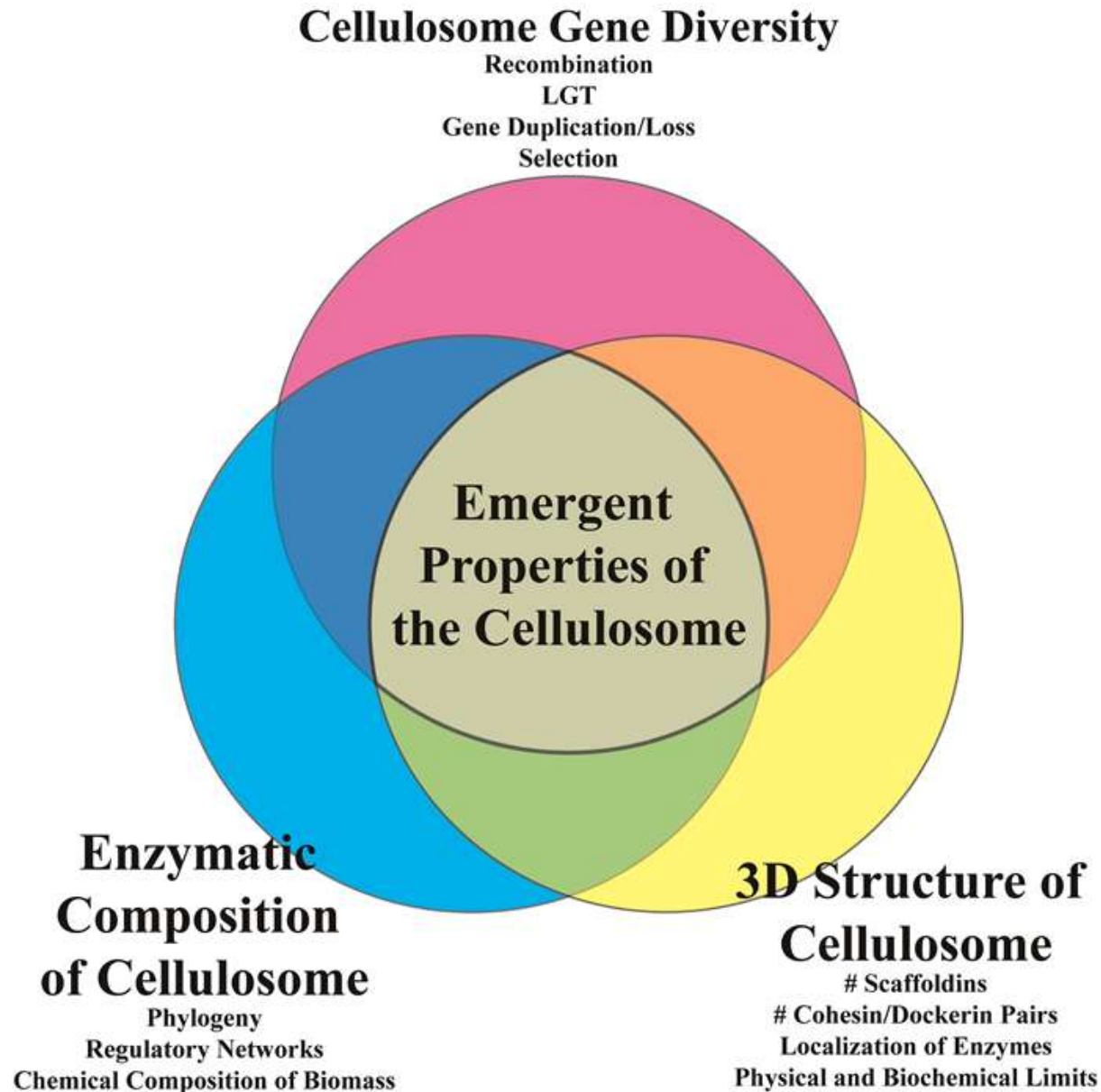12. Ruminococcus flavefaciens FD-1

Clostridium lineage

Acetivibrio-like
Clostridium lineage

Ruminococcus lineage

# The Cellulosome

Catalytic module
(Glycoside hydrolase, CE etc.)

CBM
(Carbohydrate Binding Module)

Dockerin

Cohesin

X module

Cell surface
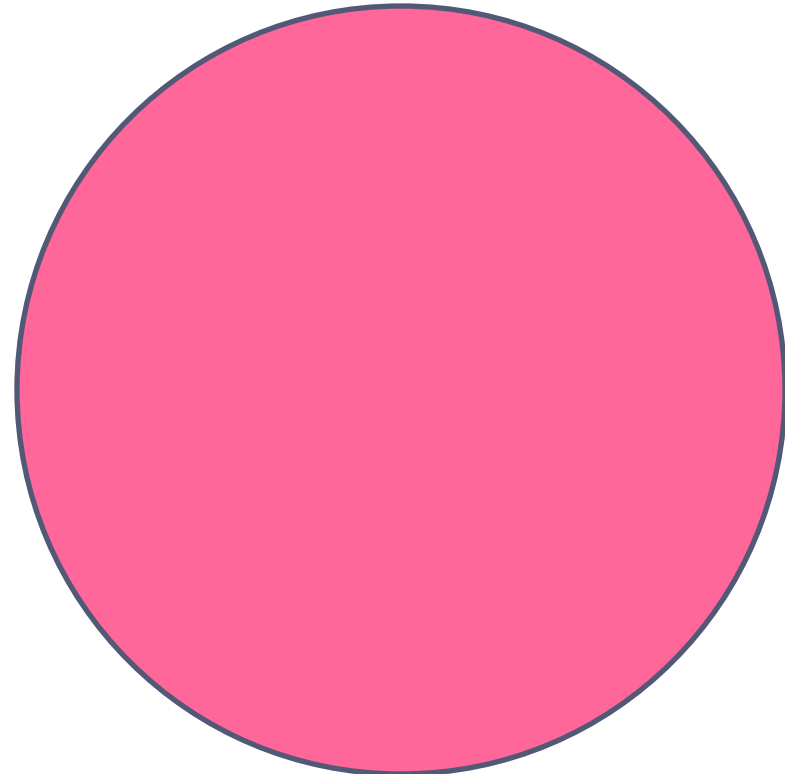
?

SLH
(Surface Layer Homology domain)

# Emergent properties of the cellulosome

# Cellulosome gene diversity

- Diversity of functional gene categories

- Diversity within a functional gene category

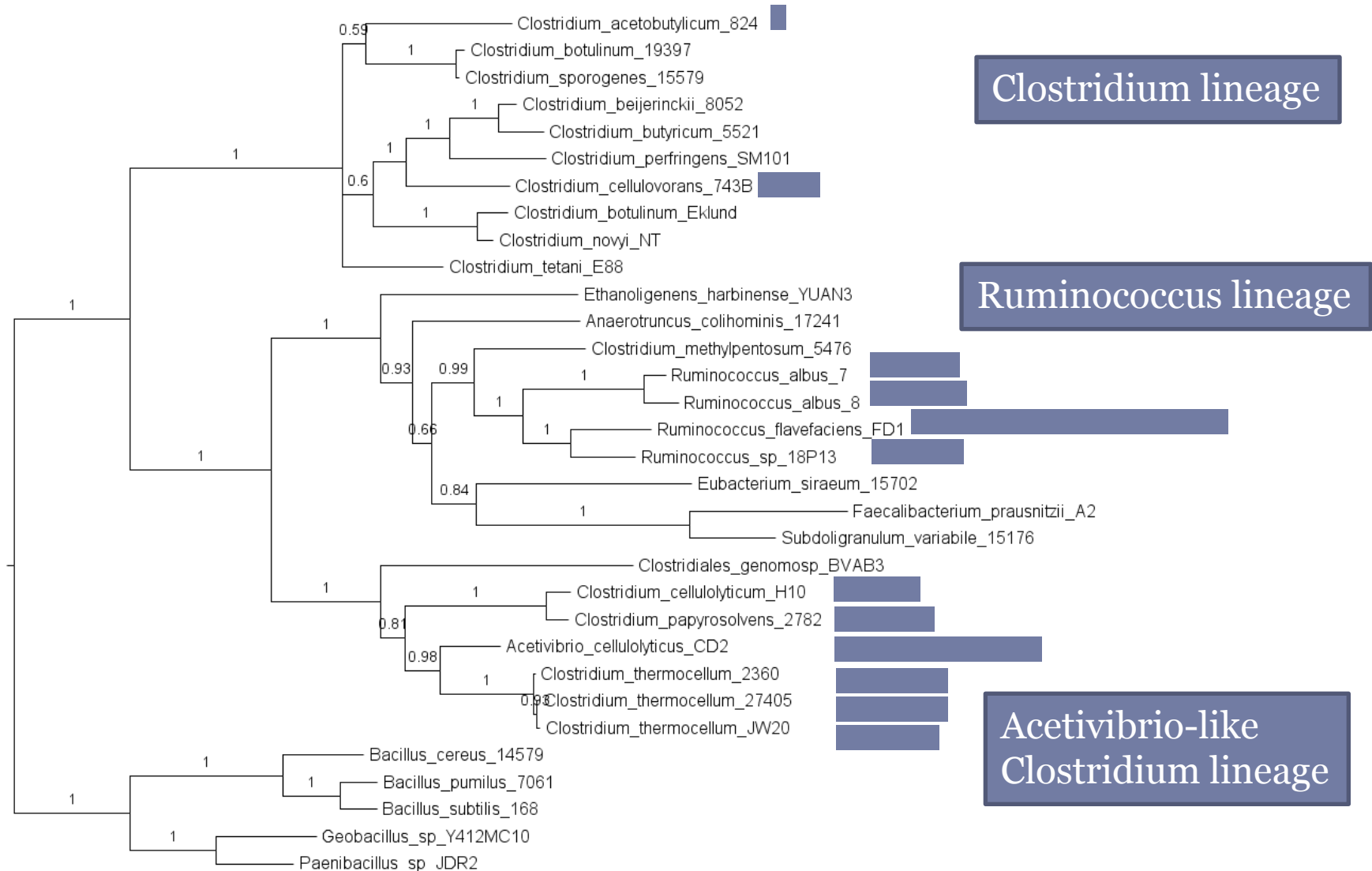- Evolutionary history within a functional gene category

# Cellulosome-related genes

| | Dockerin-containing genes | Cohesin-containing genes | Total unique | Protein-coding genes | Percentage |
|---|---|---|---|---|---|
| *C. acetobutylicum* ATCC 824 | 10 | 2 | 12 | 3847 | 0.3 % |
| *C. cellulovorans* 743B | 41 | 4 | 45 | 4254 | 1.1 % |
| *C. papyrosolvens* DSM 2782 | 66 | 5 | 70 | 4423 | 1.6 % |
| *C. Cellulolyticum* H10 | 57 | 2 | 59 | 3390 | 1.7 % |
| *C. thermocellum* ATCC LQR1 | 74 | 11 | 84 | 3091 | 2.7 % |
| *C. thermocellum* JW20 | 68 | 8 | 76 | 3076 | 2.5 % |
| *C. thermocellum* ATCC 27405 | 74 | 8 | 82 | 3189 | 2.6 % |
| ***Acetivibrio*** | **137** | **16** | **150** | **5144** | **2.9 %** |
| *R. albus* 7 | 59 | 1 | 59 | 4157 | 1.4 % |
| *R. albus* 8 | 62 | 0 | 62 | 3872 | 1.6 % |
| *Ruminococcus* sp. 18P13 | 61 | 6 | 63 | 2114 | 3.0 % |
| ***R.Flavifaciens*** | **209** | **9** | **212** | **4162** | **5.1 %** |

# Phylogenetic distribution of the cellulosome



Clostridium lineage

Ruminococcus lineage

Acetivibrio-like Clostridium lineage
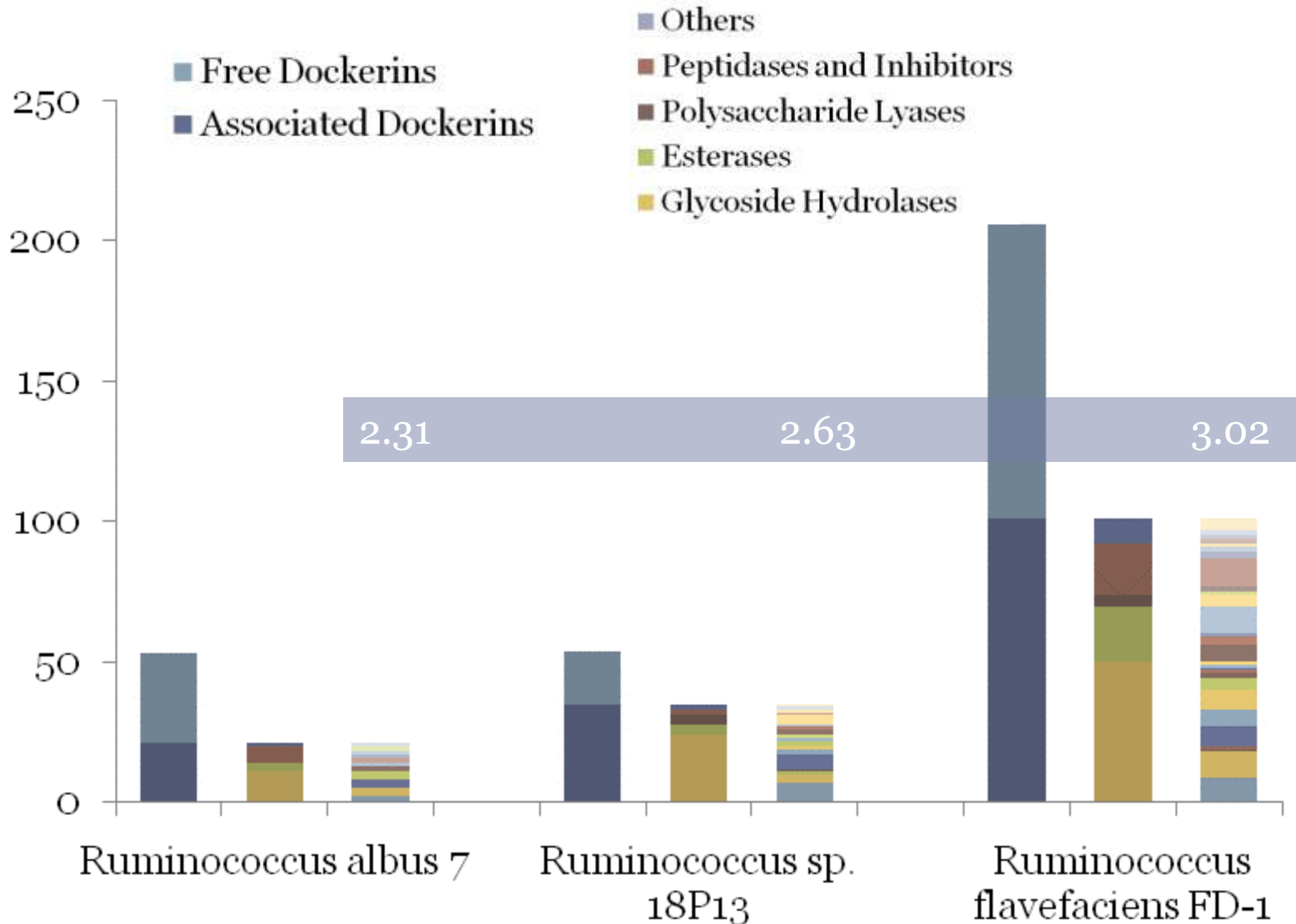
16S Phylogeny with posterior probability
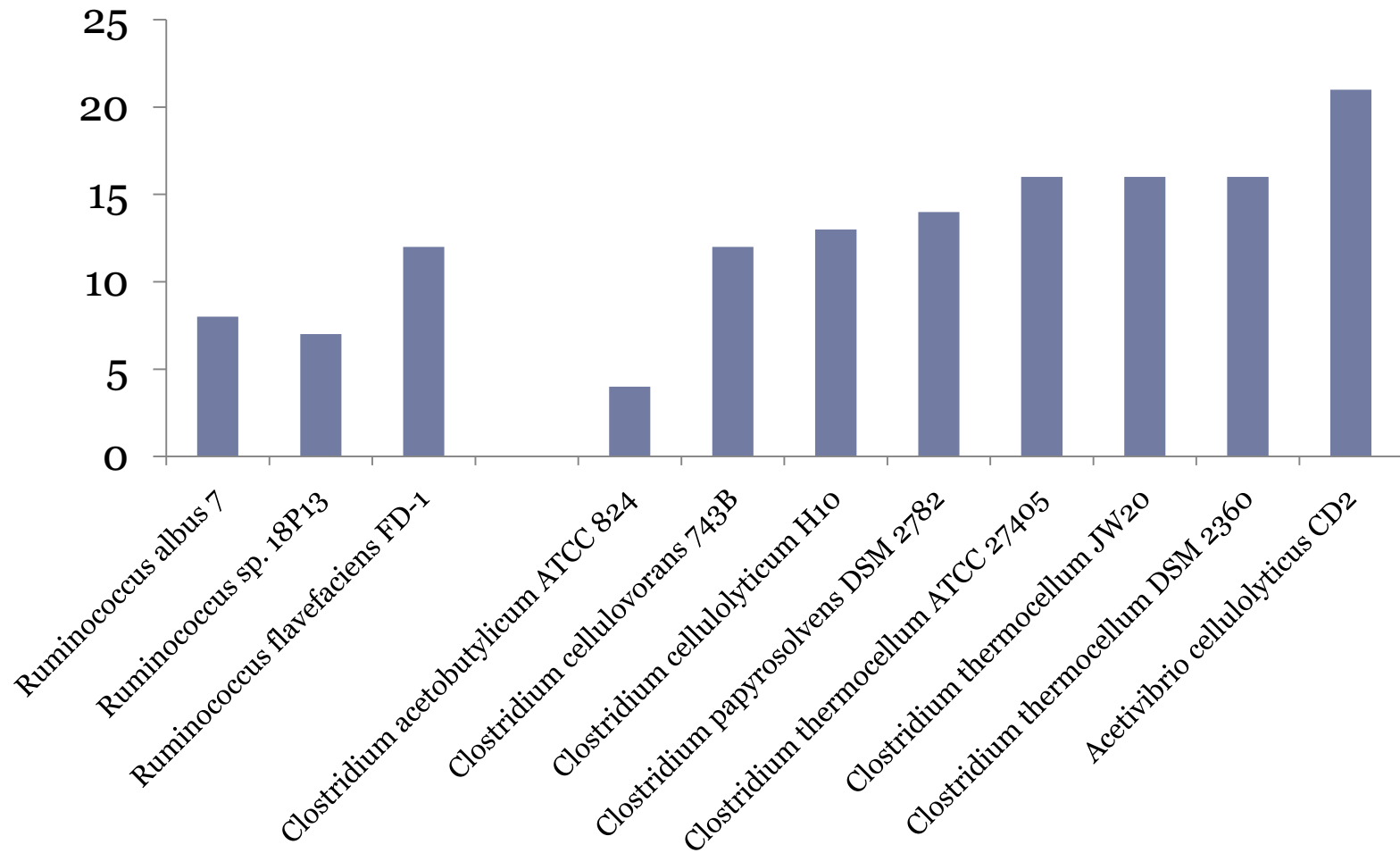
# Diversity of functional gene categories
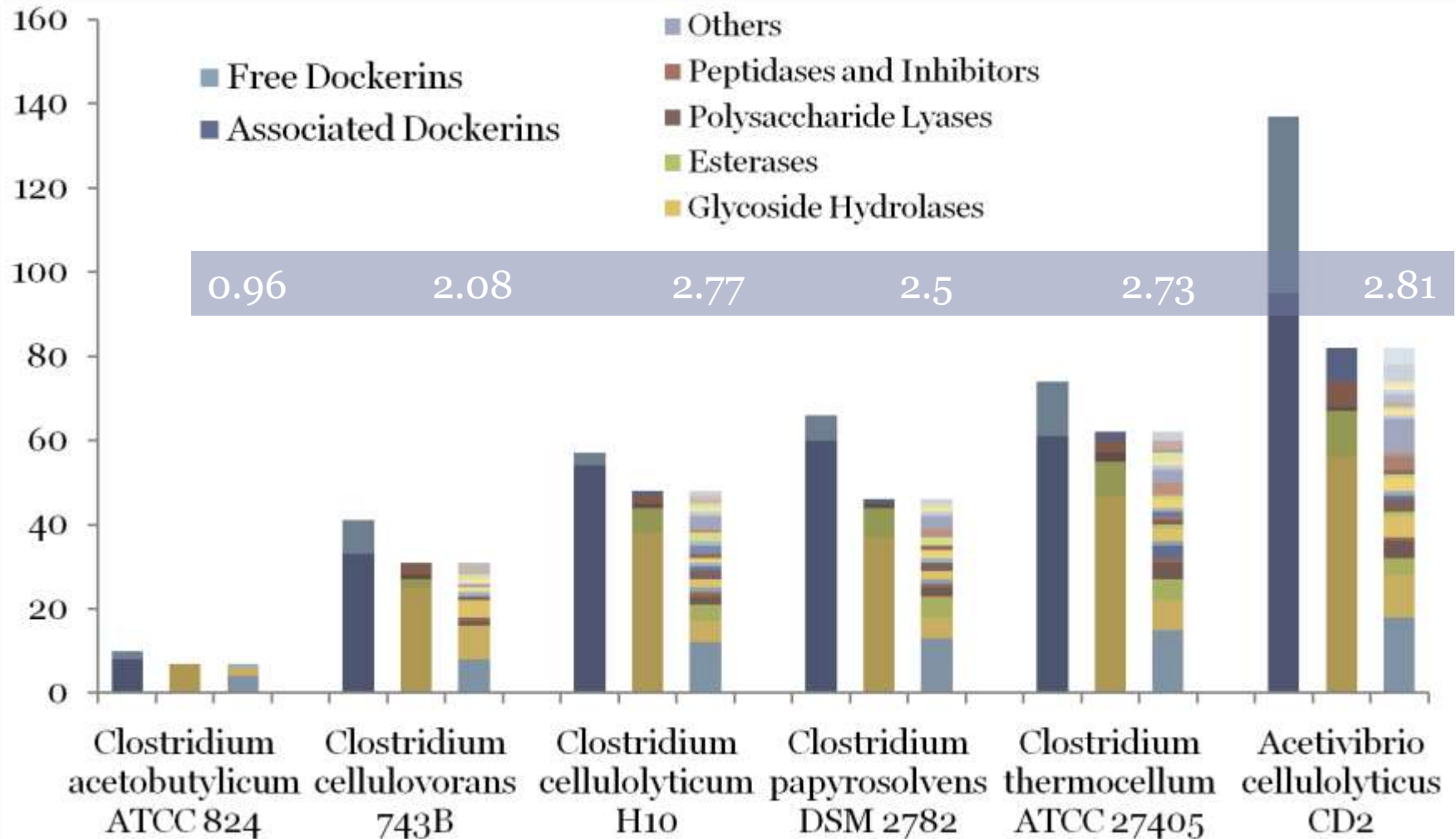


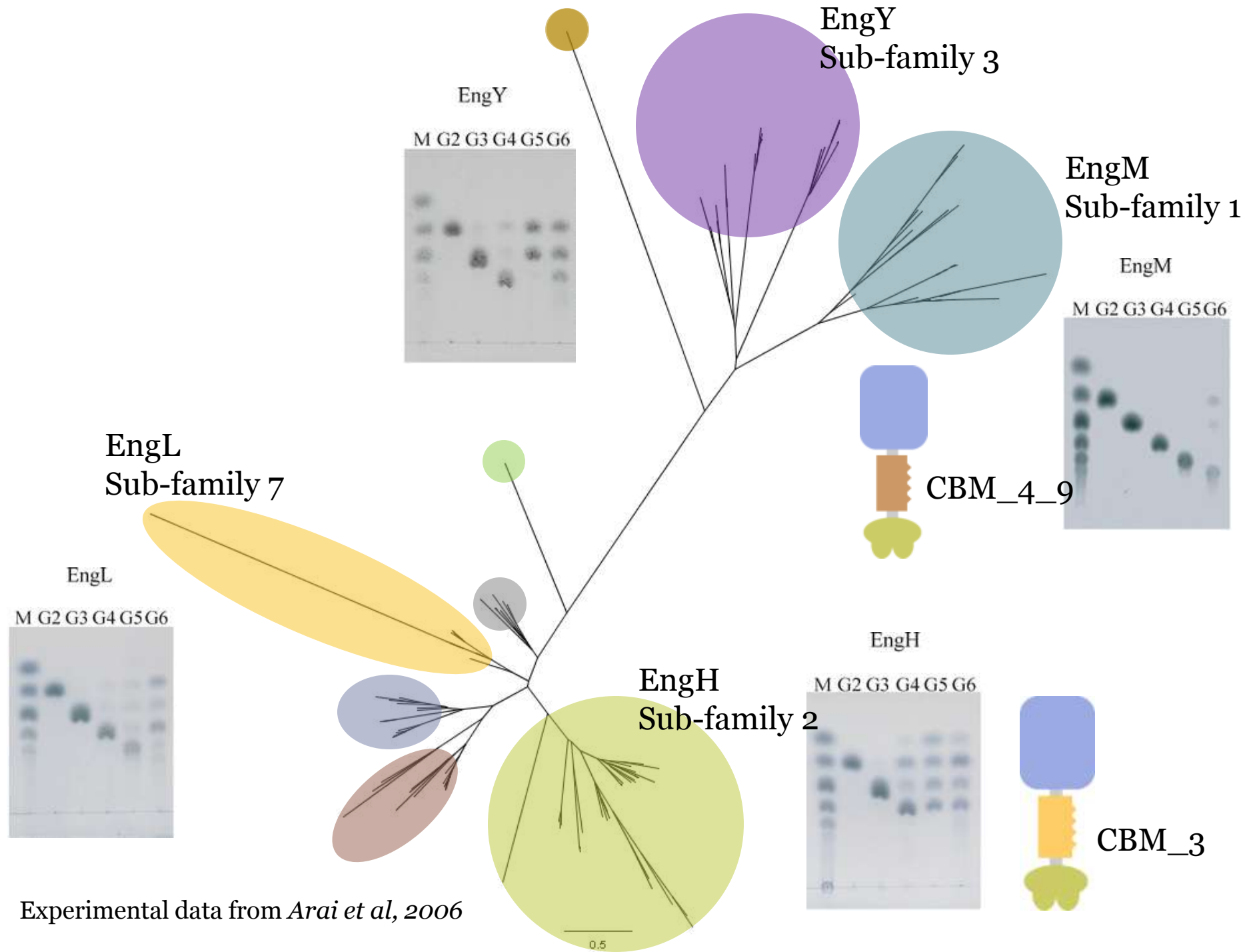Shannon diversity index = -sum($p_{i*} \ln p_i$)

# Diversity of functional gene categories

# Diversity within a functional gene category

## Glycoside Hydrolase family 9
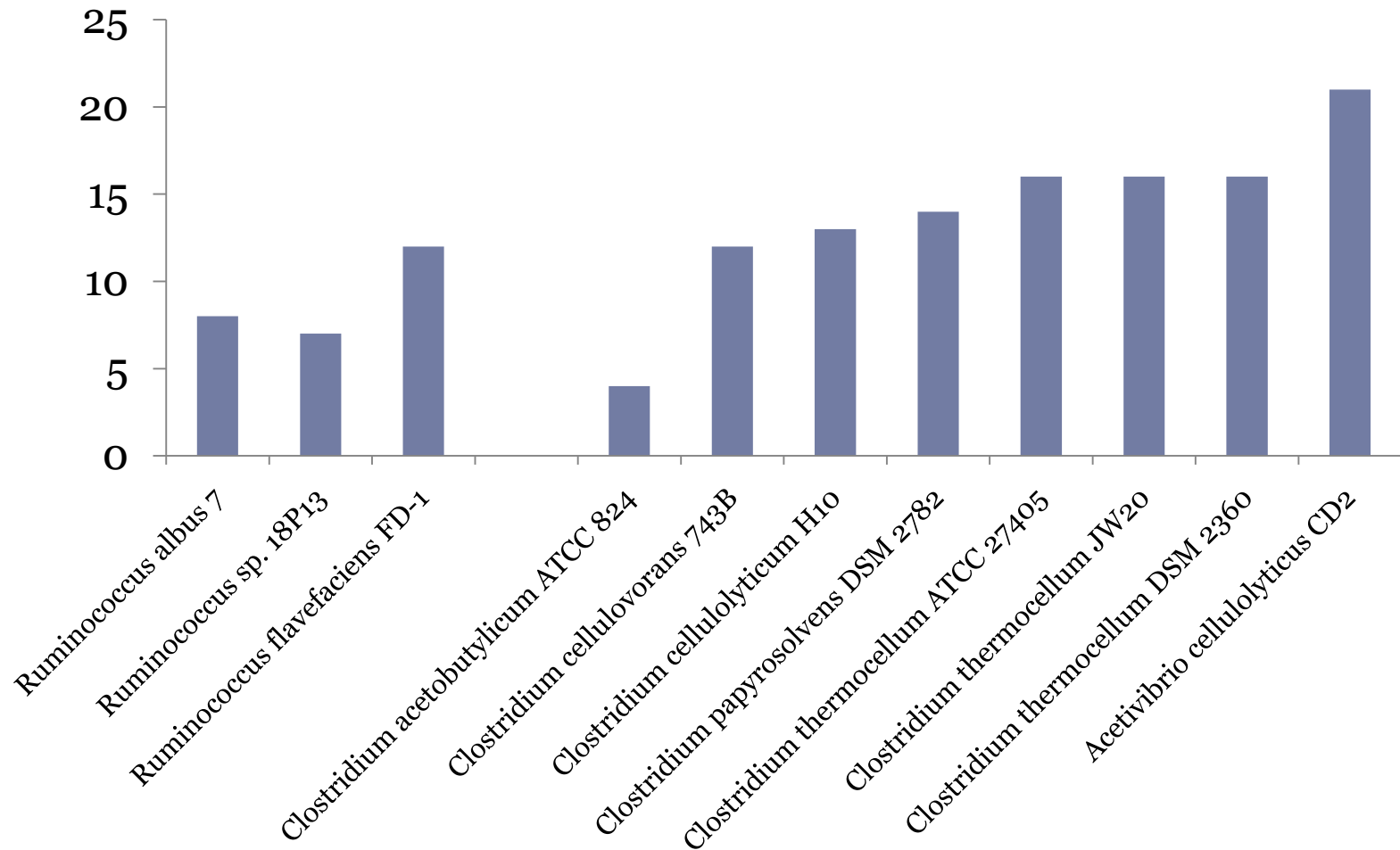
EngY
Sub-family 3

EngM
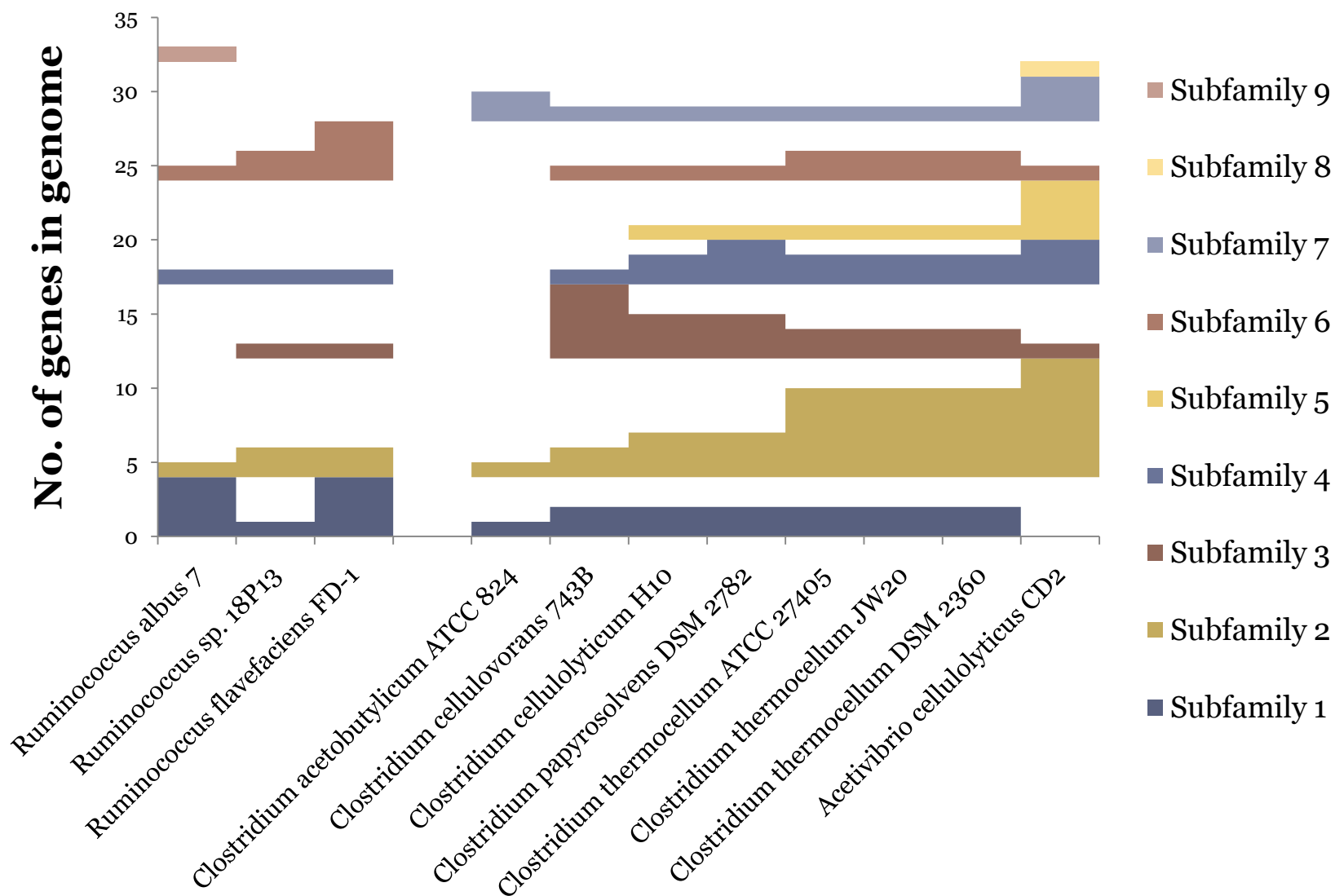Sub-family 1

EngY

M G2 G3 G4 G5 G6

EngM

M G2 G3 G4 G5 G6

CBM_4_9

EngL
Sub-family 7

EngL

M G2 G3 G4 G5 G6

EngH
Sub-family 2

EngH

M G2 G3 G4 G5 G6

CBM_3

Experimental data from *Arai et al, 2006*

0.5

gi|256003097|Clostridium_thermocellum_DSM_2360
gi|125975242|Clostridium_thermocellum_ATCC_27405
gi|281419233|Clostridium_thermocellum_JW20
gi|303241877|Acetivibrio_cellulolyticus_CD2
gi|220929070|Clostridium_cellulolyticum_H10
gi|256755594|Clostridium_papyrosolvens_DSM_2782
gi|256755593|Clostridium_papyrosolvens_DSM_2782
gi|304401130|Ruminococcus_albus_7
gi|268608866|Ruminococcus_flavefaciens_FD-1
gi|302873829|Clostridium_cellulovorans_743B
gi|220928185|Clostridium_cellulolyticum_H10
gi|256756507|Clostridium_papyrosolvens_DSM_2782
gi|303239557|Acetivibrio_cellulolyticus_CD2
gi|303238713|Acetivibrio_cellulolyticus_CD2
gi|291543938|Ruminococcus_sp._18P13
gi|125974845|Clostridium_thermocellum_ATCC_27405
gi|256003898|Clostridium_thermocellum_DSM_2360
gi|281418700|Clostridium_thermocellum_JW20
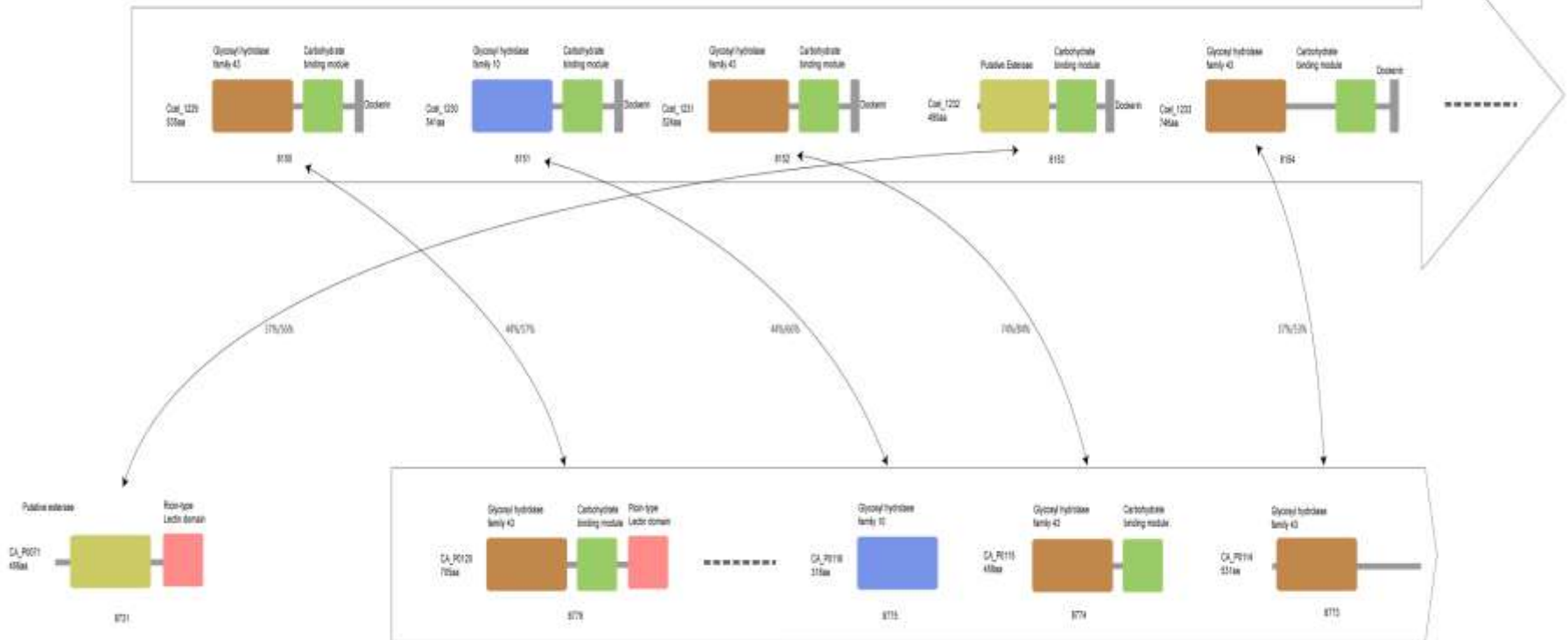
0.5

Glycoside Hydrolase family 9
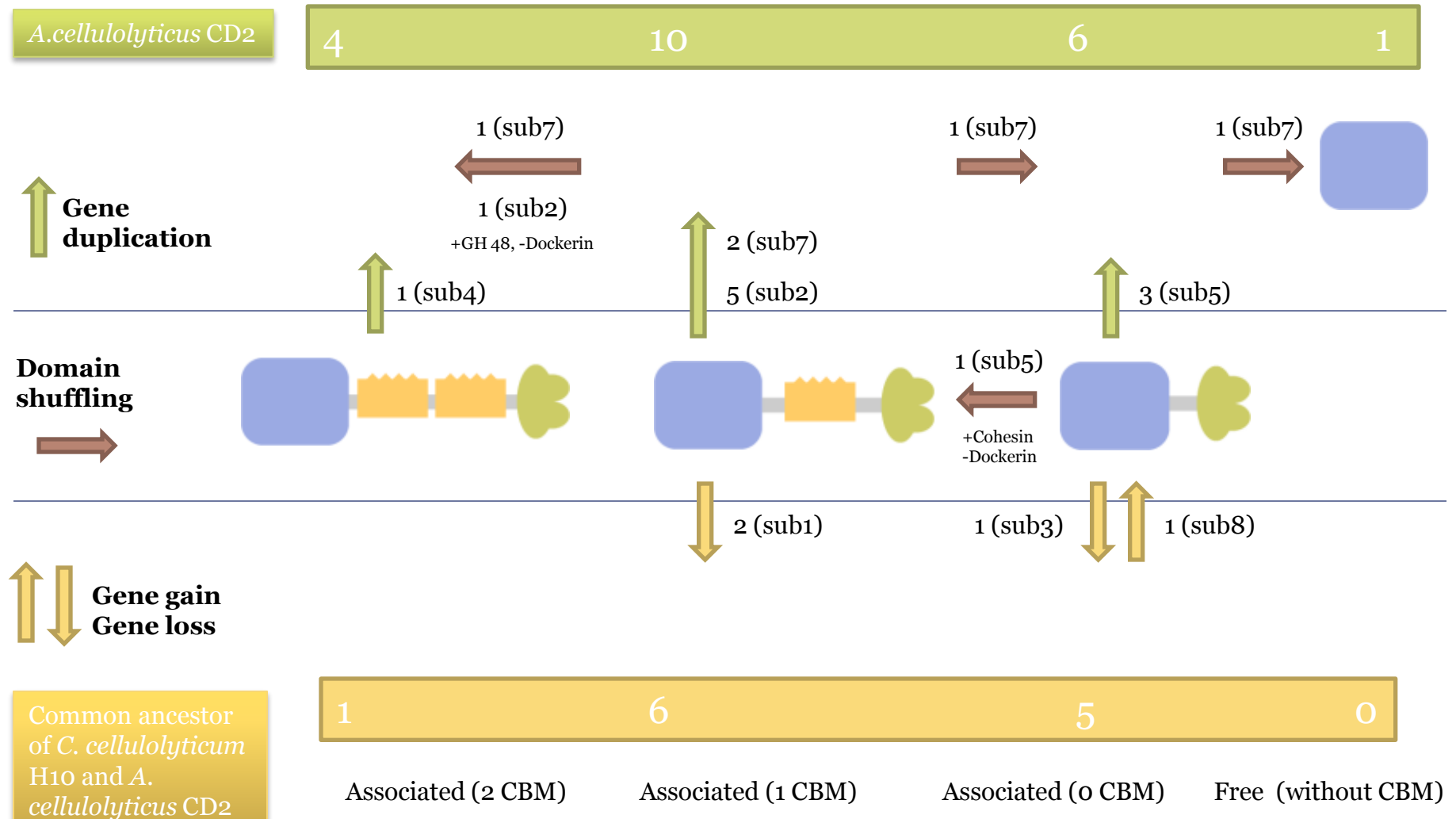
# Orthologs and paralogs of dockerin-containing genes

# Conservation of function but not gene structure
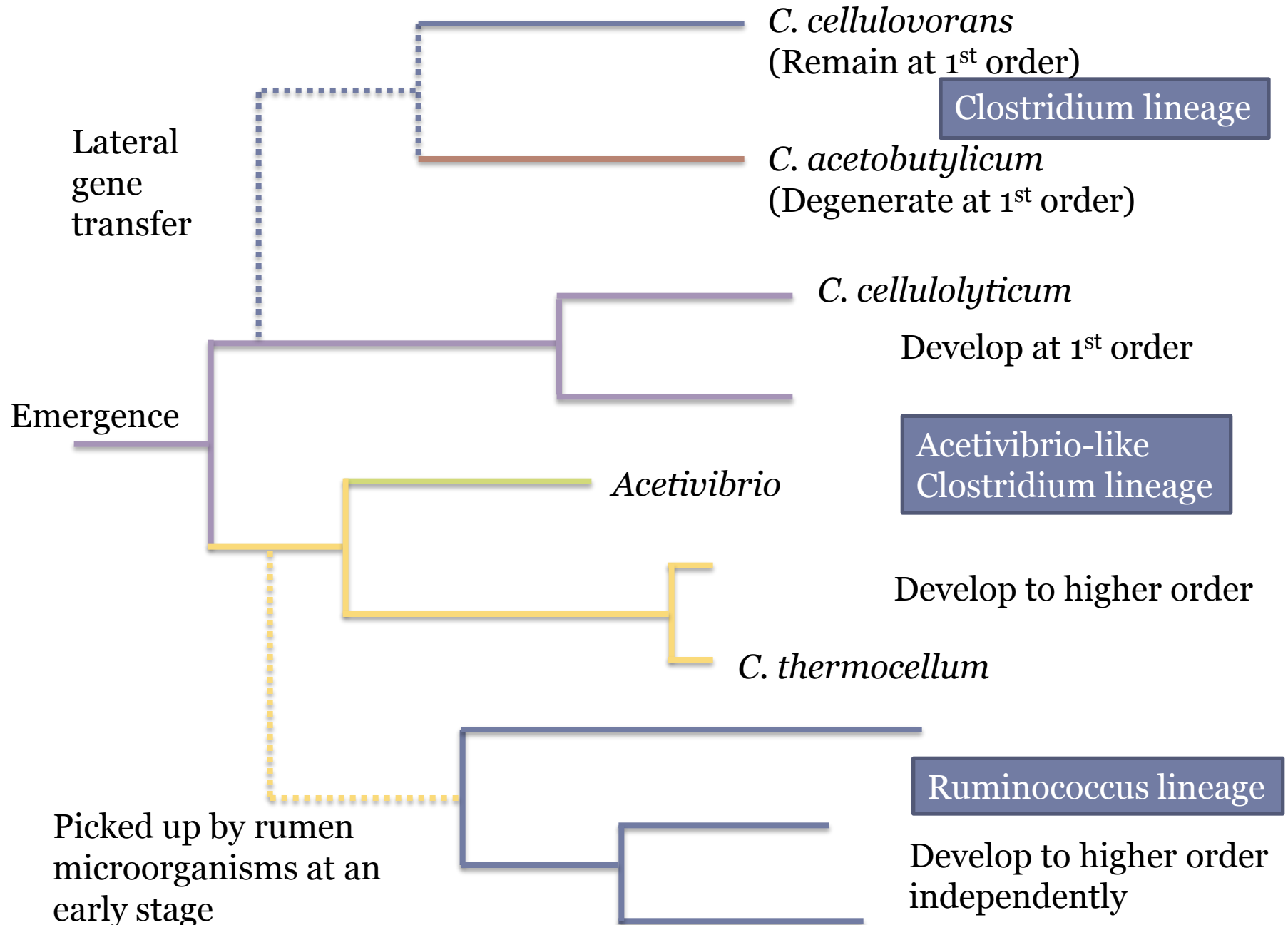
## Clostridium cellulolyticum H10



## Clostridium acetobutylicum ATCC 824
plasmid pSOL1

How modular evolution produces functional strength?

*C. cellulovorans*
(Remain at 1st order)

Clostridium lineage

*C. acetobutylicum*
(Degenerate at 1st order)

Lateral
gene
transfer

*C. cellulolyticum*

Develop at 1st order

Emergence

Acetivibrio-like
Clostridium lineage

*Acetivibrio*

Develop to higher order

*C. thermocellum*

Ruminococcus lineage

Picked up by rumen
microorganisms at an
early stage

Develop to higher order
independently

# Emergent properties of the cellulosome



**Cellulosome Gene Diversity**
Recombination
LGT
Gene Duplication/Loss
Selection

**Emergent Properties of the Cellulosome**

**Enzymatic Composition of Cellulosome**
Phylogeny
Regulatory Networks
Chemical Composition of Biomass

**3D Structure of Cellulosome**
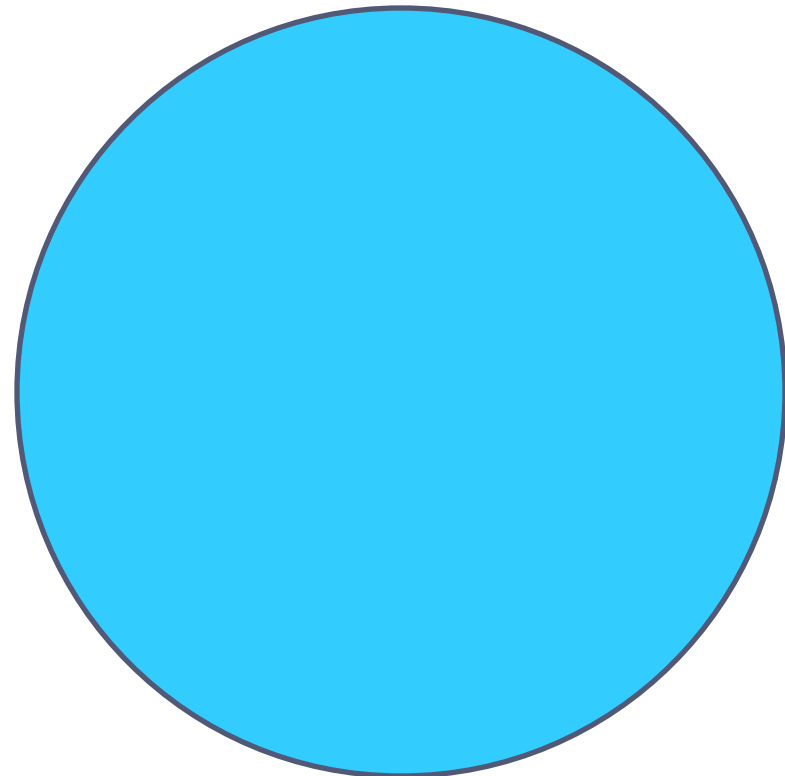# Scaffoldins
# Cohesin/Dockerin Pairs
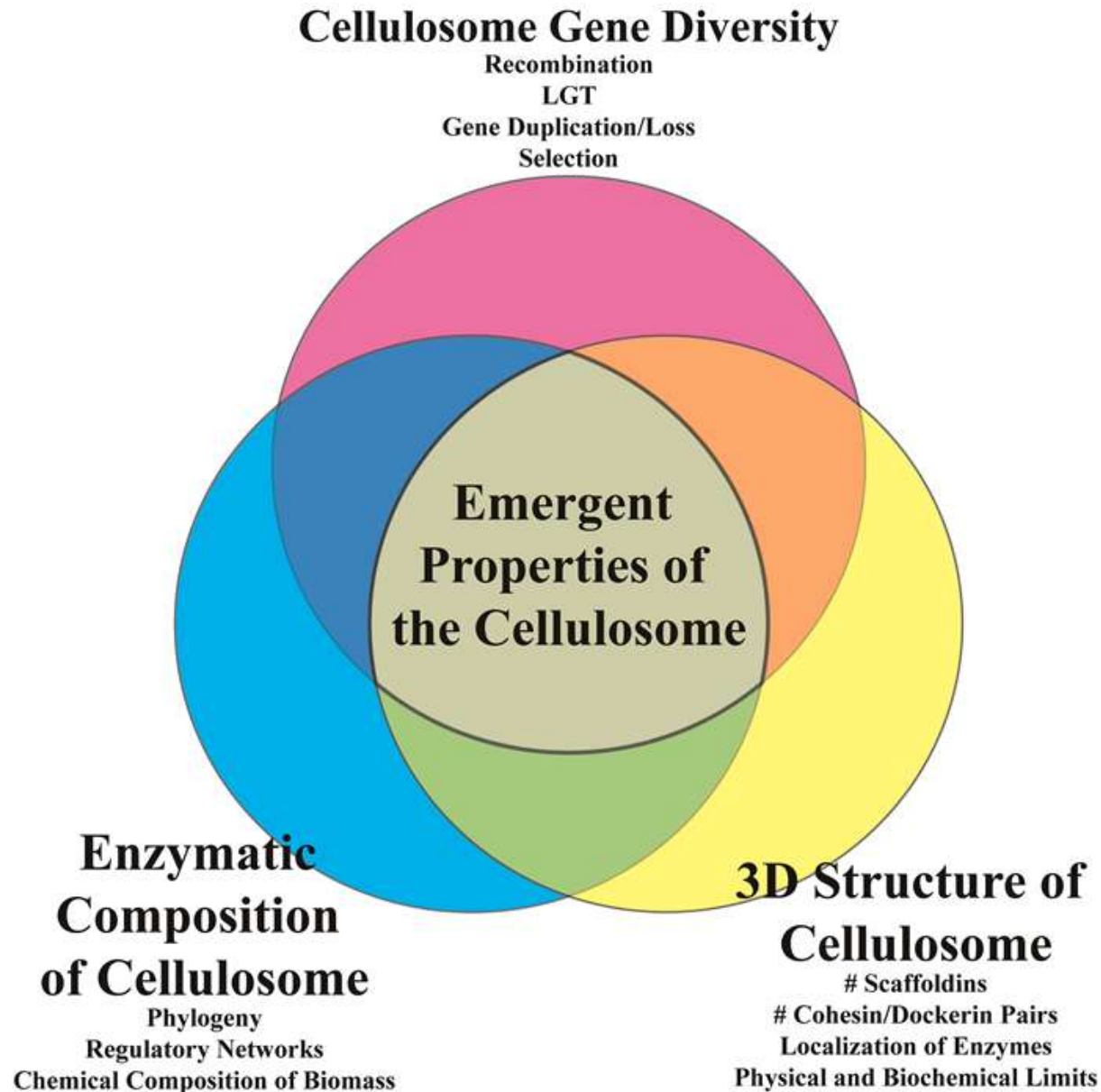Localization of Enzymes
Physical and Biochemical Limits

# Enzymatic composition of the cellulosome

- We are using Microarray, RNA-seq, and iTRAQ to study the transcriptome and proteome of the 12 strains under cellulolytic conditions

# 3D structure of the cellulosome

- \# of scaffoldins and cohesin domains (i.e. binding sites)

- \# of cohesin subfamilies

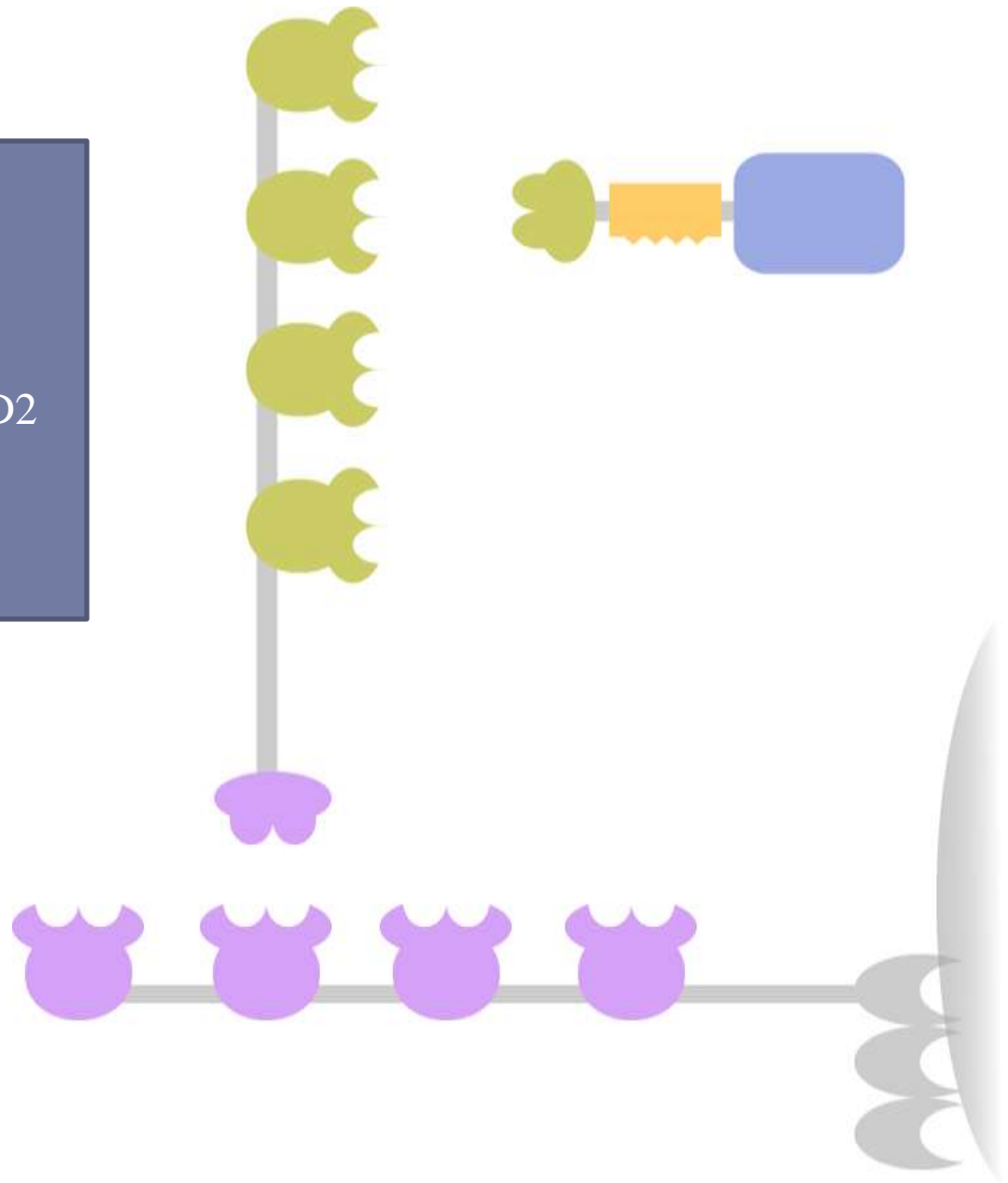  (How does modular evolution produce structural complexity?)
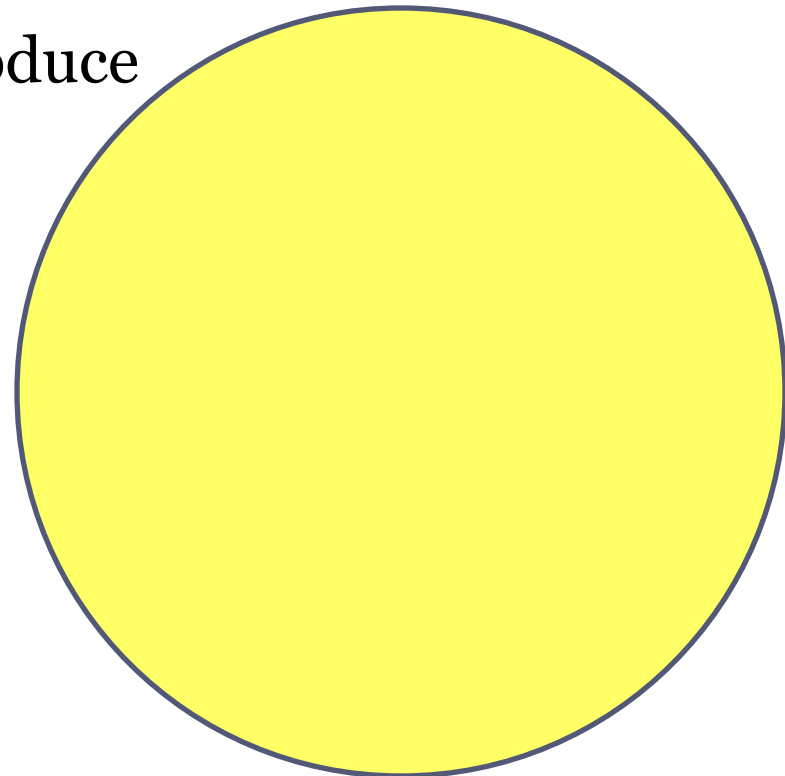
*Clostridium thermocellum*
ATCC 27405
Second order: 7*9 = 63

*Acetivibrio cellulolyticus* CD2
Third order: 3*4*7 = 84

Second order cellulosome

# # of scaffoldins and cohesin domains

| | Scaffoldins | Cohesin domains |
|---|---|---|
| *C. acetobutylicum* ATCC 824 | 2 | 6 |
| *C. cellulovorans* 743B | 4 | 11 |
| *C. papyrosolvens* DSM 2782 | 5 | 6 |
| *C. cellulolyticum* H10 | 2 | 9 |
| *C. thermocellum* ATCC LQR1 | 11 | 20 |
| *C. thermocellum* JW20 | 8 | 26 |
| *C. thermocellum* ATCC 27405 | 8 | 29 |
| ***Acetivibrio*** | **16** | **47** |
| *R. Albus* 7 | 1 | 1 |
| *R. Albus* 8 | 0 | 0 |
| *Ruminococcus* sp. 18P13 | 6 | 9 |
| ***R. Flavifaciens* FD-1** | **9** | **18** |

How modular evolution produces structural complexity?

Acetivibrio cellulolyticus CD2

Clostridium thermocellum ATCC 27405

Clostridium thermocellum ATCC 27405

2nd order

3rd order or higher

Clostridium cellulovorans 743B

Acetivibrio cellulolyticus CD2

Clostridium cellulolyticum H10
Clostridium papyrosolvens DSM 2782

Acetivibrio cellulolyticus CD2

Clostridium acetobutylicum ATCC 824

Bayesian majority consensus tree of all cohesin domains

# Emergent properties of the cellulosome



**Cellulosome Gene Diversity**
Recombination
LGT
Gene Duplication/Loss
Selection

**Emergent Properties of the Cellulosome**

**Enzymatic Composition of Cellulosome**
Phylogeny
Regulatory Networks
Chemical Composition of Biomass

**3D Structure of Cellulosome**
# Scaffoldins
# Cohesin/Dockerin Pairs
Localization of Enzymes
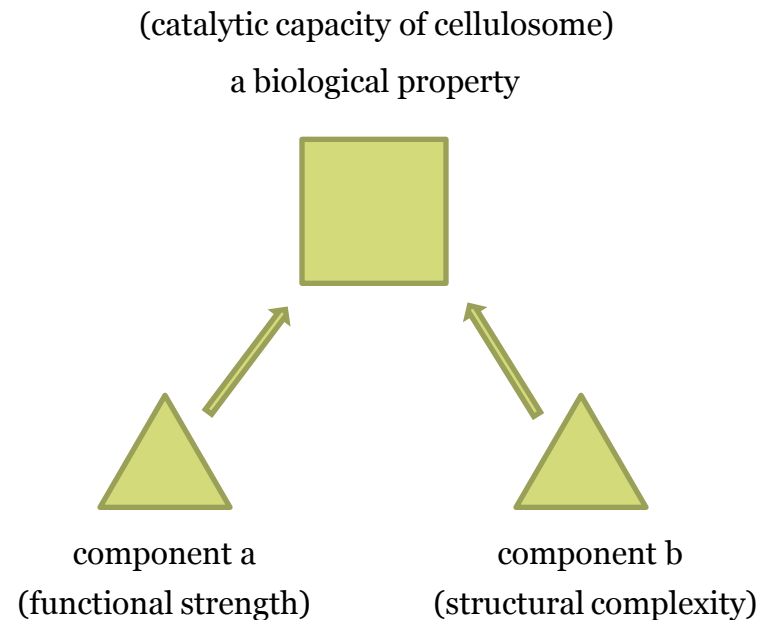Physical and Biochemical Limits

# Questions

- How modular evolution produces functional strength?

- How modular evolution produces structural complexity?

- What's the correlation between functional strength and structural complexity, and how does natural selection drive the co-evolution?

- How does the cellulosome evolve in different lineages?

# Modes of co-evolution

Co-evolution between two objects
that interact with each other
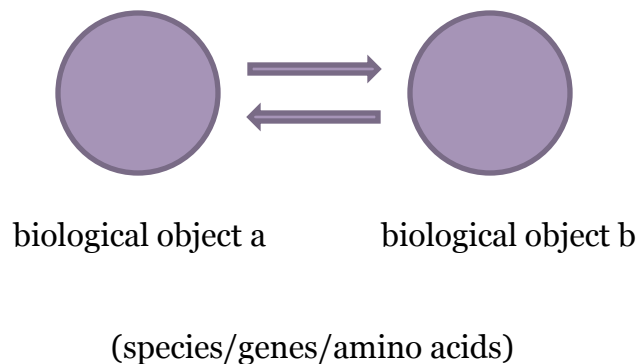
Co-evolution between two components
that contribute to a biological property
in an interdependent manner
but don't interact directly

(catalytic capacity of cellulosome)

a biological property

biological object a        biological object b

(species/genes/amino acids)

component a                    component b
(functional strength)      (structural complexity)

# Correlation between structural complexity and functional strength
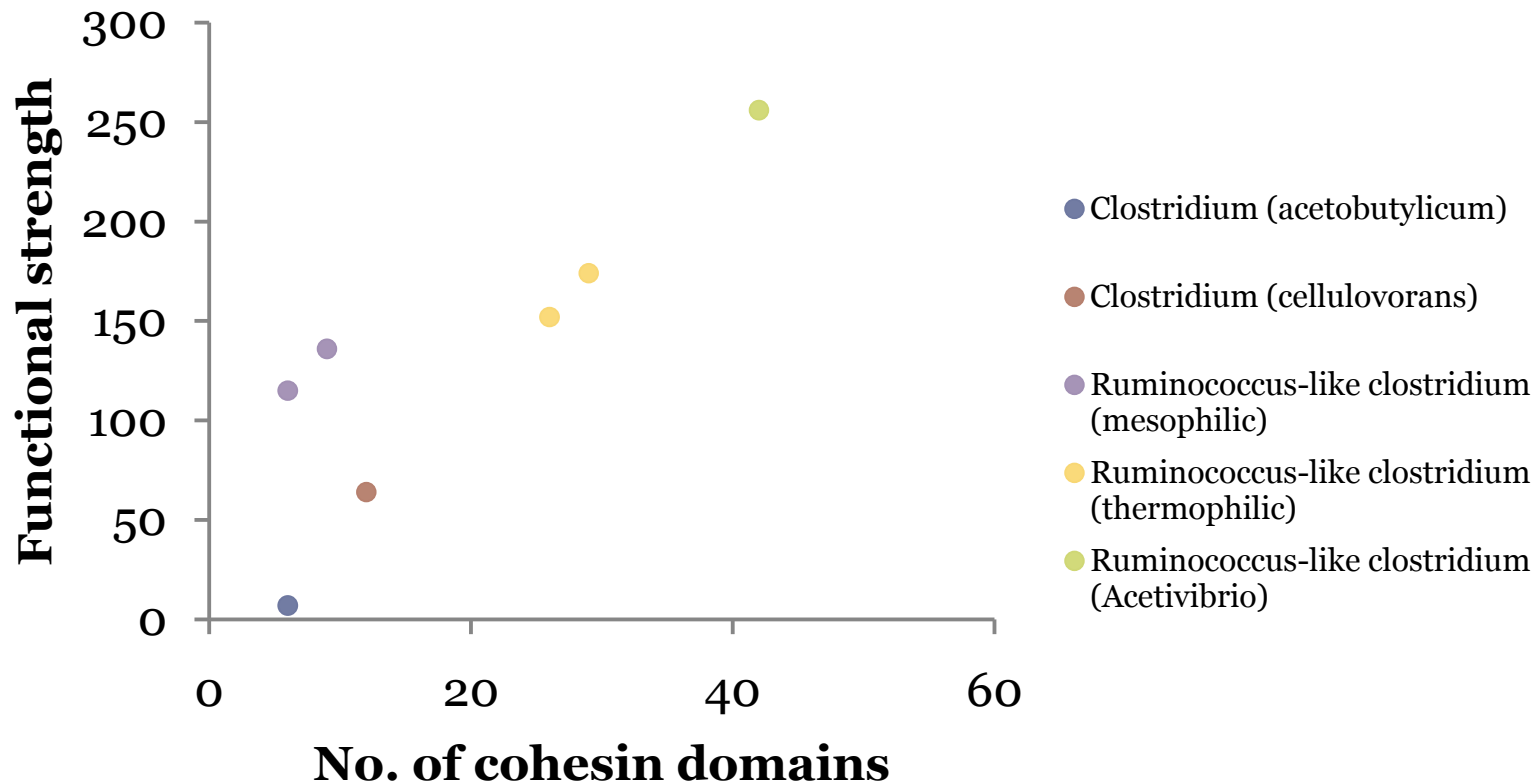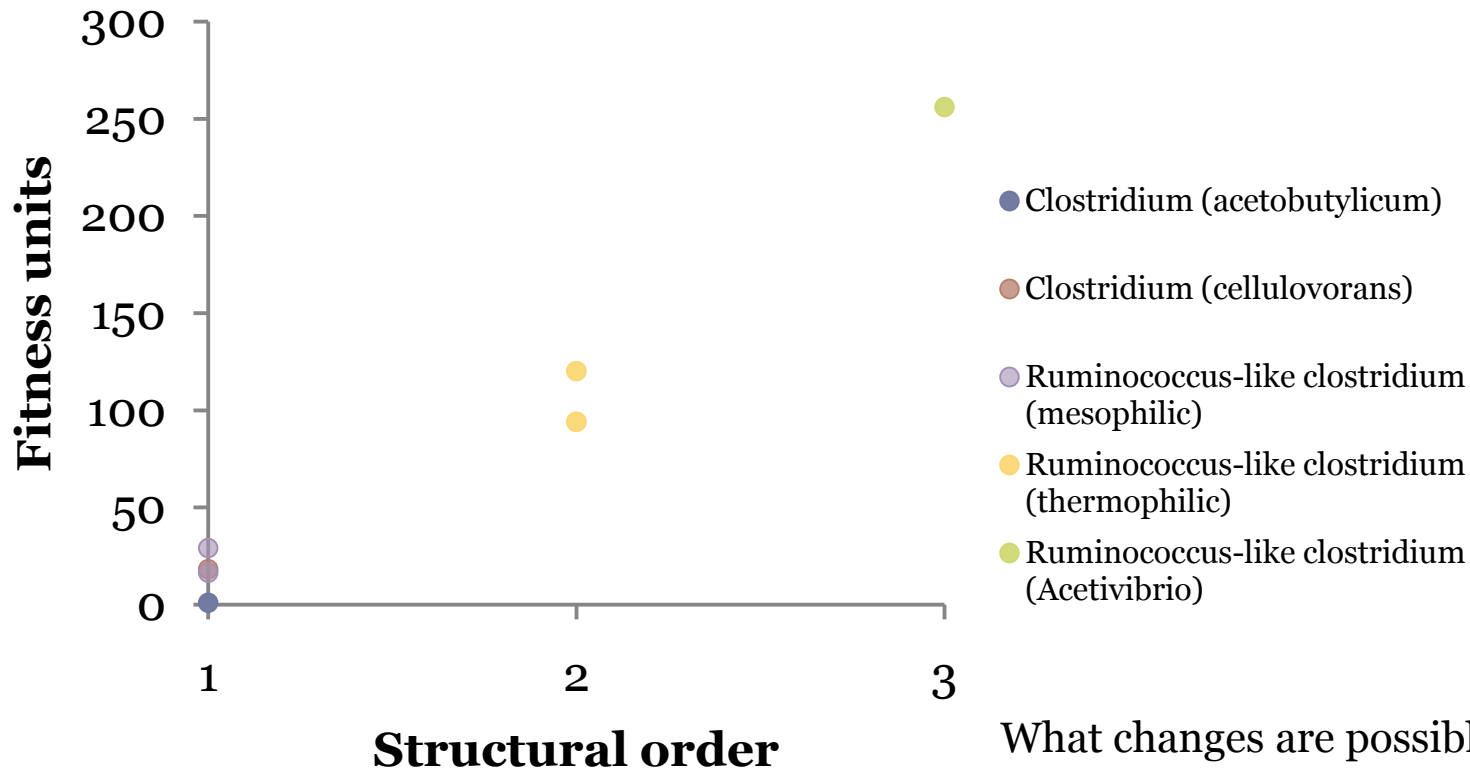


Functional strength =
No. of catalytic domains * Shannon diversity index

Evolution proceeds by small mutual adjustments of the two components in order to function properly.

# Correlation between structural complexity and functional strength



**Fitness units** (y-axis): 0, 50, 100, 150, 200, 250, 300

**Structural order** (x-axis): 1, 2, 3

- Clostridium (acetobutylicum)
- Clostridium (cellulovorans)
- Ruminococcus-like clostridium (mesophilic)
- Ruminococcus-like clostridium (thermophilic)
- Ruminococcus-like clostridium (Acetivibrio)

Enzymatic capability (fitness) =
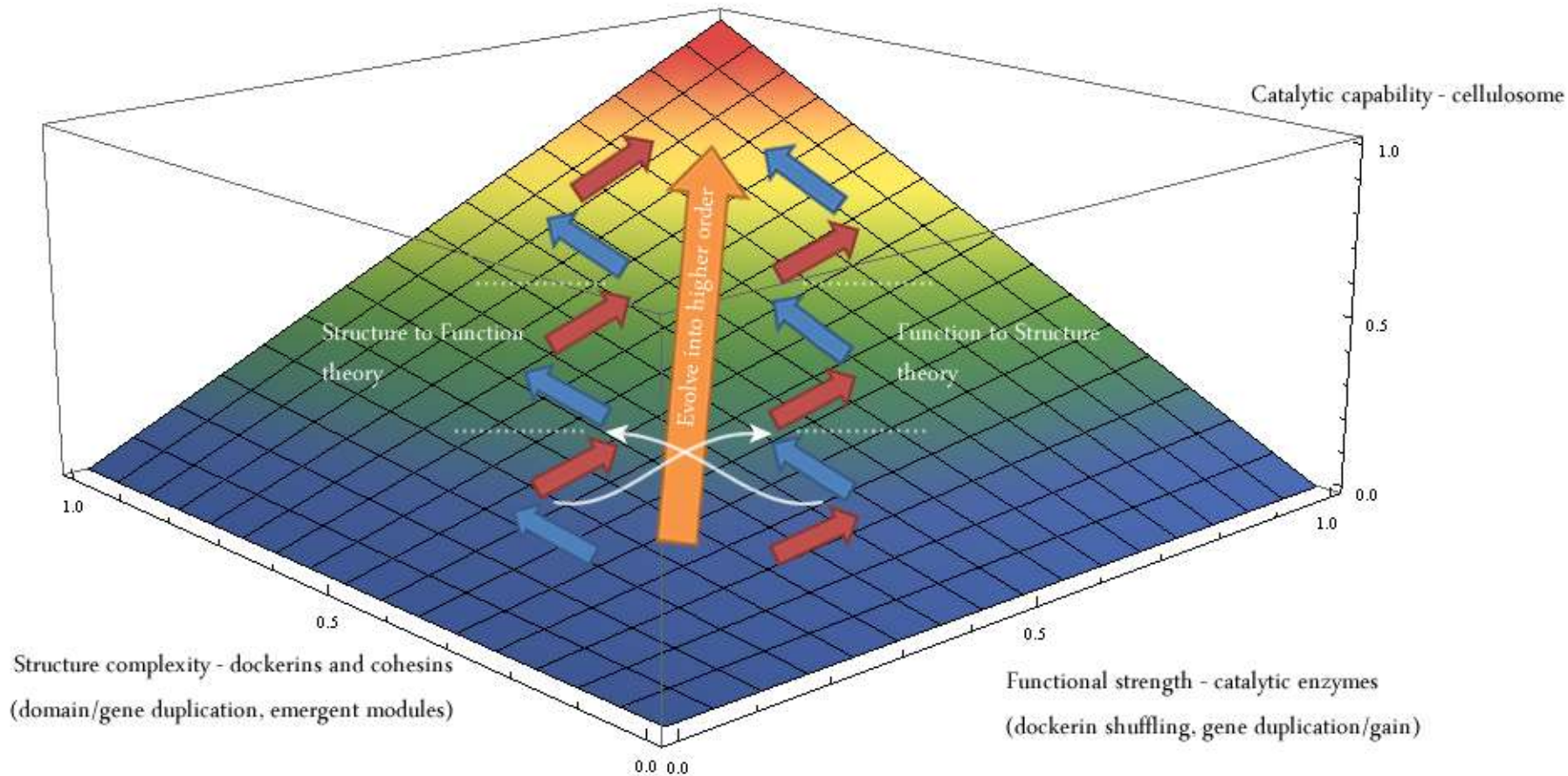No. of cohesins * Functional strength

What changes are possible depends on the pre-existing structure of the cellulosome. Without increase in structural order, fitness can only increase within certain limits, which are constrained by the underlying existing states.

# Correlation between structural complexity and functional strength

- Physical and genetic limitations to number of enzymes incorporated
  - Length of scaffoldin gene
  - Each enzyme takes up space
- Biochemical constrains
  - Enzymatic activity
  - Enzymatic diversity
  - Balance between two properties

# Correlation between structural complexity and functional strength

Catalytic capability - cellulosome

1.0

0.5

0.0

1.0

Structure to Function theory

Function to Structure theory

Evolve into higher order

1.0

0.5

Structure complexity - dockerins and cohesins
(domain/gene duplication, emergent modules)

0.5

Functional strength - catalytic enzymes
(dockerin shuffling, gene duplication/gain)

0.0 0.0

Structure to Function scenario offers a great potential for functional strength to develop and might play a major role in the evolution to higher order (e.g. the *Acetivibrio*-like *Clostridium* lineage).

# Questions

- How modular evolution produces functional strength?

- How modular evolution produces structural complexity?

- What's the correlation between functional strength and structural complexity, and how does natural selection drive the co-evolution?

- **How does the cellulosome evolve in different lineages?**

# Summary

- Domain shuffling and gene duplication play a major role in the evolution of functional strength.

- Domain duplication and divergence of novel modules contribute to the evolution of structural complexity.

- Functional strength and structural complexity are co-evolved in different lineages.

- Structure to function process might play a major role in driving the evolution to higher order.

- The cellulosome was originated in the *Acetivibrio*-like *Clostridium* lineage, and laterally transferred to the *Clostridium* lineage. The higher-order cellulosome in the *Ruminococcus* lineage was evolved independently.

# Acknowledgments

- OU
  - Jizhong Zhou
  - Huang Yue
  - Yong-Jin Lee
- Clostridium Sequencing Consortium
- US-DOE Joint Genome Institute
- NSF EPSCoR