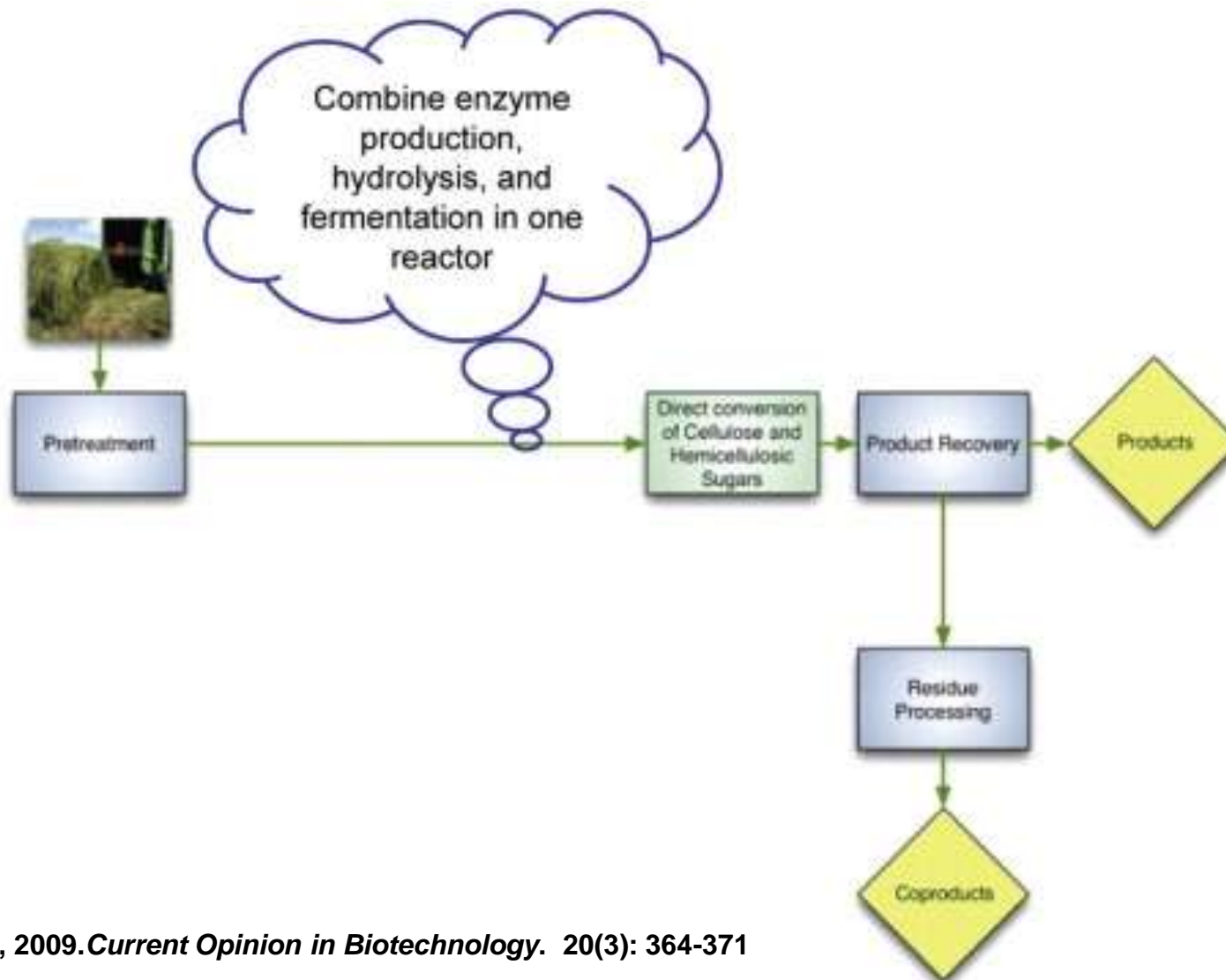


# The Structure and Evolution of the Cellulosome

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

# Consolidated bioprocessing



# Microbial strategies for enzymatic biomass degradation

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

# Cellulosome-encoding Clostridia

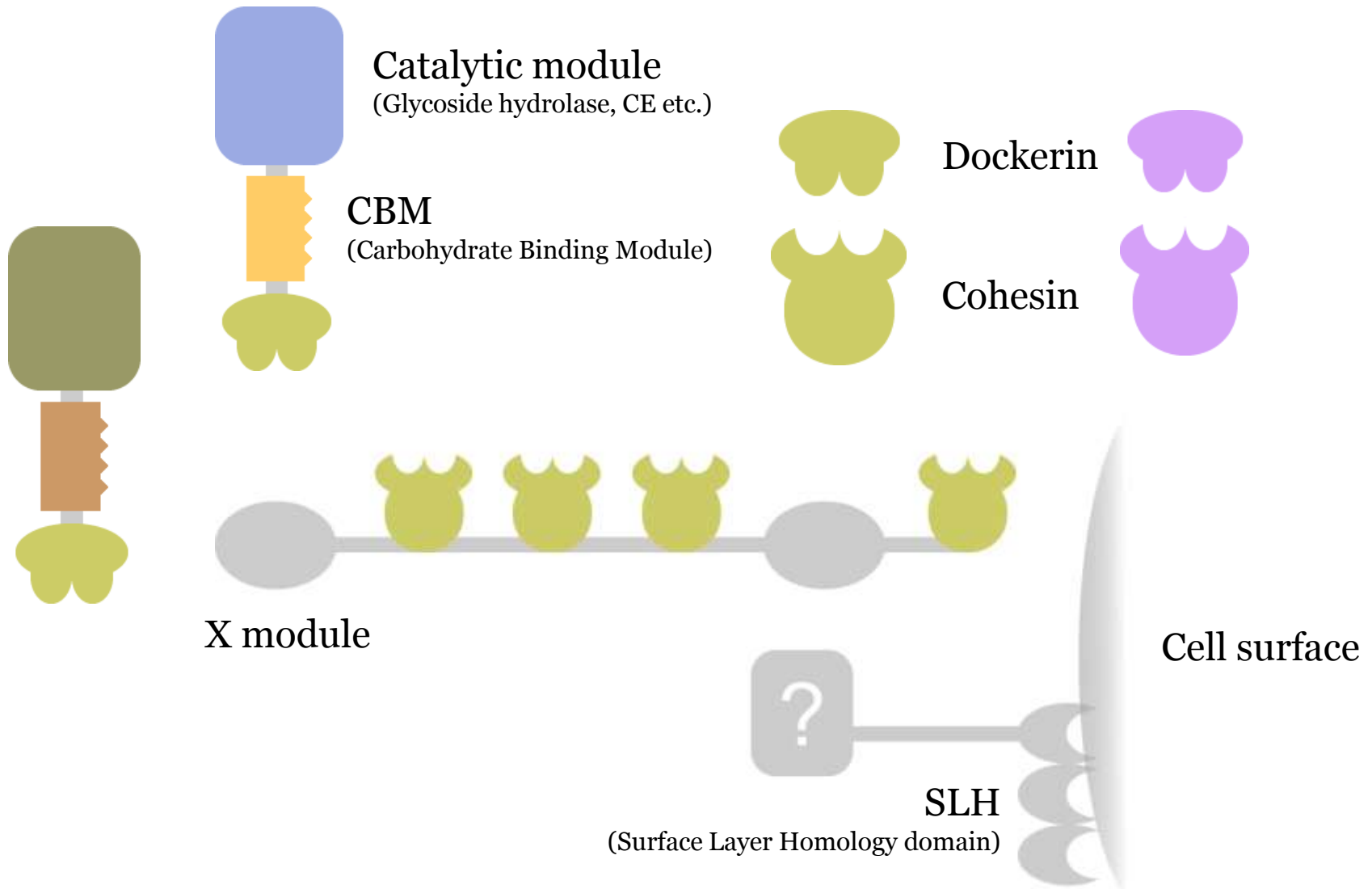
1. *Clostridium acetobutylicum* ATCC 824
2. *Clostridium cellulovorans* 743B
3. *Clostridium papyrosolvens* DSM 2782
4. *Clostridium cellulolyticum* H10
5. *Clostridium thermocellum* ATCC 27405
6. *Clostridium thermocellum* JW20
7. *Clostridium thermocellum* LQR1
8. *Acetivibrio cellulolyticus* CD2
  
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



# Emergent properties of the cellulosome

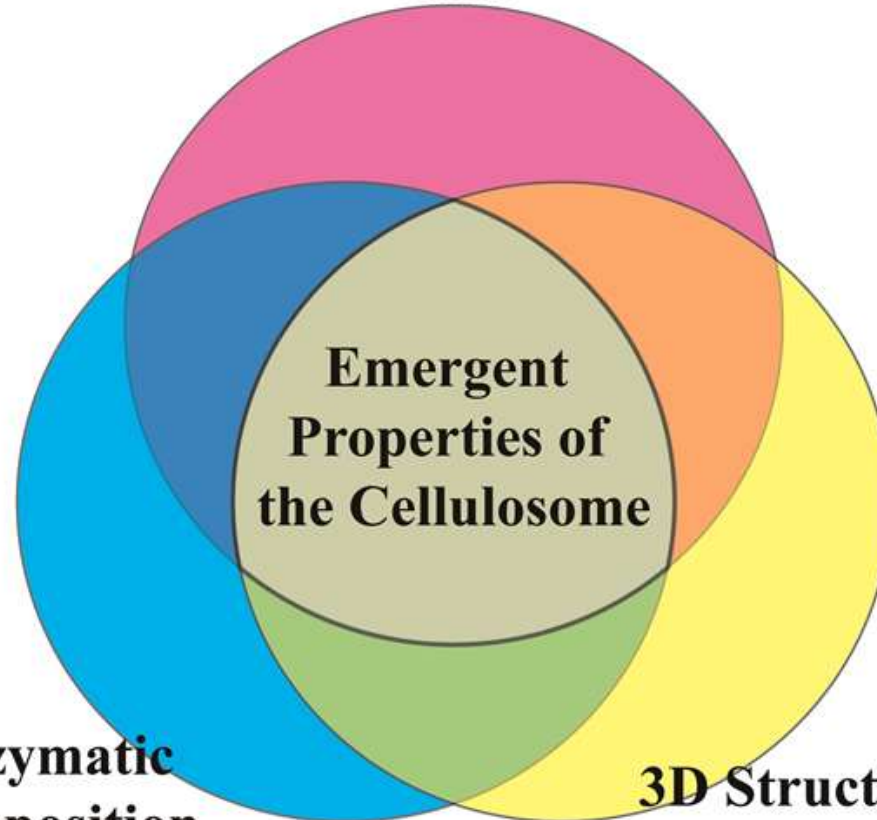
## Cellulosome Gene Diversity

Recombination

LGT

Gene Duplication/Loss

Selection



## Enzymatic Composition of Cellulosome

Phylogeny

Regulatory Networks

Chemical Composition of Biomass

## 3D Structure of Cellulosome

# Scaffoldins

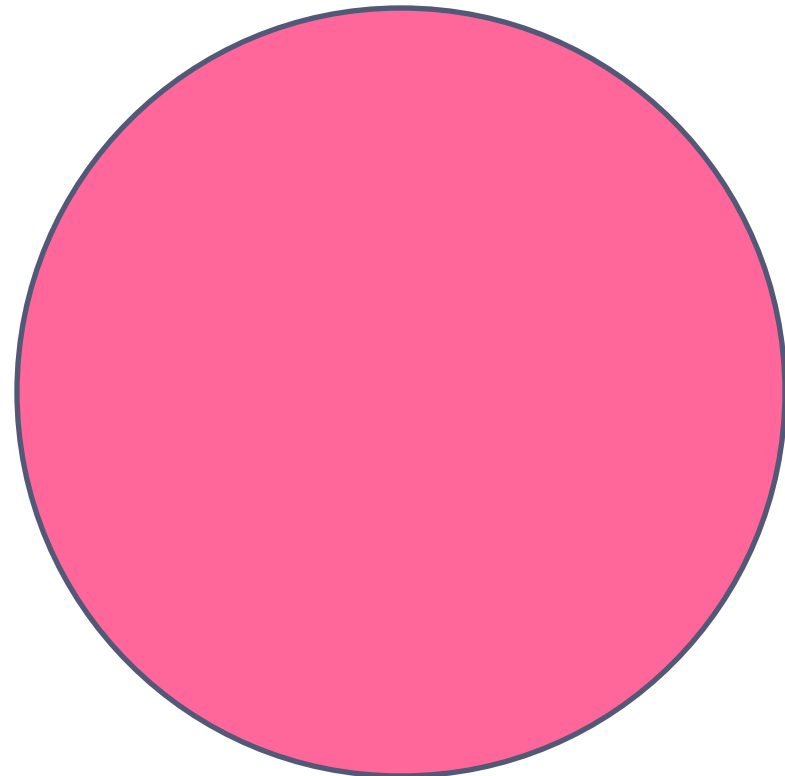
# Cohesin/Dockerin Pairs

Localization of Enzymes

Physical and Biochemical Limits

# Cellulosome gene diversity

- Diversity of functional gene categories
- Diversity within a functional gene category
- Evolutionary history within a functional gene category

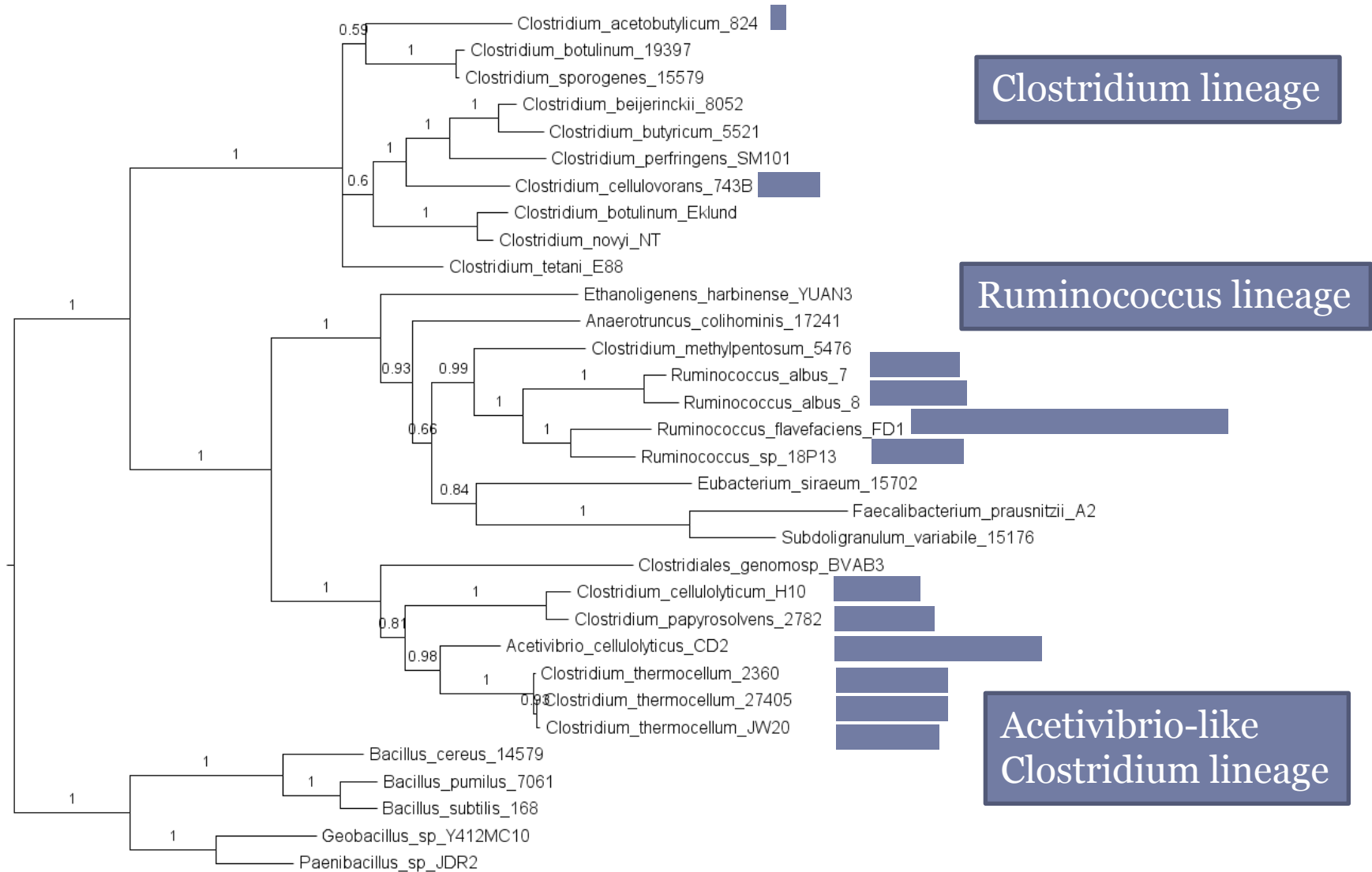


# Cellulosome-related genes

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

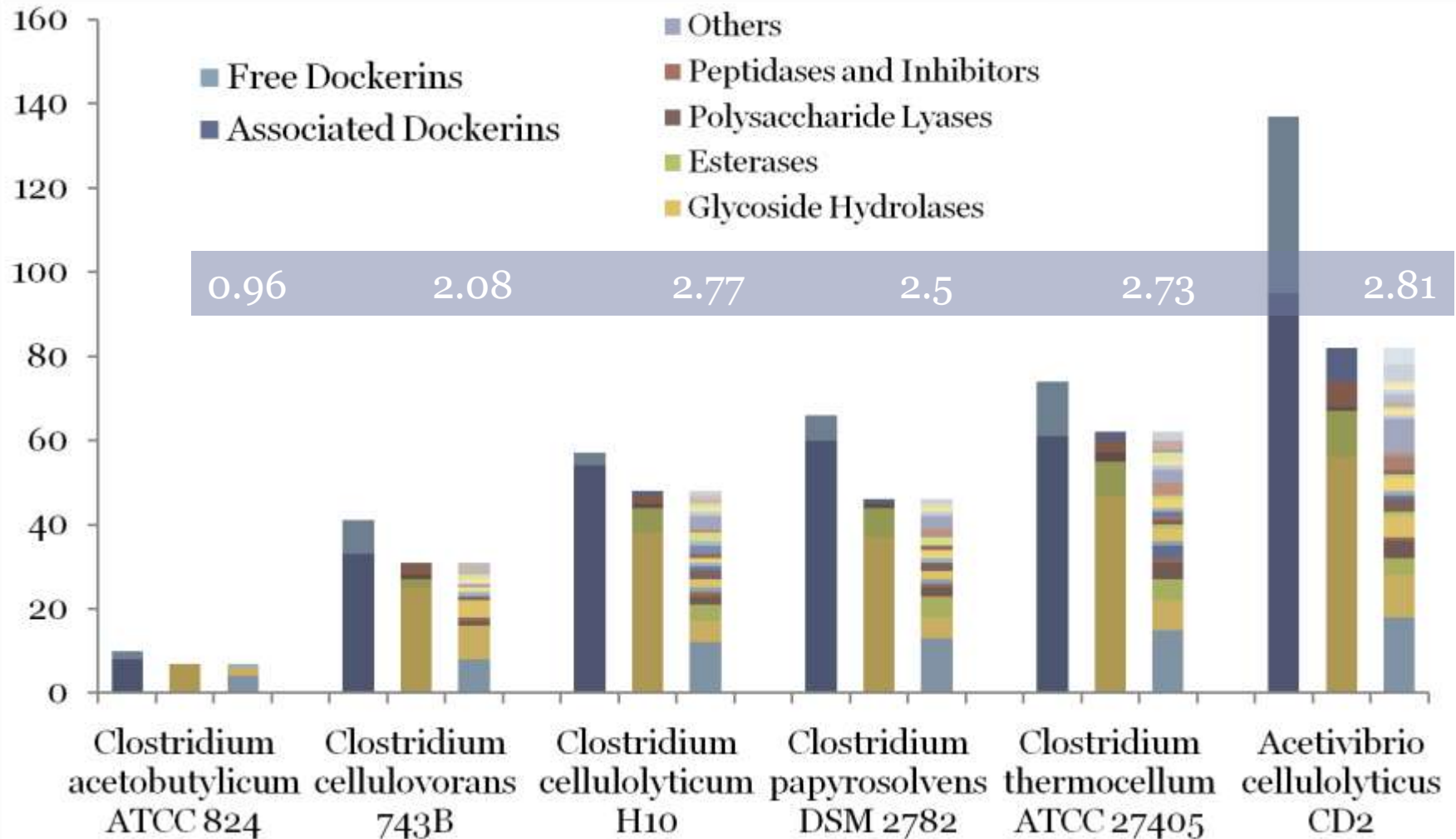


# Phylogenetic distribution of the cellulosome



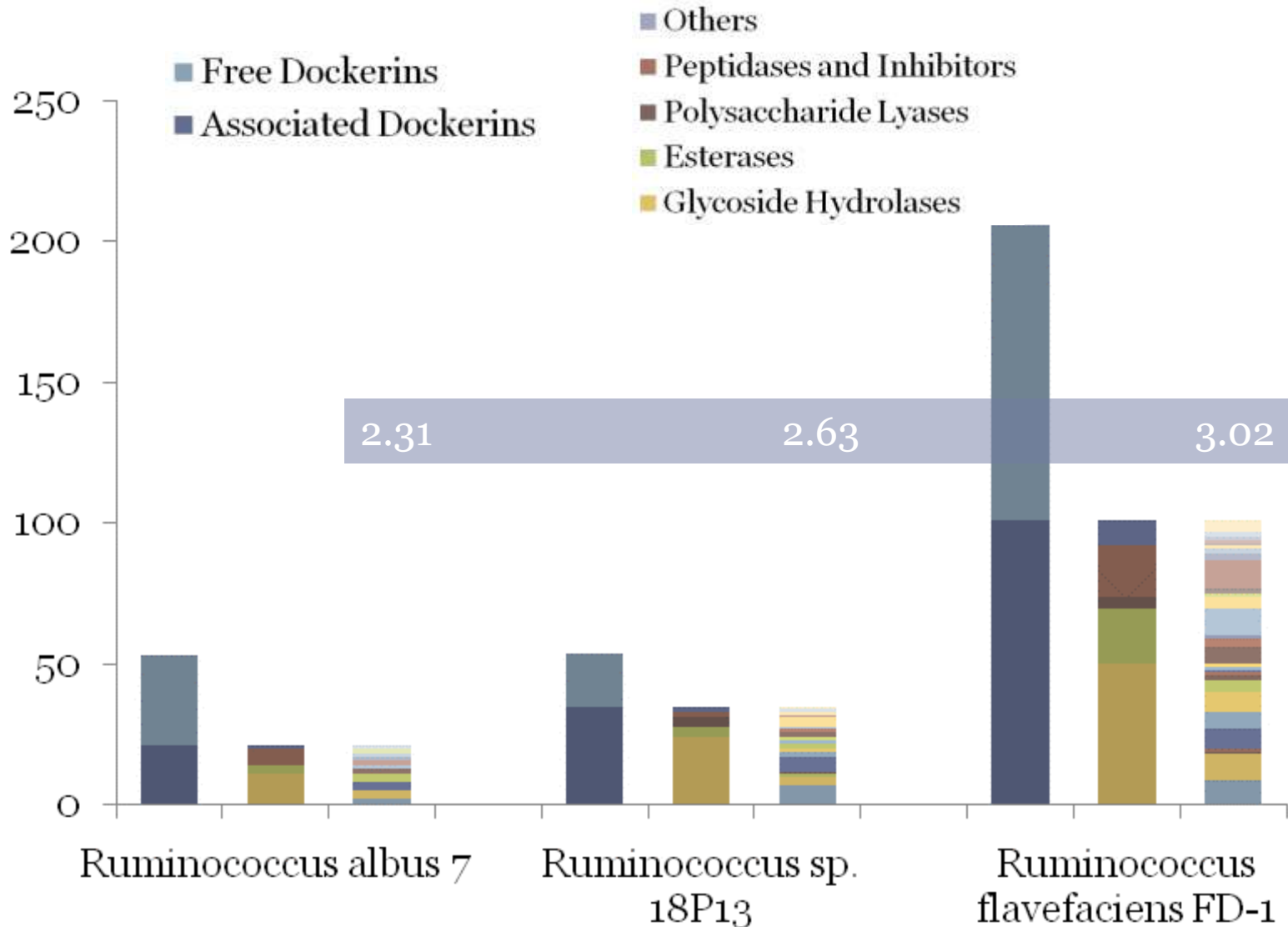
16S Phylogeny with posterior probability

# Diversity of functional gene categories



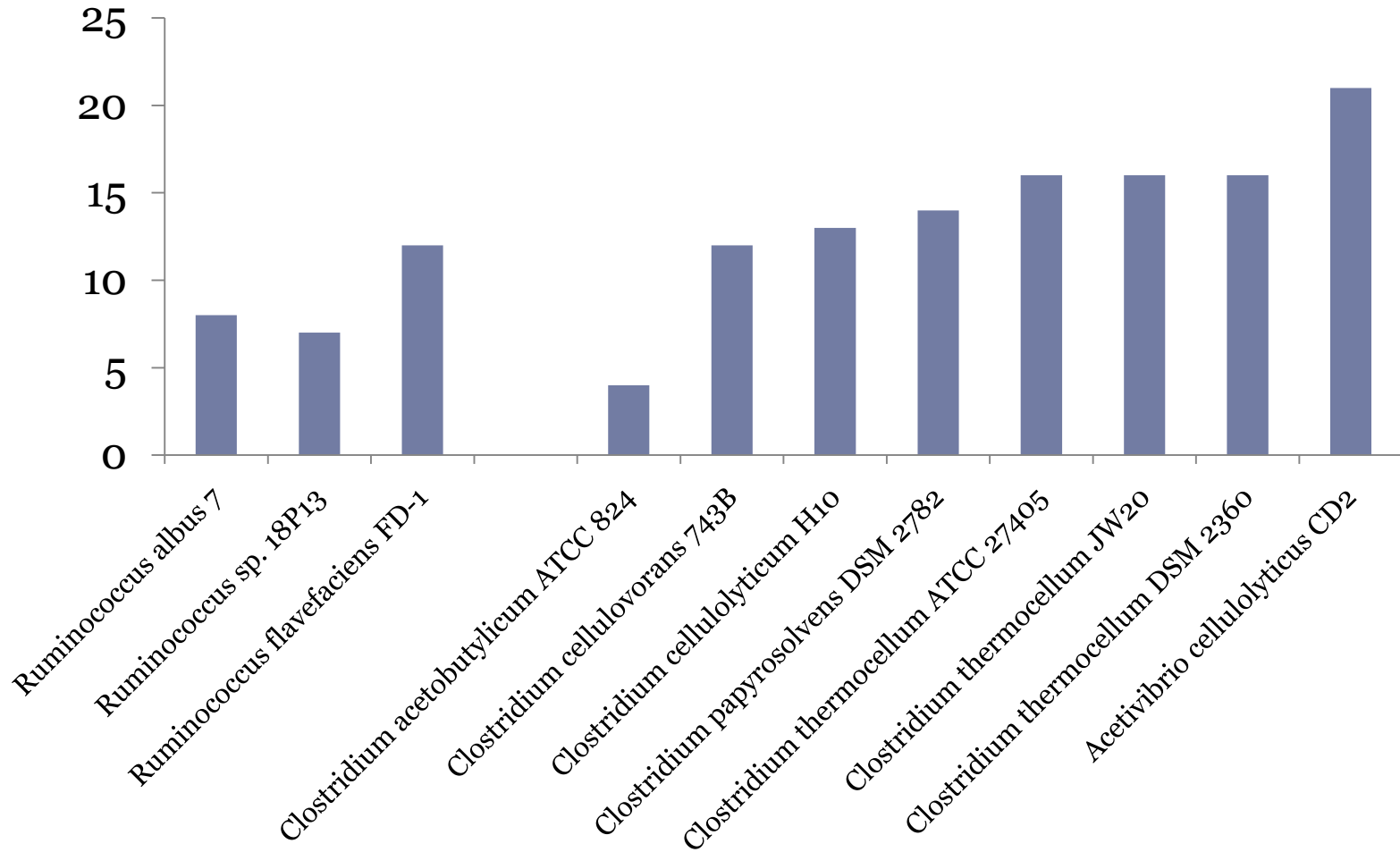
$$\text{Shannon diversity index} = -\sum(p_i * \ln p_i)$$

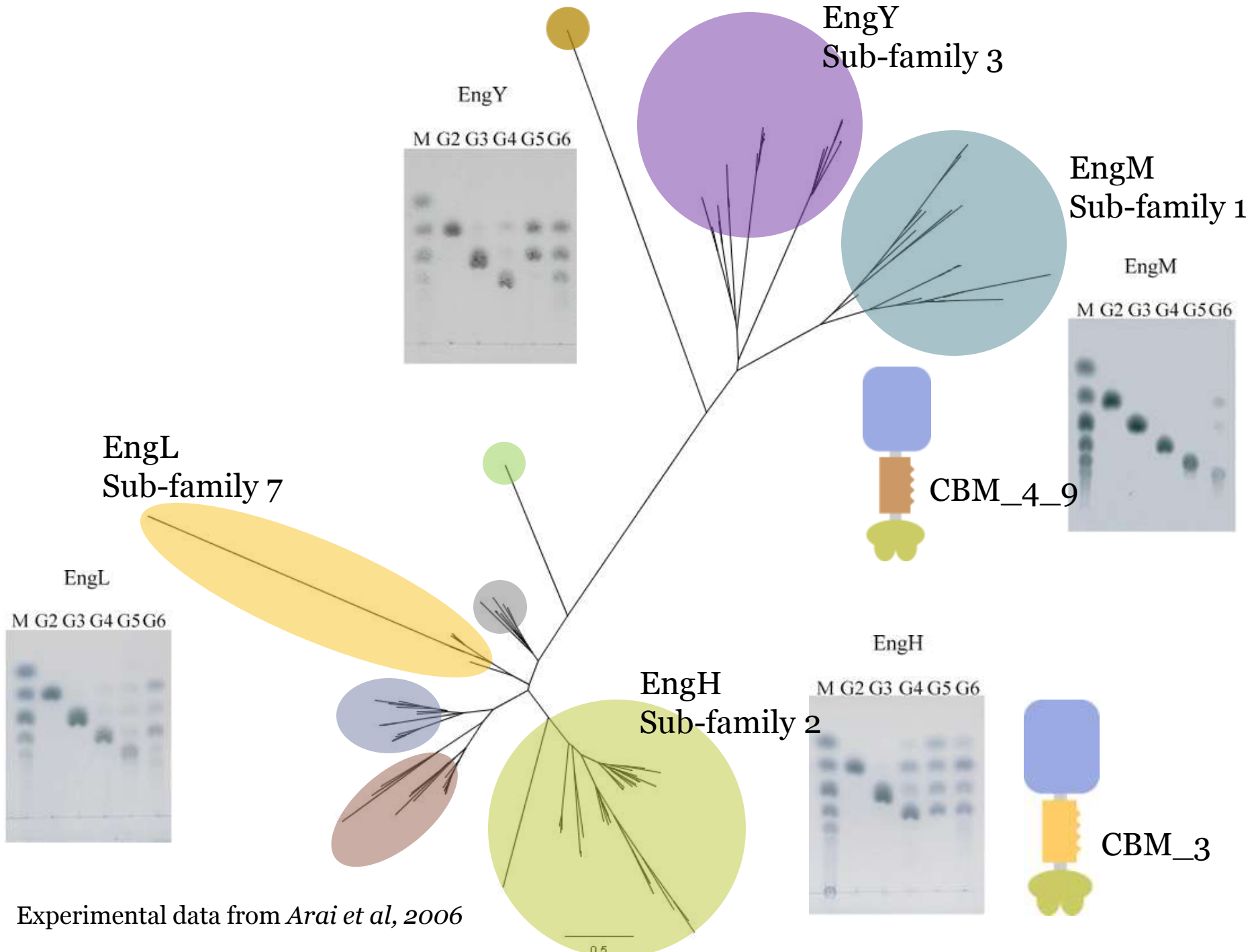
# Diversity of functional gene categories

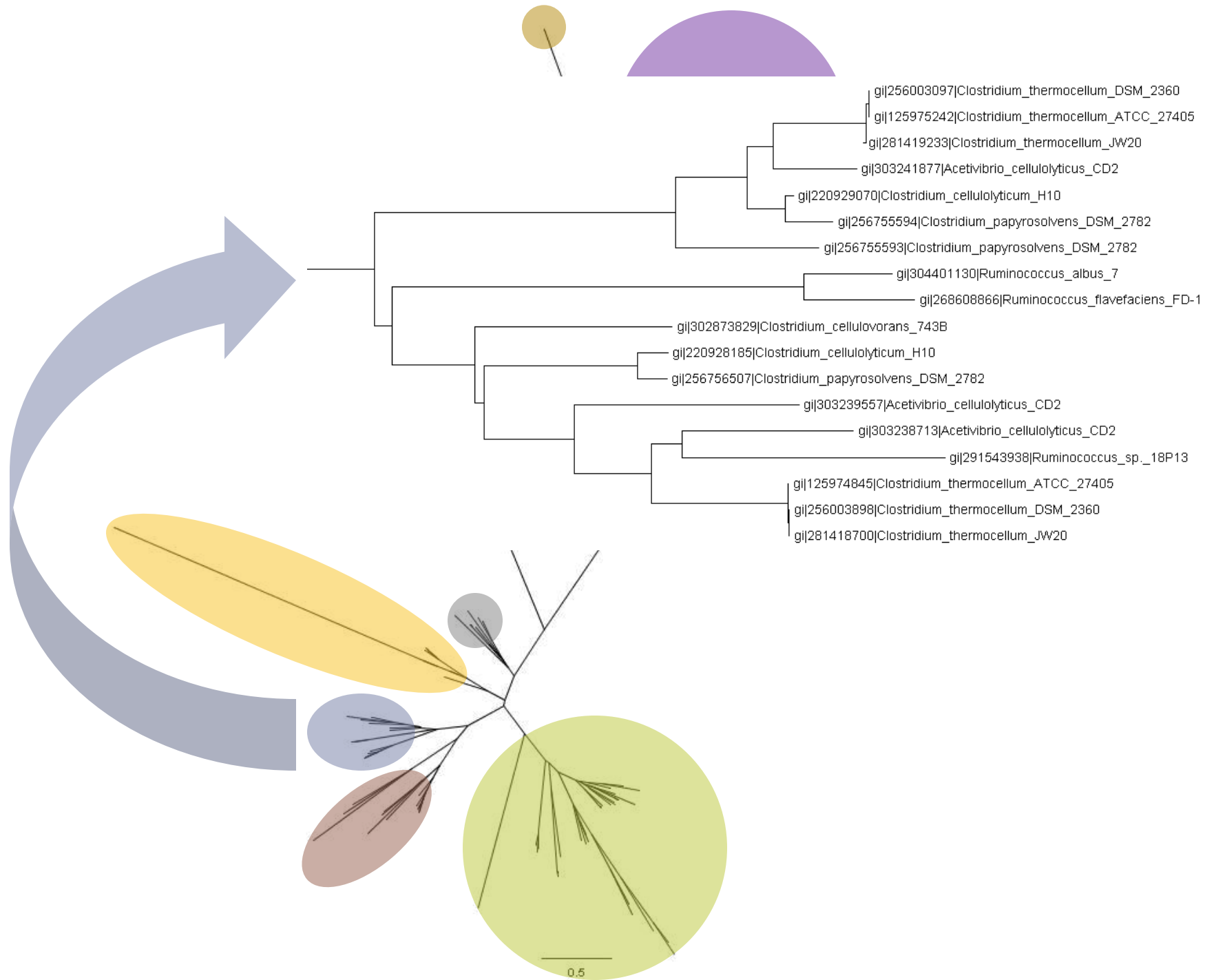


# Diversity within a functional gene category

## Glycoside Hydrolase family 9

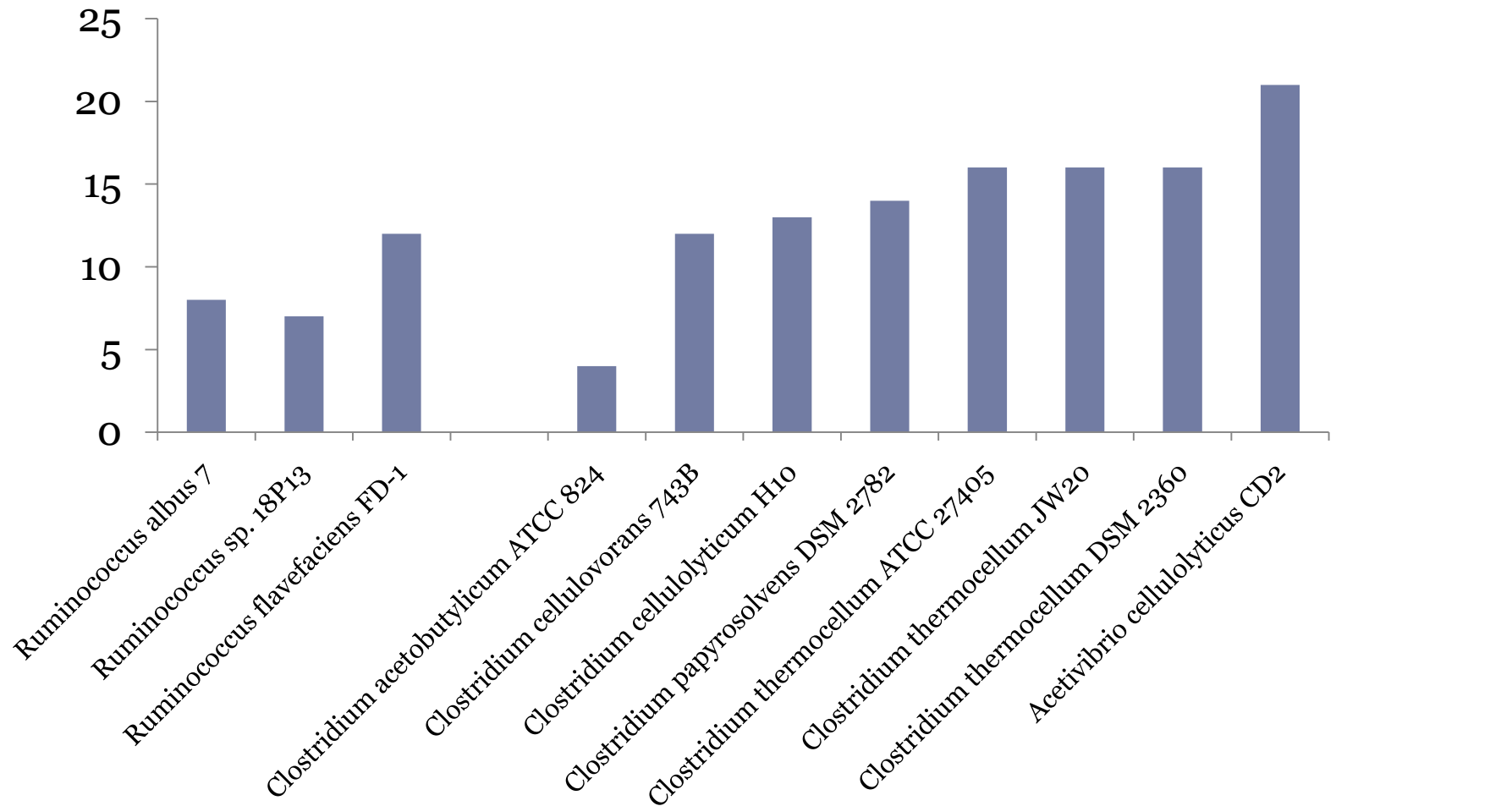




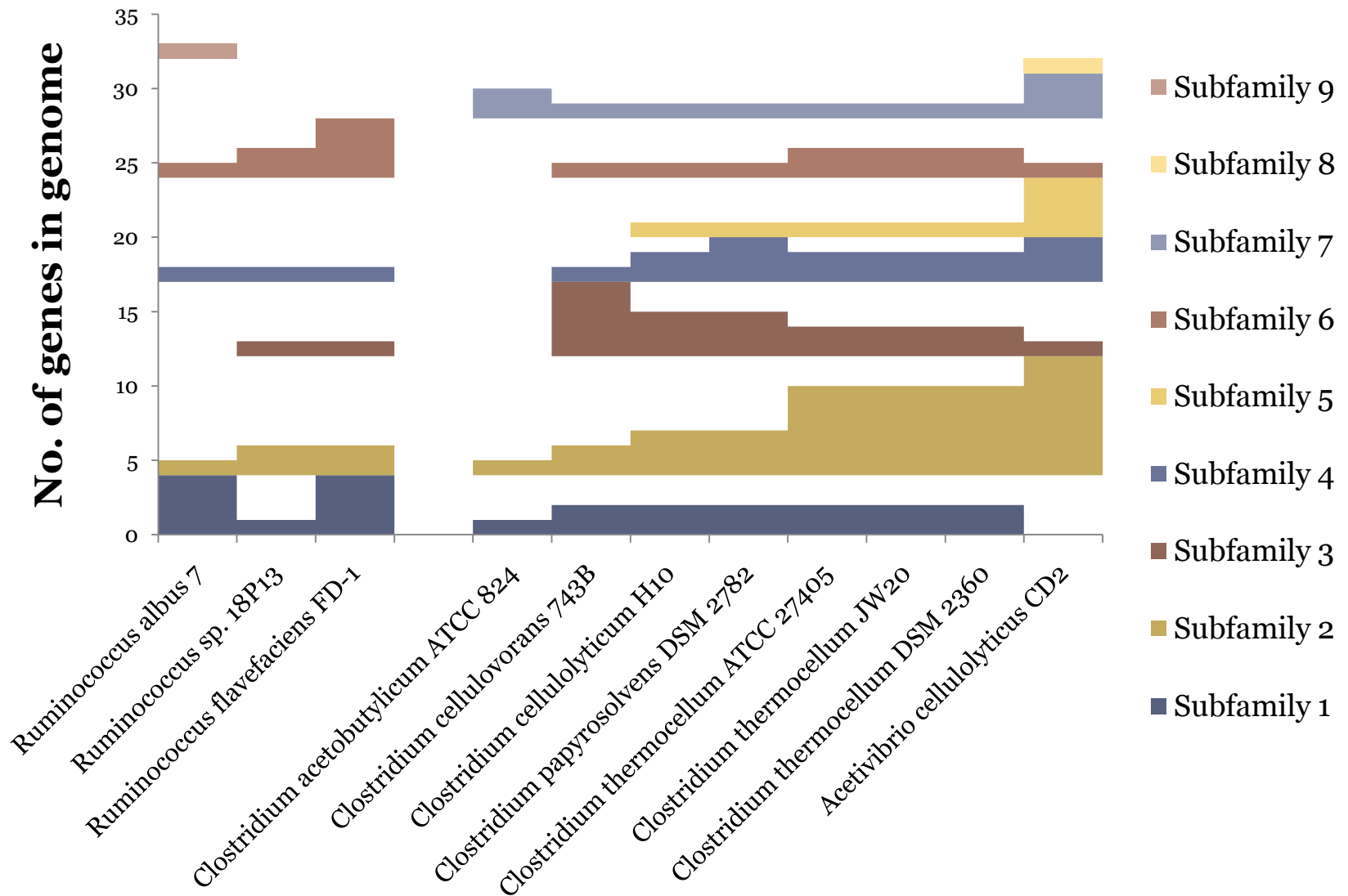


# Evolutionary history within a functional gene category

## Glycoside Hydrolase family 9



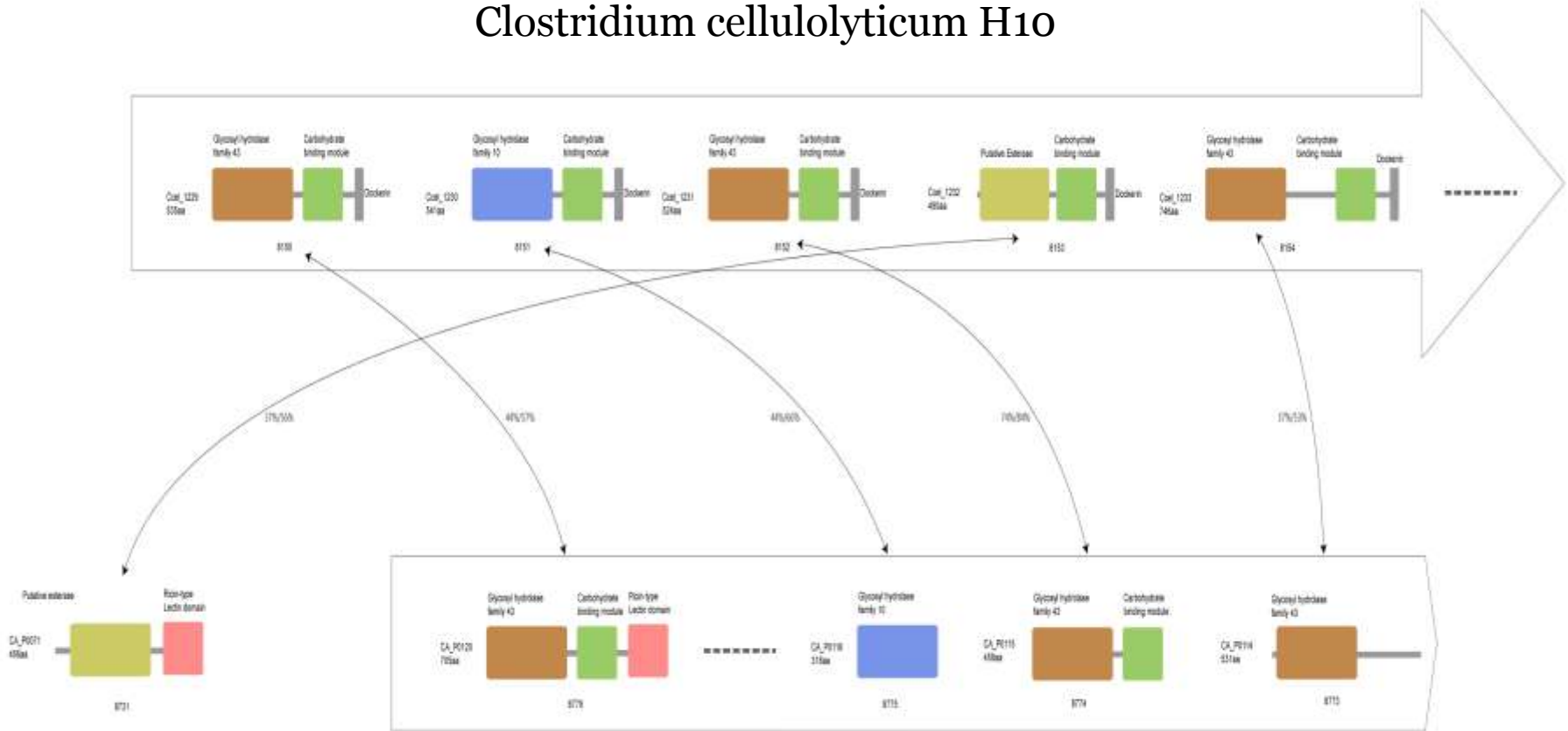
# Orthologs and paralogs of dockerin-containing genes





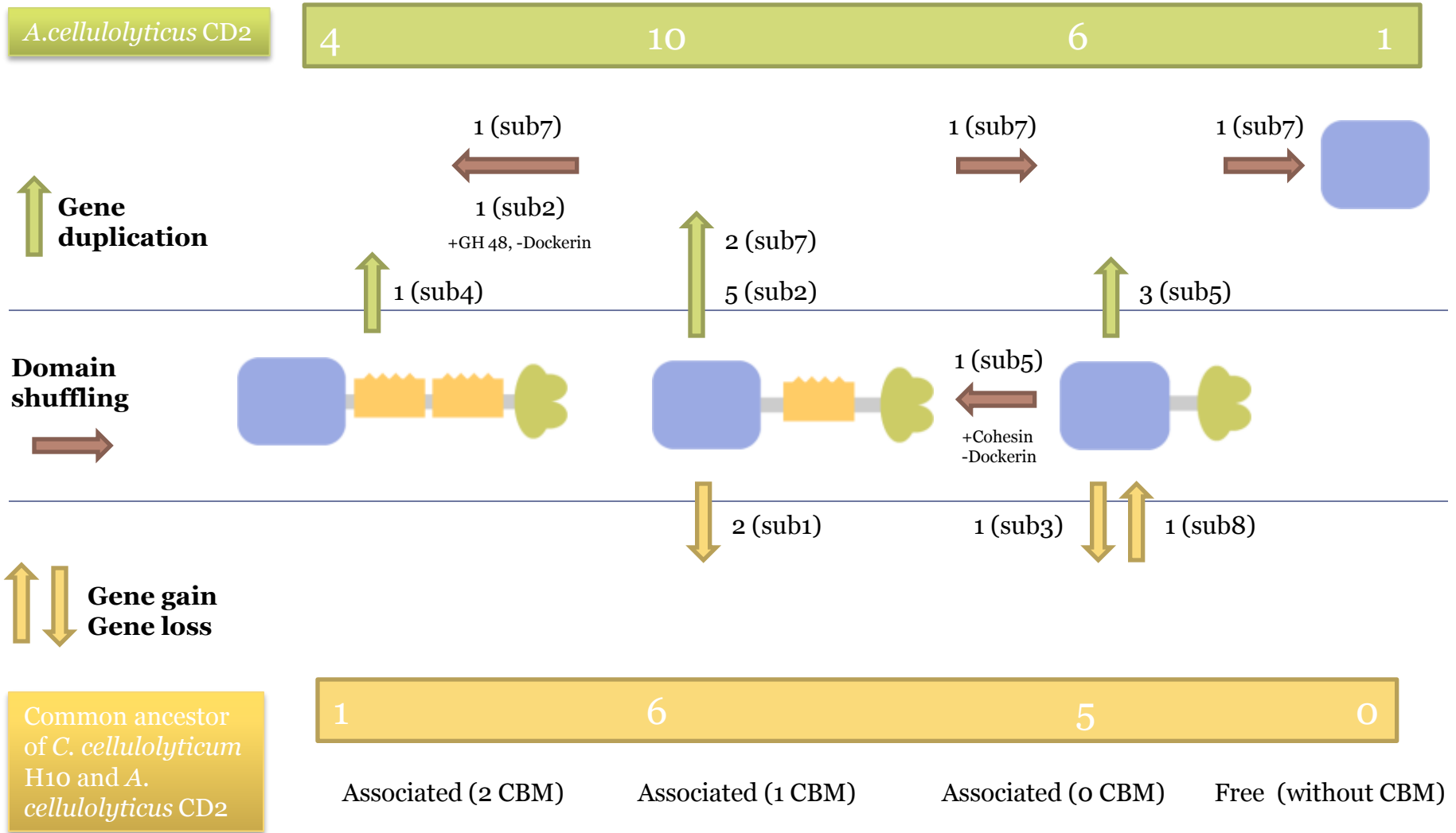
# Conservation of function but not gene structure

## Clostridium cellulolyticum H10

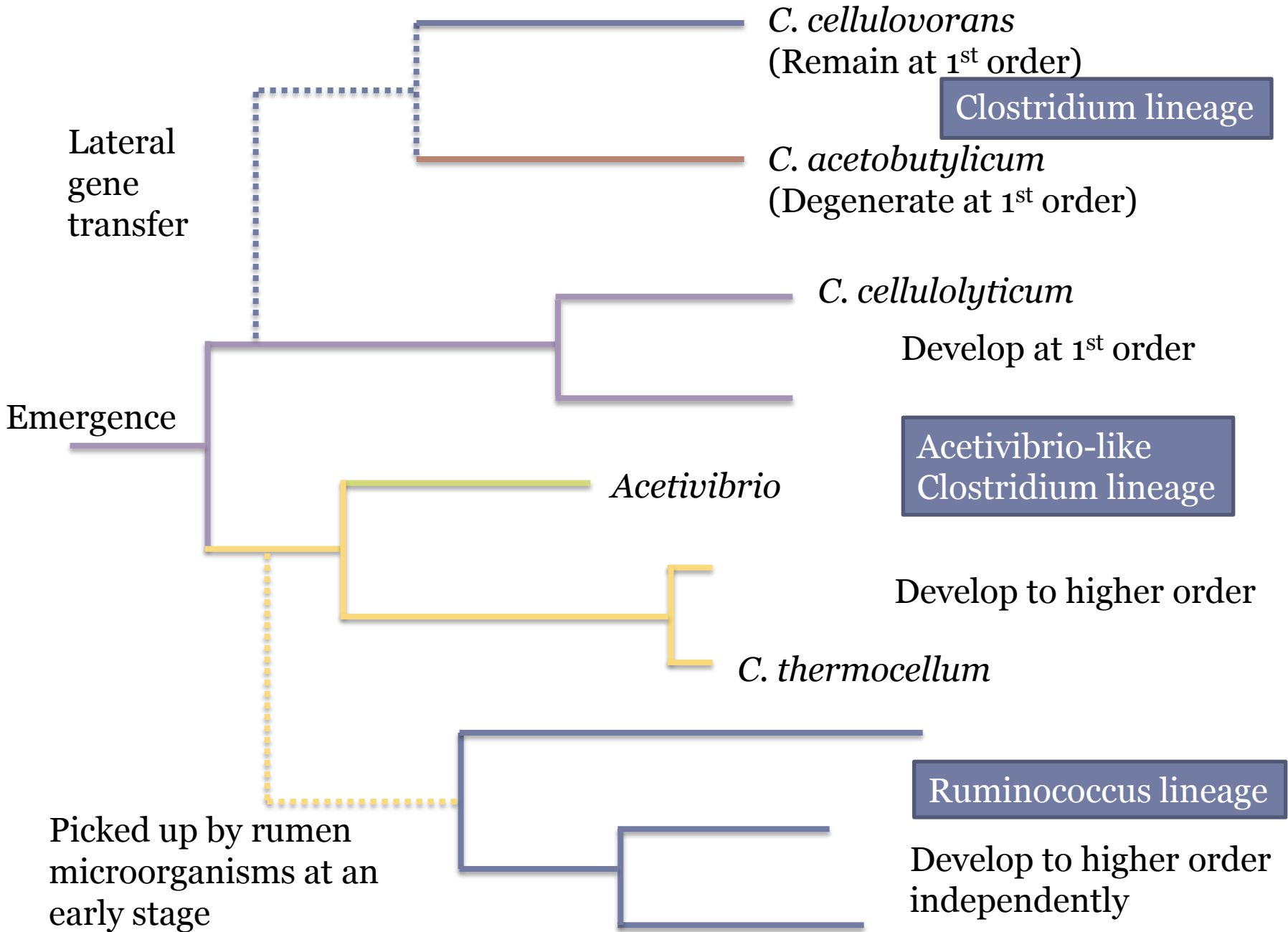


## Clostridium acetobutylicum ATCC 824 plasmid pSOL1

Gene duplication (11)  
 Domain shuffling (5 gains, 4 losses)  
 Gene gain/loss (1 gain, 3 losses)



How modular evolution produces functional strength?



# Emergent properties of the cellulosome

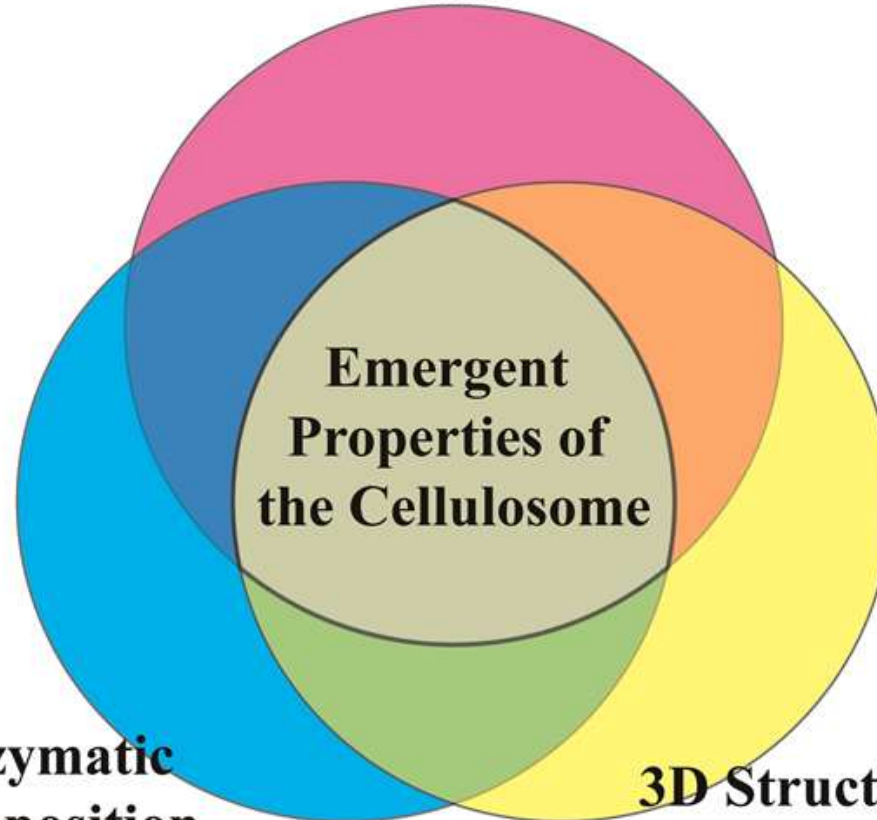
## Cellulosome Gene Diversity

Recombination

LGT

Gene Duplication/Loss

Selection



## Enzymatic Composition of Cellulosome

Phylogeny

Regulatory Networks

Chemical Composition of Biomass

## 3D Structure of Cellulosome

# Scaffoldins

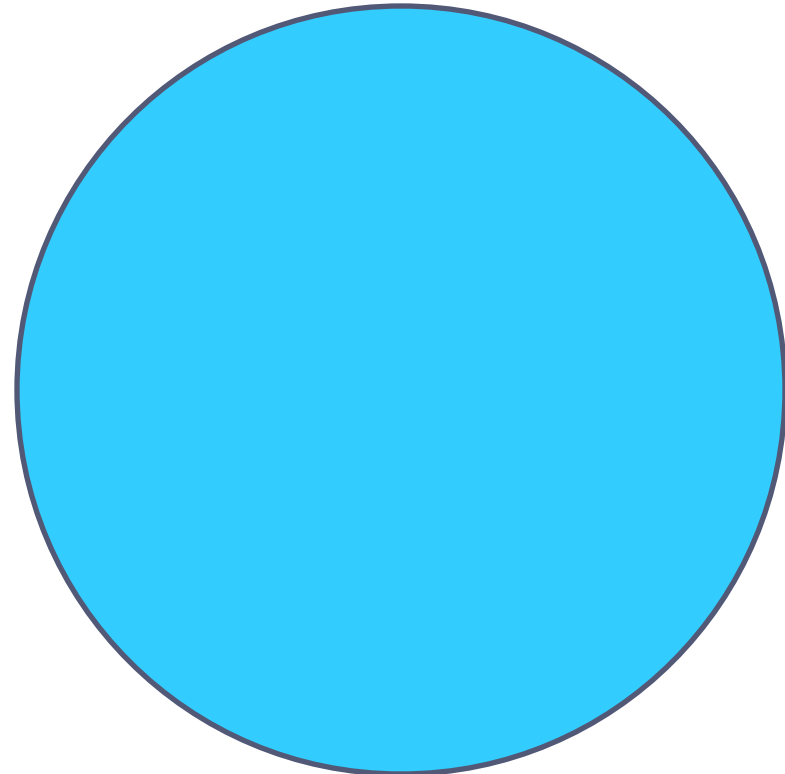
# Cohesin/Dockerin Pairs

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



# Emergent properties of the cellulosome

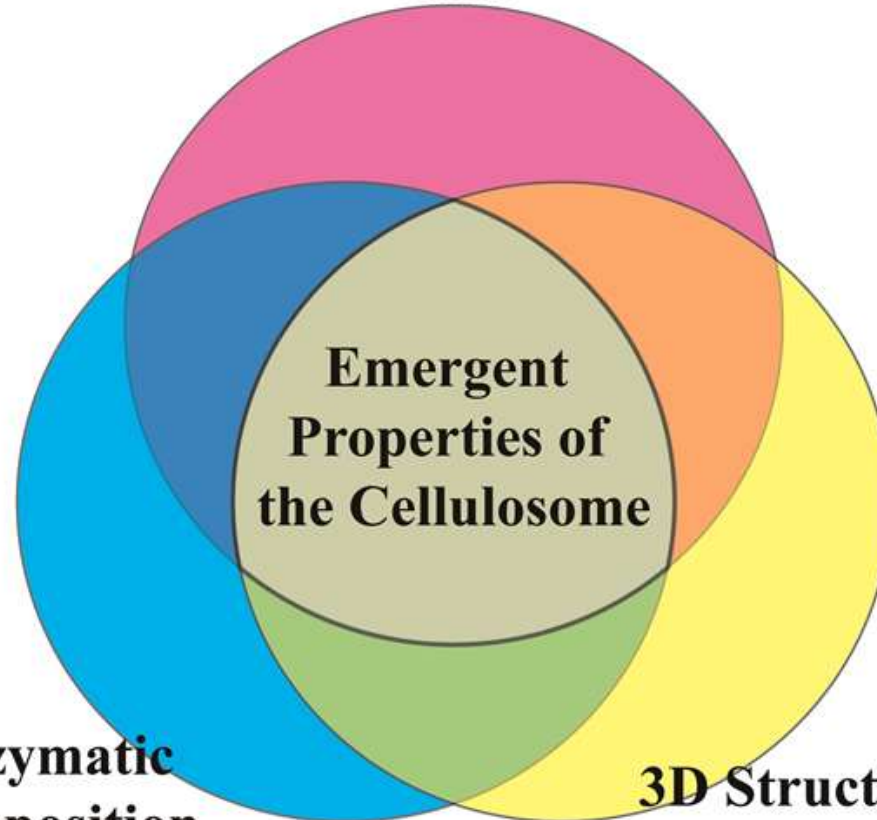
## Cellulosome Gene Diversity

Recombination

LGT

Gene Duplication/Loss

Selection



## Enzymatic Composition of Cellulosome

Phylogeny

Regulatory Networks

Chemical Composition of Biomass

## 3D Structure of Cellulosome

# Scaffoldins

# Cohesin/Dockerin Pairs

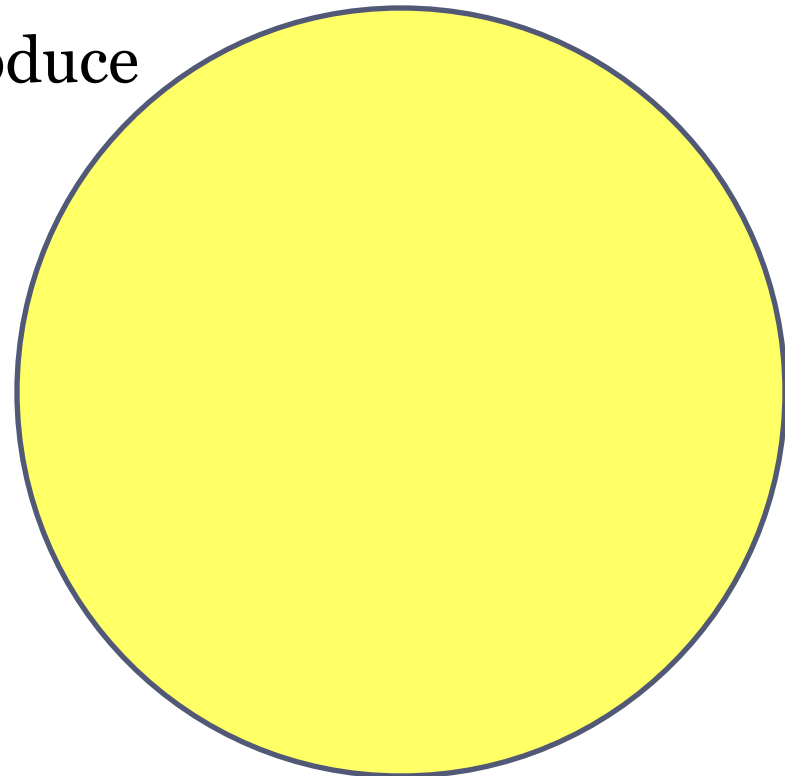
Localization of Enzymes

Physical and Biochemical Limits

# 3D structure of the cellulosome

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

(How does modular evolution produce structural complexity?)

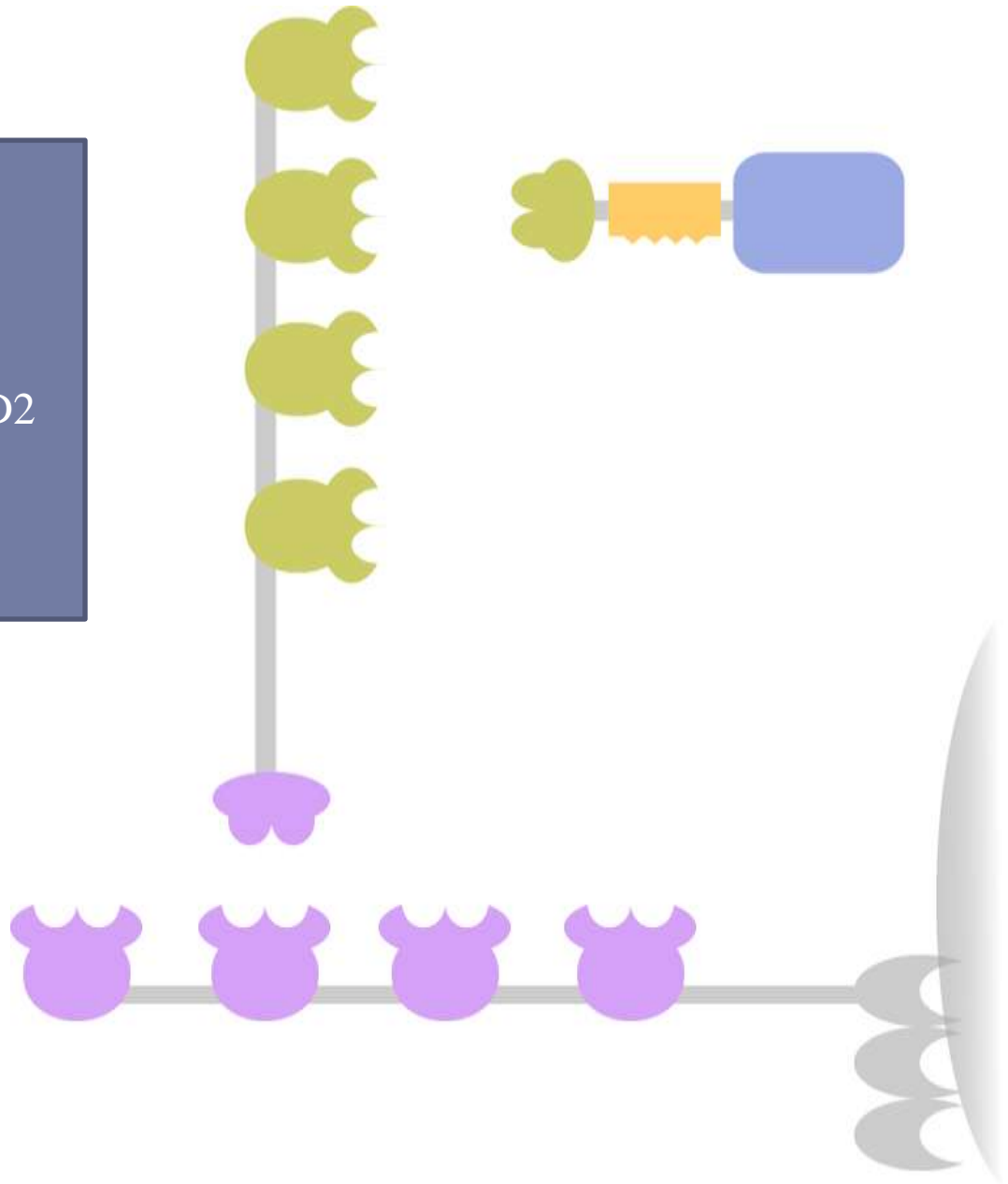


*Clostridium thermocellum*  
ATCC 27405

Second order:  $7 \times 9 = 63$

*Acetivibrio cellulolyticus* CD2

Third order:  $3 \times 4 \times 7 = 84$



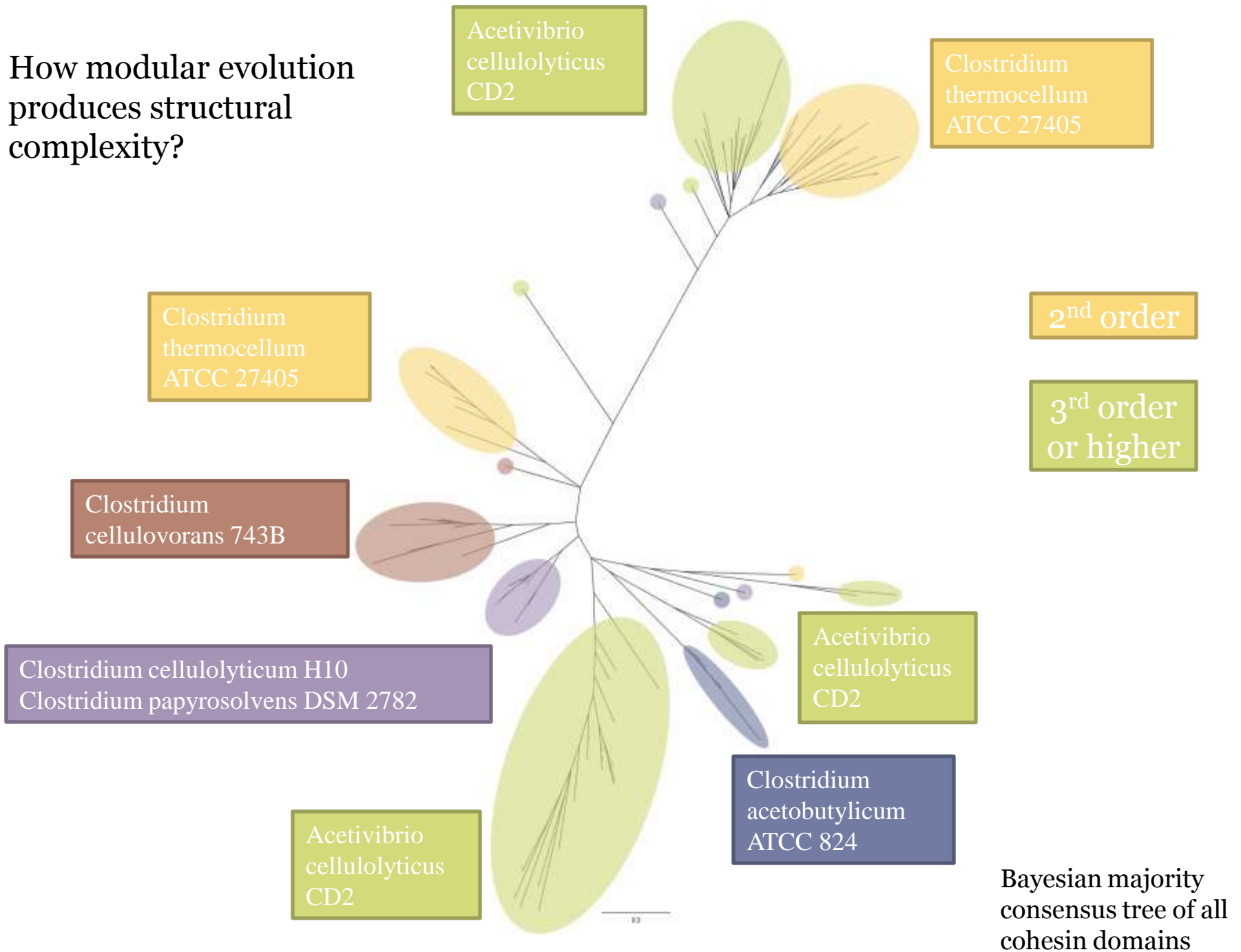
Second order cellulosome



# # of scaffoldins and cohesin domains

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

# How modular evolution produces structural complexity?



# Emergent properties of the cellulosome

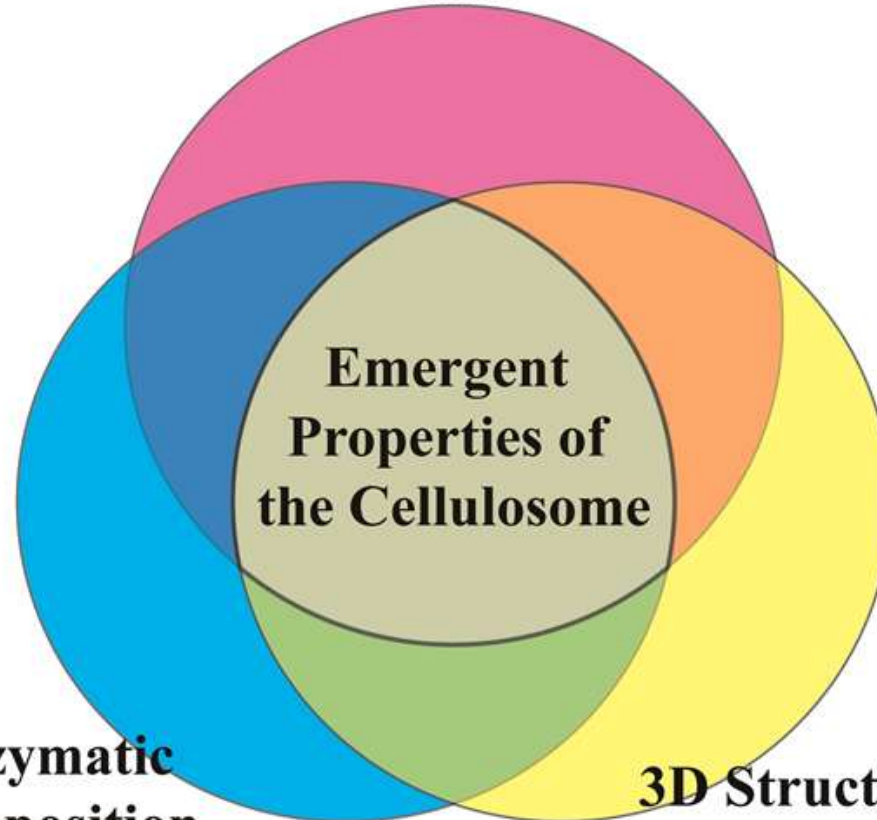
## Cellulosome Gene Diversity

Recombination

LGT

Gene Duplication/Loss

Selection



## Enzymatic Composition of Cellulosome

Phylogeny

Regulatory Networks

Chemical Composition of Biomass

## 3D Structure of Cellulosome

# Scaffoldins

# Cohesin/Dockerin Pairs

Localization of Enzymes

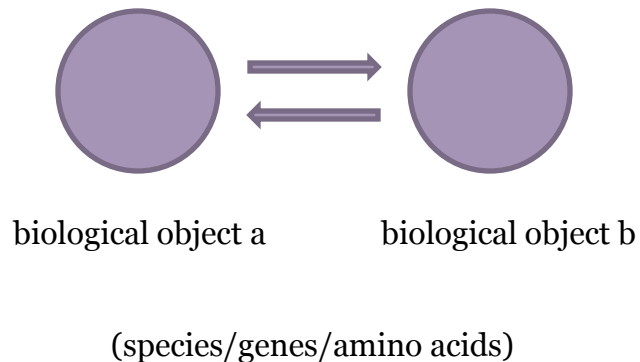
Physical and Biochemical Limits

# Questions

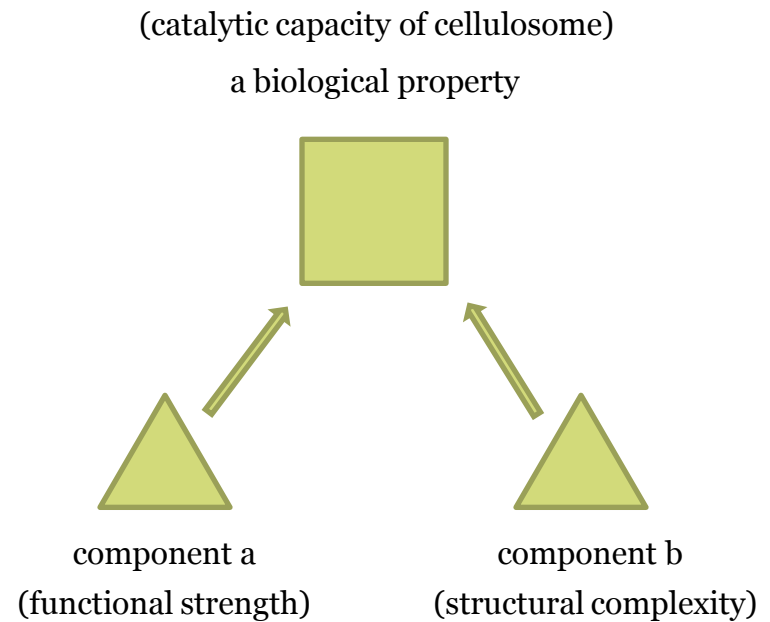
- How modular evolution produces functional strength?
- How modular evolution produces structural complexity?
- What's the correlation between functional strength and structural complexity, and how does natural selection drive the co-evolution?
- How does the cellulosome evolve in different lineages?

# Modes of co-evolution

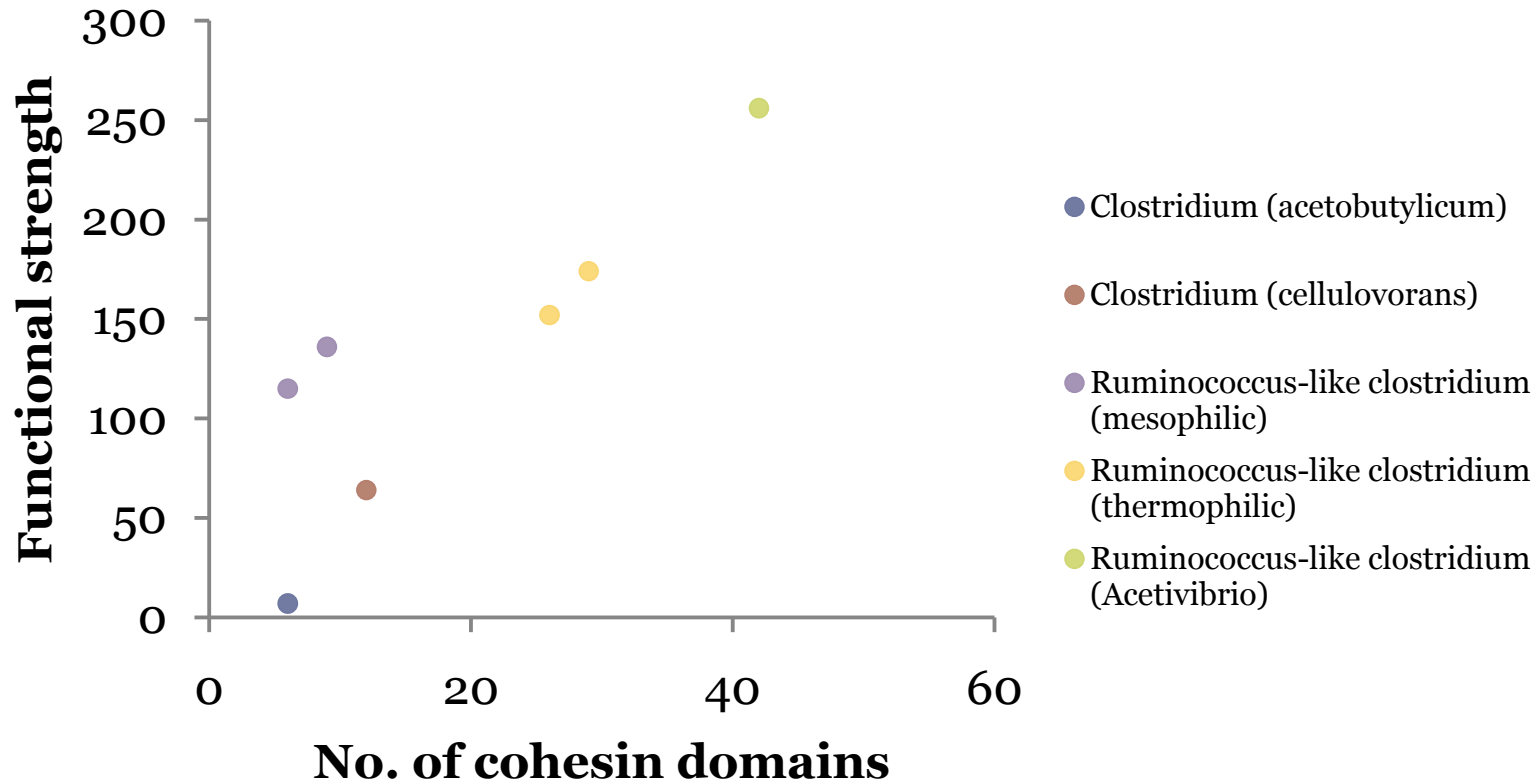
Co-evolution between two objects that interact with each other



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



# Correlation between structural complexity and functional strength



1<sup>st</sup> order

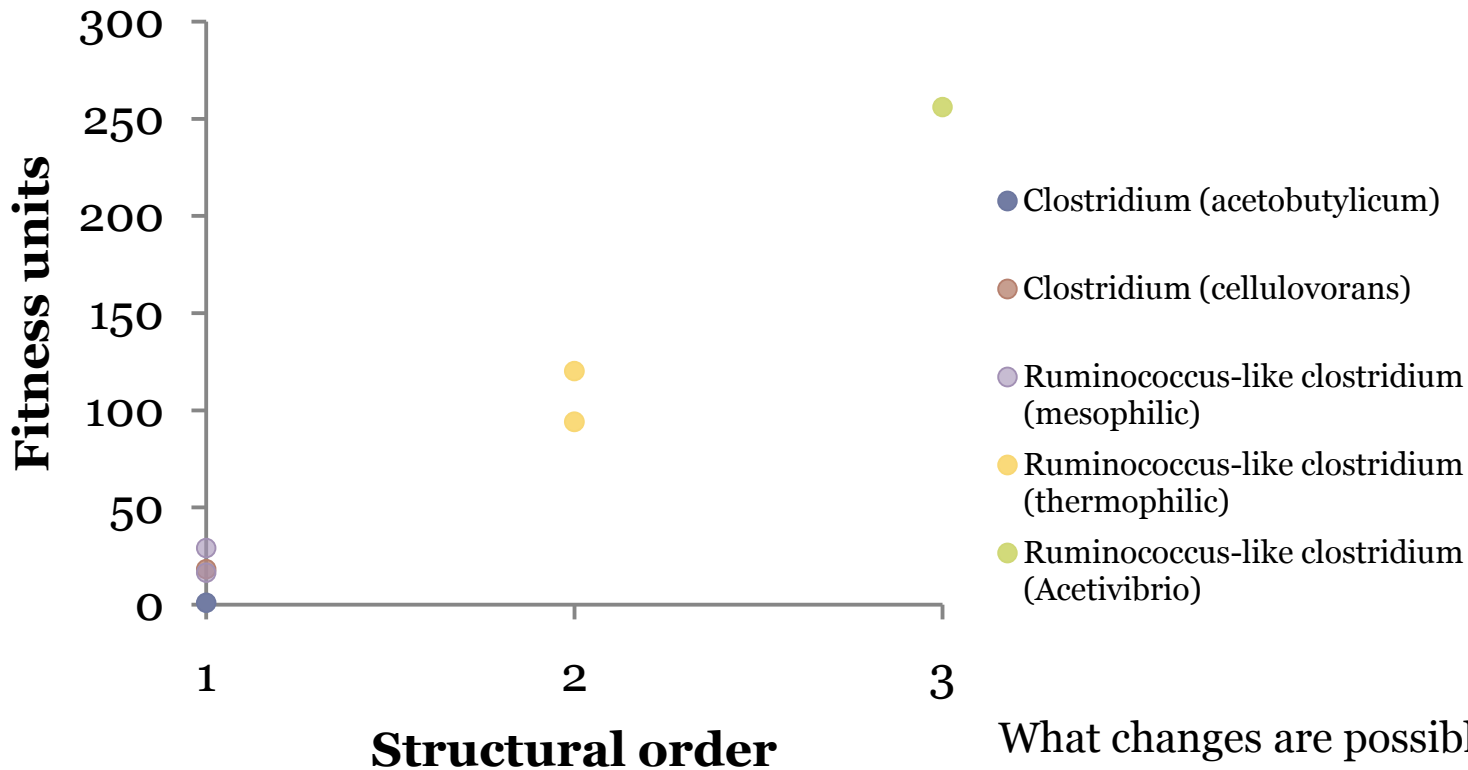
2<sup>nd</sup> order

3<sup>rd</sup> order

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



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

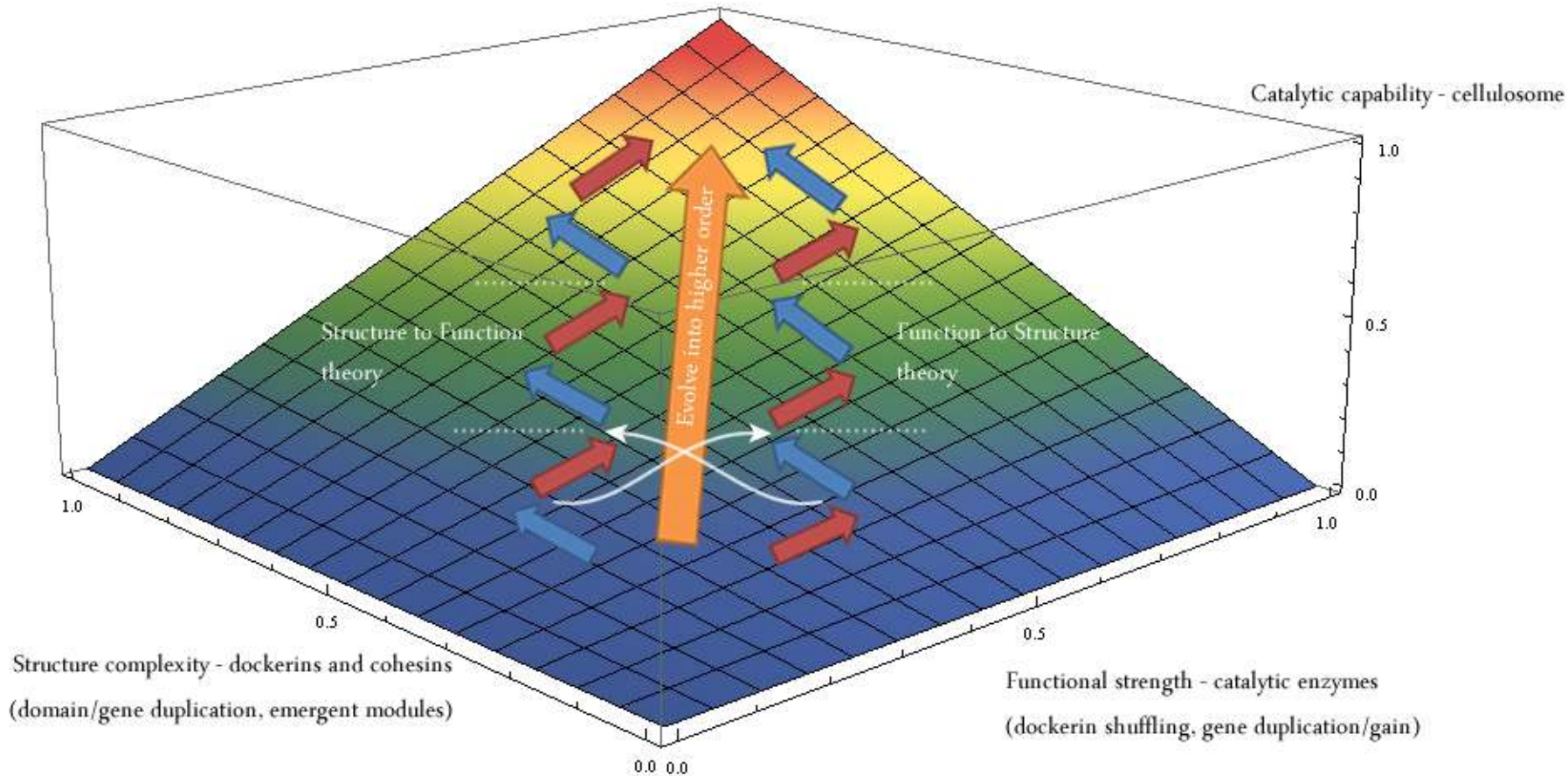
What changes are possible depends on the pre-existing structure of the cellulosome. Without increase in structural order, fitness can only increase within certain limits, which are constrained by the underlying existing states.

## 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



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