Outline

• Introduction
• Protein Network Analysis
• Metabolic Network Analysis
• Summary
Model Organism

- *Escherichia coli K-12*
Collaboration

E. coli Systems Biology

- Hirotada Mori
  - Nara Institute of Science and Technology (NAIST) 奈良先端大学
  - Institute of Advanced Biosciences, Keio University (TTCK) 慶應大学

- Systematic Approach for E. coli functional genomics

- Large-scale experimental measurements
  - Transcriptome
  - Proteome
  - Metabolome
E. coli Systems Biology

Toward complete understanding a cell

Complete genome sequence → Fundamental knowledge

Clones → Molecular mechanism

Gene disruption → Gene Function

DNA microarray, Protein array etc. → Gene network, Protein-protein interaction

Bioinformatics → Analysis, Prediction

Escherichia coli
What’s Essential Gene?

• An essential gene is one, when knocked out, renders the cell un-viable.
Experimental Design

ASKA Plasmid Clone Library

PCR amplification

DNA microarray
Protein localization

Functional analysis

Proteins

3D structure
Protein-protein interaction
Experimental Design

Deletion mutant (KO collection, Knock Out & KeiO)

47 b + 3 b (ATG) + 20 b (KmR)

21 bp

20 b (KmR) + 21 b (SD + ter) + 29 b

34 a.a. (102 bp)
MIPGIRRPAVRSSSTLSIGTSKQLQPT+X(6aa)X
Growth of Deletion Mutants

Growth of deletion mutants in LB

Time (h)

Growth (O.D. = 600nm)

- wild
- galM
- glk
- pgi
- pfkB
- pfkA
- pfp
- fbaB
- tpiA
- gapC
- gpmB
- pykA
- pykE
- ppsA

Growth of deletion mutants in LB
## Gene Knock-out Results

<table>
<thead>
<tr>
<th>Total IS, etc.</th>
<th>Targeted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes 4432</td>
<td>150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validated (out of 8)</th>
<th>Ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genes 7</td>
<td>1843</td>
</tr>
<tr>
<td>%</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survival-Essentiality (%)</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3146</td>
</tr>
<tr>
<td>%</td>
<td>73.5</td>
</tr>
<tr>
<td></td>
<td>495</td>
</tr>
<tr>
<td></td>
<td>543</td>
</tr>
<tr>
<td></td>
<td>12.7</td>
</tr>
<tr>
<td></td>
<td>328</td>
</tr>
<tr>
<td></td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>303</td>
</tr>
<tr>
<td></td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>808</td>
</tr>
<tr>
<td></td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KO collection</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3972</td>
<td>4129</td>
</tr>
<tr>
<td>-</td>
<td>303</td>
</tr>
</tbody>
</table>

**Gene A**
- **FRT**
- **geneA**
- **k1**
- **upstream**

**Gene B**
- **k2**
- **FRT**
- **Kanamycin resistance gene**
- **downstream**

**Gene C**
- **C1**
- **geneC**
- **C2**

### Genomic PCR Validation
- **A1**
- **A2**
- **C1**
- **C2**
## Essential Genes

### Knock-out Efficiency (%)

<table>
<thead>
<tr>
<th>No. genes</th>
<th>Knock-out efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1946</td>
<td>100</td>
</tr>
<tr>
<td>729</td>
<td>87.5</td>
</tr>
<tr>
<td>499</td>
<td>75.0</td>
</tr>
<tr>
<td>316</td>
<td>62.5</td>
</tr>
<tr>
<td>219</td>
<td>50.0</td>
</tr>
<tr>
<td>116</td>
<td>37.5</td>
</tr>
<tr>
<td>160</td>
<td>25.0</td>
</tr>
<tr>
<td>1</td>
<td>12.5</td>
</tr>
<tr>
<td>302</td>
<td>0</td>
</tr>
</tbody>
</table>

### PCR Verification of Deletion Mutants

<table>
<thead>
<tr>
<th>Upstream</th>
<th>Downstream</th>
</tr>
</thead>
<tbody>
<tr>
<td>geneA</td>
<td>Kanamycin resistance gene</td>
</tr>
<tr>
<td>A1</td>
<td>k1</td>
</tr>
<tr>
<td>A2</td>
<td>k2</td>
</tr>
<tr>
<td>FRT</td>
<td>FRT</td>
</tr>
<tr>
<td>C1</td>
<td>geneC</td>
</tr>
<tr>
<td>k1</td>
<td>C2</td>
</tr>
<tr>
<td>k2</td>
<td></td>
</tr>
</tbody>
</table>

- **Gene A**
  - Non-essential regions: 1, 2, 3, 4, 5, 6, 7, 8
  - Essential regions: 1, 2

- **Gene C**
  - Non-essential regions: 1, 2, 3
  - Essential regions: 1, 2, 3, 4, 5, 6, 7, 8

- **Kanamycin resistance gene**

- **FRT**

- **C1**

- **C2**

- **M**
  - A1 - A2 - k1 - k2 - FRT - FRT - C1 - C2
Essential Genes

<table>
<thead>
<tr>
<th>COG Classification</th>
<th>Essential Genes</th>
<th>Total Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>C Energy production and conversion</td>
<td>6</td>
<td>269</td>
</tr>
<tr>
<td>D Cell division and chromosome partitioning</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>E Amino acid transport and metabolism</td>
<td>6</td>
<td>378</td>
</tr>
<tr>
<td>F Nucleotide transport and metabolism</td>
<td>11</td>
<td>91</td>
</tr>
<tr>
<td>G Carbohydrate transport and metabolism</td>
<td>9</td>
<td>341</td>
</tr>
<tr>
<td>H Coenzyme metabolism</td>
<td>15</td>
<td>288</td>
</tr>
<tr>
<td>I Lipid metabolism</td>
<td>26</td>
<td>141</td>
</tr>
<tr>
<td>J Translation, ribosomal structure and biogenesis</td>
<td>24</td>
<td>92</td>
</tr>
<tr>
<td>K Transcription</td>
<td>77</td>
<td>170</td>
</tr>
<tr>
<td>L DNA replication, recombination and repair</td>
<td>26</td>
<td>141</td>
</tr>
<tr>
<td>M Cell envelope biogenesis, outer membrane</td>
<td>24</td>
<td>92</td>
</tr>
<tr>
<td>N Cell motility and secretion</td>
<td>6</td>
<td>269</td>
</tr>
<tr>
<td>O Posttranslational modification, protein turnover, chaperones</td>
<td>19</td>
<td>34</td>
</tr>
<tr>
<td>P Inorganic ion transport and metabolism</td>
<td>15</td>
<td>288</td>
</tr>
<tr>
<td>Q Secondary metabolites biosynthesis, transport and catabolism</td>
<td>14</td>
<td>235</td>
</tr>
<tr>
<td>R General function prediction only</td>
<td>11</td>
<td>151</td>
</tr>
<tr>
<td>S Signal transduction mechanisms</td>
<td>9</td>
<td>132</td>
</tr>
<tr>
<td>T Function unknown (S, U, V)</td>
<td>1</td>
<td>250</td>
</tr>
<tr>
<td>U Not in COG</td>
<td>1</td>
<td>250</td>
</tr>
<tr>
<td>V Not in COG</td>
<td>5</td>
<td>93</td>
</tr>
<tr>
<td>W Not in COG</td>
<td>15</td>
<td>446</td>
</tr>
<tr>
<td>X Not in COG</td>
<td>9</td>
<td>172</td>
</tr>
<tr>
<td>Y Not in COG</td>
<td>18</td>
<td>444</td>
</tr>
<tr>
<td>Z Not in COG</td>
<td>26</td>
<td>1166</td>
</tr>
</tbody>
</table>
E. coli K-12

Cellular Components

- intracellular: 36%
- external encapsulating structure: 12%
- membrane: 18%
- organelle: 12%
- protein complex: 13%

All: 44%
Biological Processes

- protein metabolism: 30%
- nucleotide and nucleic acid metabolism: 27%
- carbohydrate metabolism: 12%
- development: 10%
- amino acid and derivative metabolism: 10%
- lipid metabolism: 9%
- regulation of physiological process: 7%
- response to stimulus: 4%

Total: 100%
Overlap of essential genes and orthologs

- **E. coli K-12**
  - 4,432 genes
  - Essential: 303 genes
  - Orthologs: 2,985 genes

- **B. subtilis**
  - 4,105 genes
  - Essential: 271 genes
  - Orthologs: 2,575 genes

- **Overlap:**
  - 1,447/1,530 orthologs
  - 67 genes
  - 75 essential genes
  - 150 genes
  - 46 essential genes

- **Total:**
  - 1,447/1,530 orthologs
  - 150 genes
  - 75 essential genes
  - 46 essential genes
Minimum Genome!? 

Construction of minimal genome

Non-essential gene: 4126  
Essential gene: 303  
128 regions; 3,243 kb  
(70% of genome)  
18 regions; 304 kb  
(7% of genome)  

Synthetic lethal?  
284 genes
Why “Essential”?

- From aspects at Systems level
- From the **Network** point of view
  - Protein-protein interaction network
  - Metabolic network
Protein Interaction Network
Experimental Measurements

Identification of protein-protein interaction

Cell growth
IPTG induction
Sonication
Affinity purification by Ni$^{2+}$-NTA
SDS-PAGE
MALDI
Protein identification
Analysis

Arifuzzaman et al.
Experimental Measurements

- **Success to be purified**
  - Cyt: 2,184 ORFs
  - Mem: 153

- **Fail to be purified**
  - Cyt: 640
  - Mem: 928

- **Number of putative complex observed**
  - 10,816
  - 695
  - **Total: 11,511**

- **No interaction observed**
  - 330 orfs
    - Cytosol (cyt): 298
    - Membrane (mem): 32
E. coli Protein Interaction Network
Network Analysis

Topological Properties

- Degree (connectivity)
- Number of links from the node to others \( (k) \)
- Clustering coefficient
- Connectivity between neighbors

\[
C = \frac{2E}{k(k - 1)}
\]

Network Models

Box 2 | Network models

A Random network

B Scale-free network

C Hierarchical network

\[ P(k) = \frac{\alpha}{k^\gamma} \]

\[ C(k) \sim k^{-\gamma} \]

Connectivity (Degree)

*E. coli* Protein Interaction Network

**Equations:**

- Non-essential genes:
  
  \[ y = 7392 x^{-2.01} \]
  
  \[ R^2 = 0.986 \]

- Essential genes:
  
  \[ y = 252 x^{-1.56} \]
  
  \[ R^2 = 0.914 \]
Protein Interaction Network

- Protein network of E. coli is also scale-free
- Essential genes have a shallower slope ➞ larger proportion of them are hubs
- Essential genes have more than twice links
- Validating earlier finding in yeast

<table>
<thead>
<tr>
<th></th>
<th>Essential</th>
<th>Non-ess.</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; degree &gt;</td>
<td>17.76</td>
<td>6.67</td>
</tr>
<tr>
<td>&lt; clust. coeff. &gt;</td>
<td>0.058</td>
<td>0.064</td>
</tr>
</tbody>
</table>
Yeast

- Essential proteins have ca. twice as many links as non-essential proteins
- Essential proteins have a shallower slope
- **Difference** in clustering coefficients

![Graph showing relationship between log number of proteins and log degree of a network.](image)

<table>
<thead>
<tr>
<th></th>
<th>Average degree (K)</th>
<th>Clustering coefficient (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essential</td>
<td>18.7</td>
<td>0.182</td>
</tr>
<tr>
<td>Synthetic lethal</td>
<td>9.2</td>
<td>0.083</td>
</tr>
<tr>
<td>Non-essential</td>
<td>7.4</td>
<td>0.095</td>
</tr>
<tr>
<td><em>P</em>-value</td>
<td>&lt;10^{-12}</td>
<td>&lt;10^{-12}</td>
</tr>
</tbody>
</table>

Metabolic Network
Metabolic Pathways

New Annotation

- KEGG pathway / reaction database
- Follow KEGG reference reactions

Newly annotated E. coli K-12 genes

E. coli Annotation Workshop II, March 19-24, 2005

729 genes

KEGG

117

new

32

1057 genes w/ EC

296
Essential Genes in Metabolic Pathways

- 153 essential genes as enzymes in metabolic reactions

![Venn Diagram]

- 153 essential genes with EC
- 30 new
- 123
- 3
## Essential Genes in Metabolic Pathways

<table>
<thead>
<tr>
<th>Unique enzyme</th>
<th>Essential</th>
<th>Non-ess.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique enzyme</td>
<td>88.8%</td>
<td>51.6%</td>
</tr>
<tr>
<td>Unique products</td>
<td>9.3%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

- Large portion of essential genes are unique enzymes in metabolic reactions.
- Essential genes tend to be un-replaceable edges and nodes.
### Frequency of counts of compounds & reactions in *E. coli*

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>1000</td>
</tr>
</tbody>
</table>

**Power-law**

- **Compounds**
  
  \[ y = 136.9 x^{-1.05} \]
  
  \[ R^2 = 0.761 \]

- **Reactions**
  
  \[ y = 892 x^{-2.42} \]
  
  \[ R^2 = 0.887 \]
Analysis of “Affected” Network by Gene Deletion
Affected Metabolites

- Affected number of metabolites by deletion of essential genes is significantly higher.
Affected Reactions

- Affected number of metabolic reactions by deletion of essential genes also tend to be higher.
Summary

• Essential genes have significantly more links than the non-essential genes in protein interaction network.

• Essential genes tend to be unreplaceable in metabolic network.

• Affected metabolites and reactions by in silico deletion of essential genes are significantly higher than non-ess.
Outlook

• More topological properties will be investigated
• Flux balance analysis for metabolic network
• Relationship to gene expression profiles
• Evolutionary roles of essential genes
Acknowledgement

- Yue-Yang Bao (鮑岳洋)
- Jen-Yun Chang (張人勳)
- Yih-Chii Hwang (黃奕綺)
- Chen-Ching Lin (林振慶)

- Prof. Hirotada Mori (NAIST, Japan)
- Prof. Hsueh-Fen Juan (NTU Life-Sci)
Thank YOU!

Story continued ...
The 7th International Conference on Systems Biology
ICSB 2006

October 9-13, 2006
in Yokohama, Japan
Pacifico Yokohama

http://icsb-2006.org/
陽明大學
生物醫學資訊研究所
專業研究型大學

數理組：歡迎數理及工程背景學生加入
系統生物學的研究