

Evaluation of Functional Gene for Classification of Pathogenic Bacteria

Department of Microbiology
Regeneration and Advanced Medical Science ,
Gifu University, Graduate school of Medicine
Yanagido 1-1, Gifu, Japan

MLST (multilocus sequence typing) analysis information is rapidly accumulated, but we need more convincing data to satisfy medical community and veterinarian to change BSL3 classification.

- It is no doubt that DNA based identification and detection of pathogens have a powerful potential to control infectious diseases because of the rapiness and sensitiveness.

However, current taxonomy of BSL3 pathogens have many problems to use genetic identification and detection methods.

Problems of classification of BSL₃ Pathogens

Examples:

1. *Bacillus cereus* group: *B. anthracis*, *B. cereus*, *B. thuringiensis*, and *B. mycooides* carry identical 16S rRNA sequences.

They were classified different organisms because of their pathogenic factors.

2. *Escherichia coli* and *Shigella species* shares identical 16S rRNA. They were classified different organisms because of their pathogenic factors.

3. *Clostridium botulinum* species is clusters of phylogenetically different organisms.

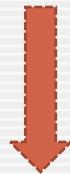
They only shares common feature to produce botulinum toxins.

Historical Errors

Strain which does not produce gas from glucose, lactose negative and nonmotile and etc.



Classied as *Shigella* species and serologically identified as *Shigella* species



No

Biochemically identified as *Escherichia group*



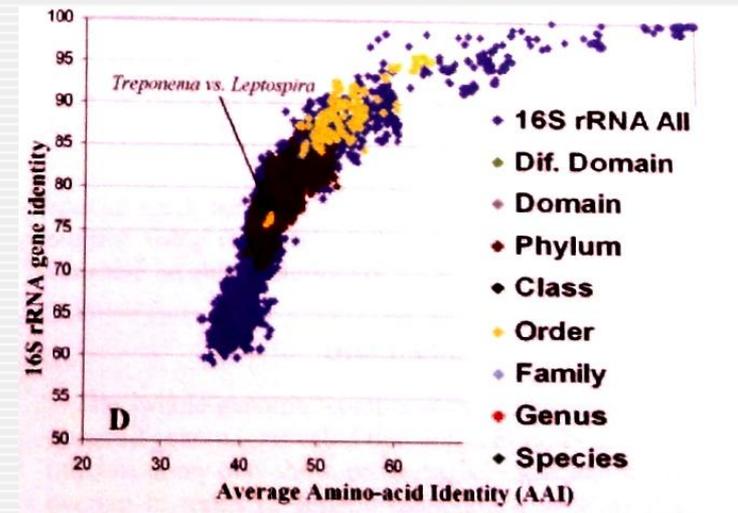
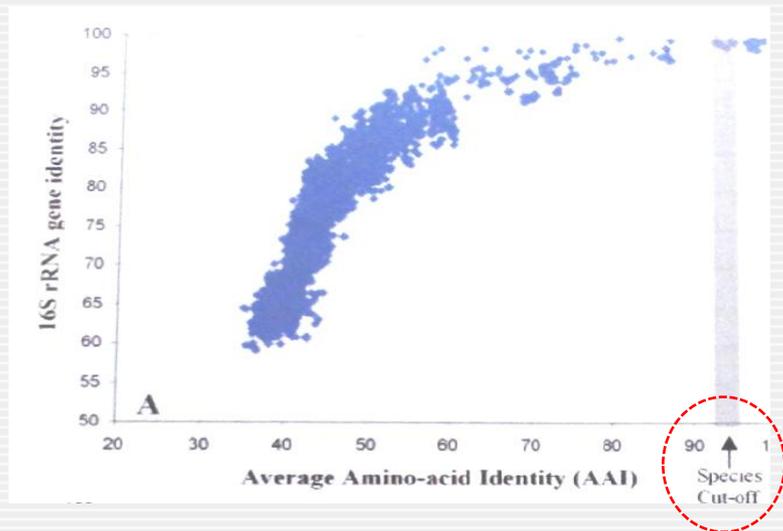
Later, none motile, lactose negative, were found and described as *Pathogenic E.coli*

***E.coli* O antigen: O1 to O180**

- E.coli O7 = S.boydii 12
- E.coli 28a,c = S.boydii 13
- E.coli O112a,c = S.dysenteriae 2
- E.coli O124 = S.dysenteriae 3
- E.coli O143 = S.boydii 8
- E.coli O144 = S.dysenteriae 10
- E.coli O152 = S.dysenteriae 12

Average 16S rRNA SNP within a species is less than 1.3% (Constantinidis and Tiedje, 2005)

Species
cut off



16S rDNA cannot differentiate BSL3 pathogens and Plant pathogens from their closely related species because their sequence are related more than 99.5%.

Need complete protein sequence analysis to use housekeeping genes as a taxonomic tool

Whole genome sequenced strains(1230str.)

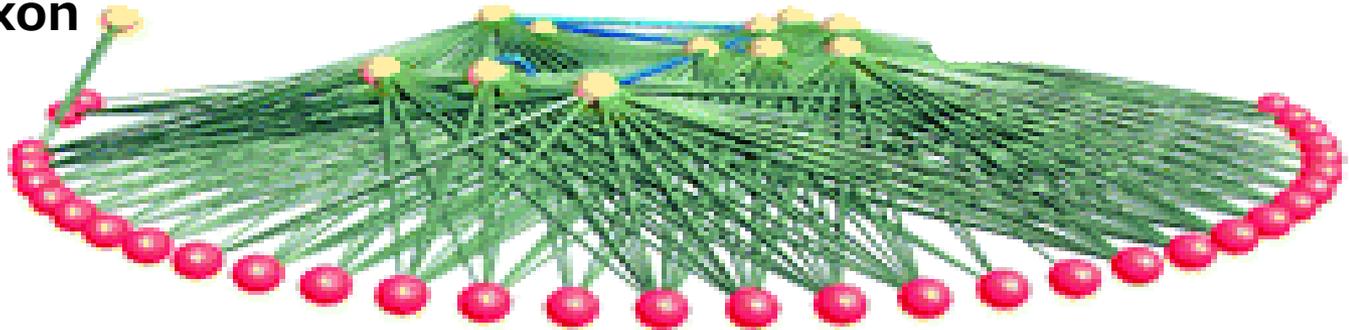
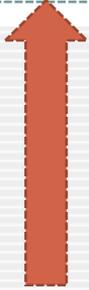
(Million species>7000species>1400genus>280Family)

- 1. Only 500 dollar to determine whole genome of one strain.**
- 2. Shotgun sequencing produced many full sequenced genes and protein information. Strain level information is bursting.**

In medical microbiology, pathogenic species are well defined and their differentiation methods from non-pathogenics are well established in detail through 100 year history of pathogenic bacterial history.

Cannot reach to Common ancestor

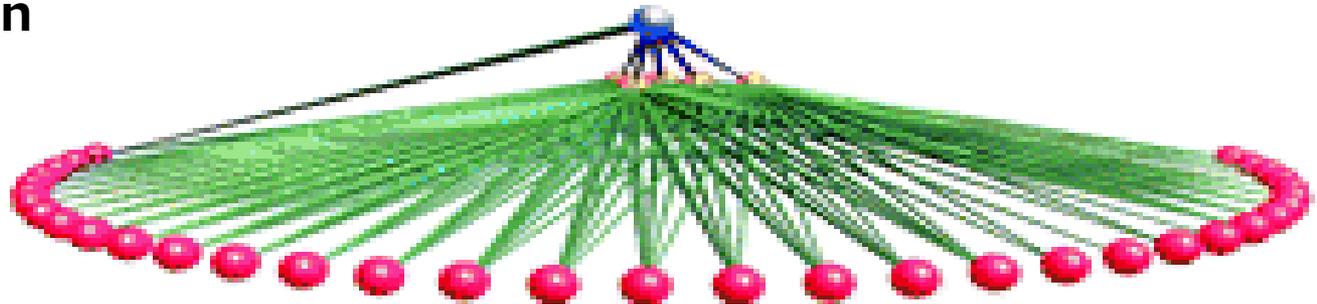
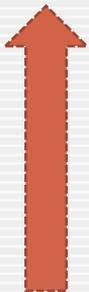
Higher taxon



Species level :pick up only specific genes, pathogenic factors among species

Common ancestor

Higher taxon



Species level: compare commonly conserved house keeping genes

Commonly Conserved Genes among whole genome sequenced strains

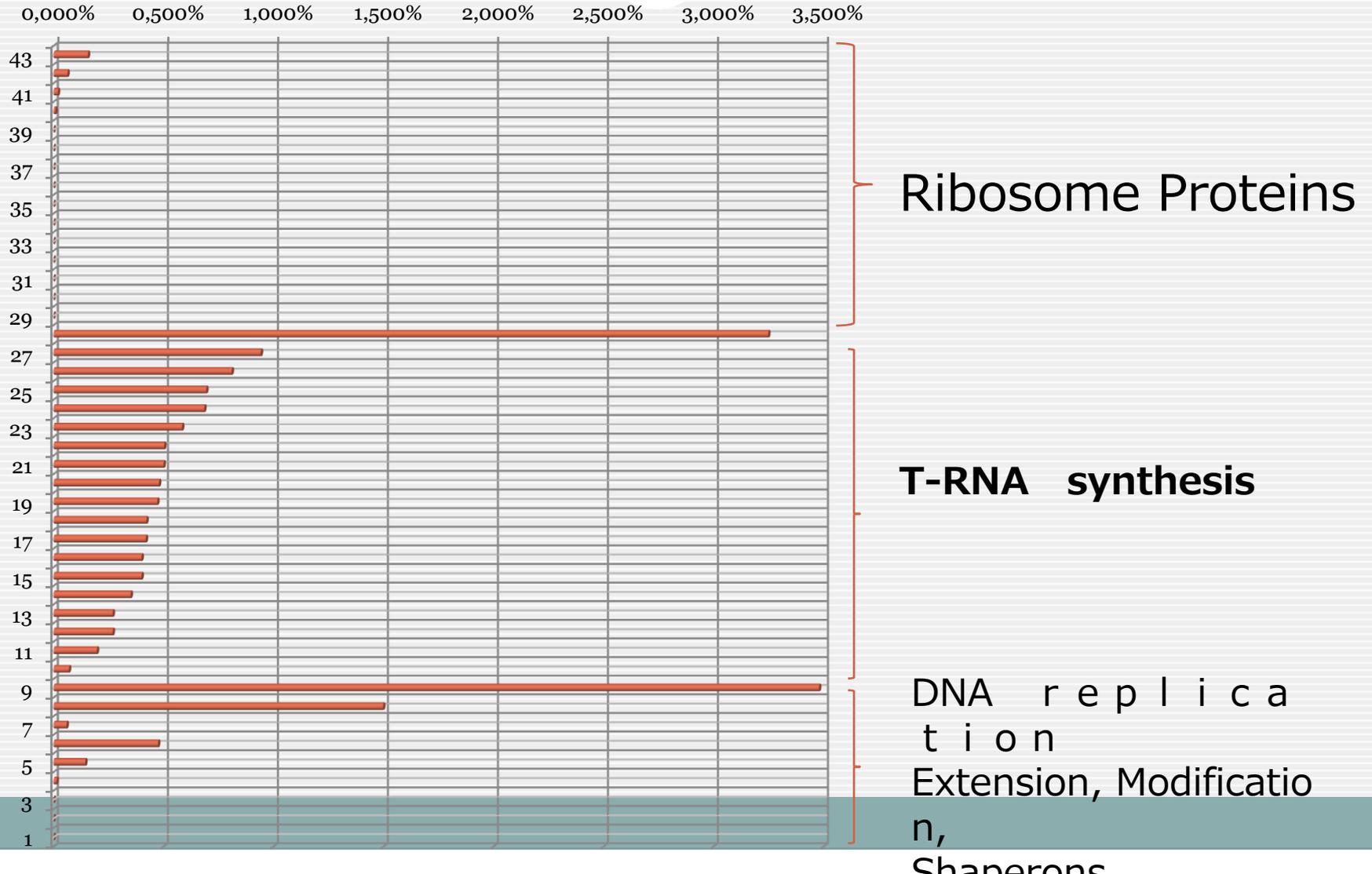
| | |
|---------------------------------------------------------------------------------------------|-----|
| UJ COG0048 Ribosomal protein S12 | 124 |
| UJ COG0049 Ribosomal protein S7 | 32 |
| UJ COG0051 Ribosomal protein S10 | 103 |
| UJ COG0052 Ribosomal protein S2 | 241 |
| UJ COG0088 Ribosomal protein L4 | 201 |
| UJ COG0089 Ribosomal protein L23 | 100 |
| UJ COG0090 Ribosomal protein L2 | 273 |
| UJ COG0091 Ribosomal protein L22 | 110 |
| UJ COG0100 Ribosomal protein S11 | 129 |
| UJ COG0102 Ribosomal protein L13 | 142 |
| UJ COG0103 Ribosomal protein S9 | 130 |
| UJ COG0244 Ribosomal protein L10 | 165 |
| UJ COG0255 Ribosomal protein L29 | 63 |
| UJ COG0256 Ribosomal protein L18 | 117 |
| UJ COG0231 Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A) | 188 |
| UJ COG0361 Translation initiation factor 1 (IF-1) | 72 |
| UJ COG0532 Translation initiation factor 2 (IF-2; GTPase) | 890 |
| LJ COG0550 Topoisomerase IA | 653 |
| UJ COG0124 Histidyl-tRNA synthetase | 414 |
| UJ COG0143 Methionyl-tRNA synthetase | 677 |
| UJ COG0162 Tyrosyl-tRNA synthetase | 424 |
| UJ COG0441 Threonyl-tRNA synthetase | 642 |
| UJ COG0442 Protyl-tRNA synthetase | 572 |
| UJ COG0495 Leucyl-tRNA synthetase | 891 |
| UJ COG0522 Ribosomal protein S4 | 206 |
| UJ COG0525 Valyl-tRNA synthetase | 951 |
| EJ COG0006 Xaa-Pro aminopeptidase1 | 361 |
| EJ COG0006 Xaa-Pro aminopeptidase2 | 441 |
| OJ COG0492 Thioredoxin reductase | 321 |
| EJ COG0112 Glycine/serine hydroxymethyltransferase | 417 |
| FJ COG0504 CTP synthase (UTP-ammonia lyase) | 545 |
| GJ COG0126 3-phosphoglycerate kinase | 387 |
| GJ COG0148 Enolase | 432 |
| GJ COG0149 Triosephosphate isomerase | 255 |
| GJ COG1109 Phosphomannomutase | 456 |
| UJ COG0030 Dimethyladenosine transferase (rRNA methylation) | 273 |
| glycine tRNA synthetase, alpha subunit | 303 |
| glycine tRNA synthetase, beta subunit | 889 |
| UJ COG0008 Glutaminyl-tRNA synthetases | 471 |
| UJ COG0013 Alanyl-tRNA synthetase | 876 |
| UJ COG0016 Phenylalanine tRNA synthetase, alpha subunit | 327 |
| UJ COG0072 Phenylalanine tRNA synthetase, beta subunit | 795 |
| UJ COG0018 Arginyl-tRNA synthetase | 577 |
| UJ COG0024 Methionine aminopeptidase | 264 |
| UJ COG0060 Isoleucyl-tRNA synthetase | 938 |

**Ribosome
Proteins,**

**Replication,
Modification,
Protein maturation**

**T-RNA
Synthesis**

House keeping Protein SAP in *Escherichia coli* (N=44strains)



Ribosome protein SAP in *E.coli*(N=44 strains)

| Ribosome protein | Amino | Average | Range | SD |
|-----------------------|-------|---------------|---------------|---------------|
| Ribosomal protein L10 | 165 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein L13 | 142 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein L18 | 117 | 0.000% | 0-0.8% | 0.000% |
| Ribosomal protein L2 | 273 | 0.000% | 0-0.3% | 0.000% |
| Ribosomal protein L22 | 110 | 0.000% | 0-0.3% | 0.000% |
| Ribosomal protein L23 | 100 | 0.000% | 0-1.0% | 0.000% |
| Ribosomal protein L4 | 201 | 0.000% | 0-0.5% | 0.000% |
| Ribosomal protein S10 | 103 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein S11 | 129 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein S4 | 206 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein S7 | 32 | 0.000% | 0.000% | 0.000% |
| Ribosomal protein S12 | 124 | 0.009% | 0-0.8% | 0.000% |
| Ribosomal protein L29 | 63 | 0.020% | 0-1.6% | 0.000% |
| Ribosomal protein S9 | 130 | 0.063% | 0-0.8% | 0.219% |
| Ribosomal protein S2 | 241 | 0.156% | 0-0.4% | 0.198% |

Ribosome Proteins

Ribosome protein SNP within a species was less than 1.6%

T-RNA and Replication associated Protein SNP

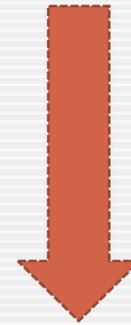
| Escherichia coli 44 株 (COG genes) | Amino | Average | Range | SD |
|----------------------------------------------------------------------------------|--------------|----------------|--------------|-----------|
| Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A) | 188 | 0.000% | 0-0.5% | 0.000% |
| Translation initiation factor 1 (IF-1) | 72 | 0.000% | 0.000% | 0.000% |
| Triosephosphate isomerase | 255 | 0.000% | 0-0.4% | 0.000% |
| Translation initiation factor 2 (IF-2; GTPase) | 890 | 0.014% | 0-0.1% | 0.000% |
| Phosphoglycerate kinase | 387 | 0.144% | 0-0.3% | 0.153% |
| Thioredoxin reductase | 321 | 0.475% | 0-1.2% | 0.321% |
| CTP synthase (UTP-ammonia lyase) | 545 | 0.059% | 0-0.6% | 0.148% |
| Topoisomerase IA | 653 | 1.498% | 0-4.6% | 1.243% |
| Phosphomannomutase | 456 | 3.479% | 0-24.0% | 3.502% |
| Glycine/serine hydroxymethyltransferase | 417 | 0.070% | 0-0.2% | 0.098% |
| Glycine tRNA synthetase | 689 | 0.197% | 0-0.4% | 0.182% |
| Dimethyladenosine transferase (rRNA methylation) | 273 | 0.269% | 0-0.7% | 0.280% |
| Threonyl-tRNA synthetase | 642 | 0.269% | 0-1.1% | 0.084% |
| Valyl-tRNA synthetase | 951 | 0.351% | 0-0.7% | 0.124% |
| Phenylalanine tRNA synthetase, beta subunit | 795 | 0.400% | 0-1.0% | 0.272% |
| Alanyl-tRNA synthetase | 876 | 0.400% | 0-2.7% | 0.185% |
| Methionine aminopeptidase | 264 | 0.419% | 0-1.5% | 0.276% |
| Methionyl-tRNA synthetase | 677 | 0.422% | 0-1.5% | 0.182% |
| Phenylalanine tRNA synthetase, alpha subunit | 327 | 0.471% | 0-0.9% | 0.209% |
| Arginyl-tRNA synthetase | 577 | 0.479% | 0-1.9% | 0.179% |
| Glutamyl-tRNA synthetases | 471 | 0.500% | 0-1.1% | 0.262% |
| Isoleucyl-tRNA synthetase | 938 | 0.503% | 0-1.0% | 0.227% |
| Leucyl-tRNA synthetase | 891 | 0.584% | 0-1.0% | 0.172% |
| Tyrosyl-tRNA synthetase | 424 | 0.685% | 0-4.7% | 0.282% |
| Xaa-Pro aminopeptidase1 | 361 | 0.694% | 0-1.8% | 0.806% |
| Prolyl-tRNA synthetase | 572 | 0.810% | 0-1.2% | 0.203% |
| Histidyl-tRNA synthetase | 414 | 0.942% | 0-6.8% | 0.409% |
| Xaa-Pro aminopeptidase2 | 441 | 3.248% | 0-6.1% | 1.634% |

DNA

T-RNA

| t-RNA | Str.K-12 W3110 | Str.K-12 DH10B | Str.K-12 MG1655 |
|----------|-------------------|-------------------|--------------------|
| Ile-tRNA | 5 | 6 | 6 |
| Ala-tRNA | 5 | 6 | 5 |
| Asp-tRNA | 5 | 2 | 3 |
| Thr-tRNA | 2 | 4 | 4 |
| Arg-tRNA | 4 | 6 | 7 |
| Gln-tRNA | 4 | 4 | 4 |
| Leu-tRNA | 8 | 8 | 7 |
| Met-tRNA | 6 | 7 | 7 |
| Lys-tRNA | 6 | 6 | 6 |
| Val-tRNA | 5 | 7 | 2 |
| Ser-tRNA | 5 | 2 | 6 |
| Tyr-tRNA | 2 | 3 | 3 |
| Val-tRNA | 2 | 2 | 5 |
| Cys-tRNA | 1 | 1 | 1 |
| Gly-tRNA | 5 | 6 | 6 |
| Asn-tRNA | 4 | 4 | 4 |
| Pro-tRNA | 3 | 4 | 4 |
| Glu-tRNA | 3 | 3 | 4 |
| Phe-tRNA | 2 | 2 | 2 |
| His-tRNA | 2 | 1 | 1 |
| Trp-tRNA | 1 | | 1 |
| Sec-tRNA | 1 | | 1 |
| Xaa-tRNA | | | 3 |

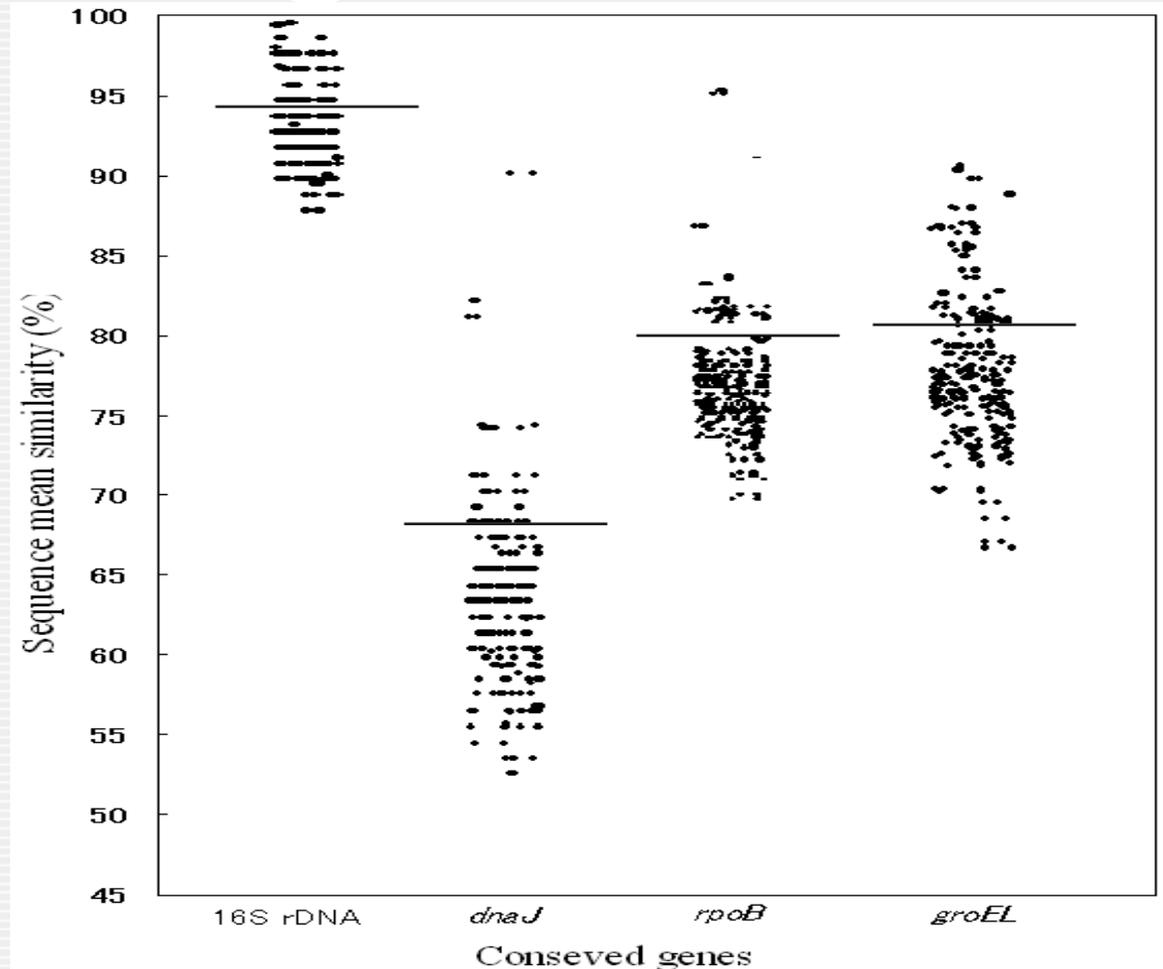
Copy Numbers of T-RNA gene within E.coli K-12 strains



Different among
strain level:

Not suitable for
classification

House keeping genes (*dnaJ*, *rpoB*, *gyrB*) are usually exist only one on their chromosome only exist one copy on their chromosome



Genus *Escherichia*

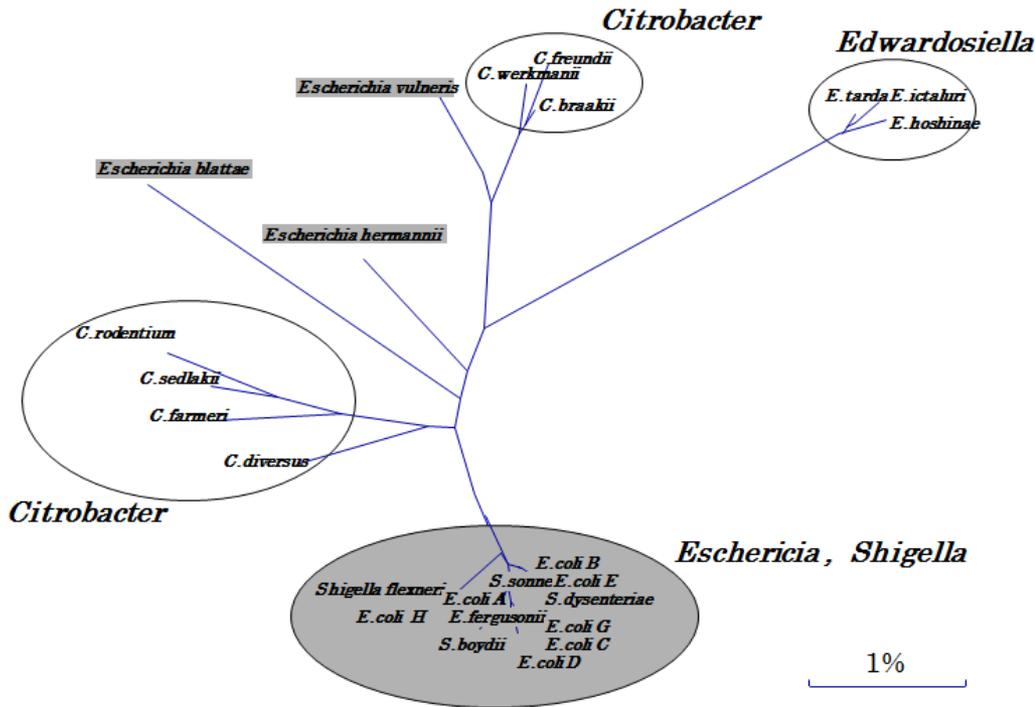


図4. 16S rDNA配列に基づいた*Escherichia*属、*Shigella*属、*Citrobacter*属、*Edwardosiella*属の系統樹

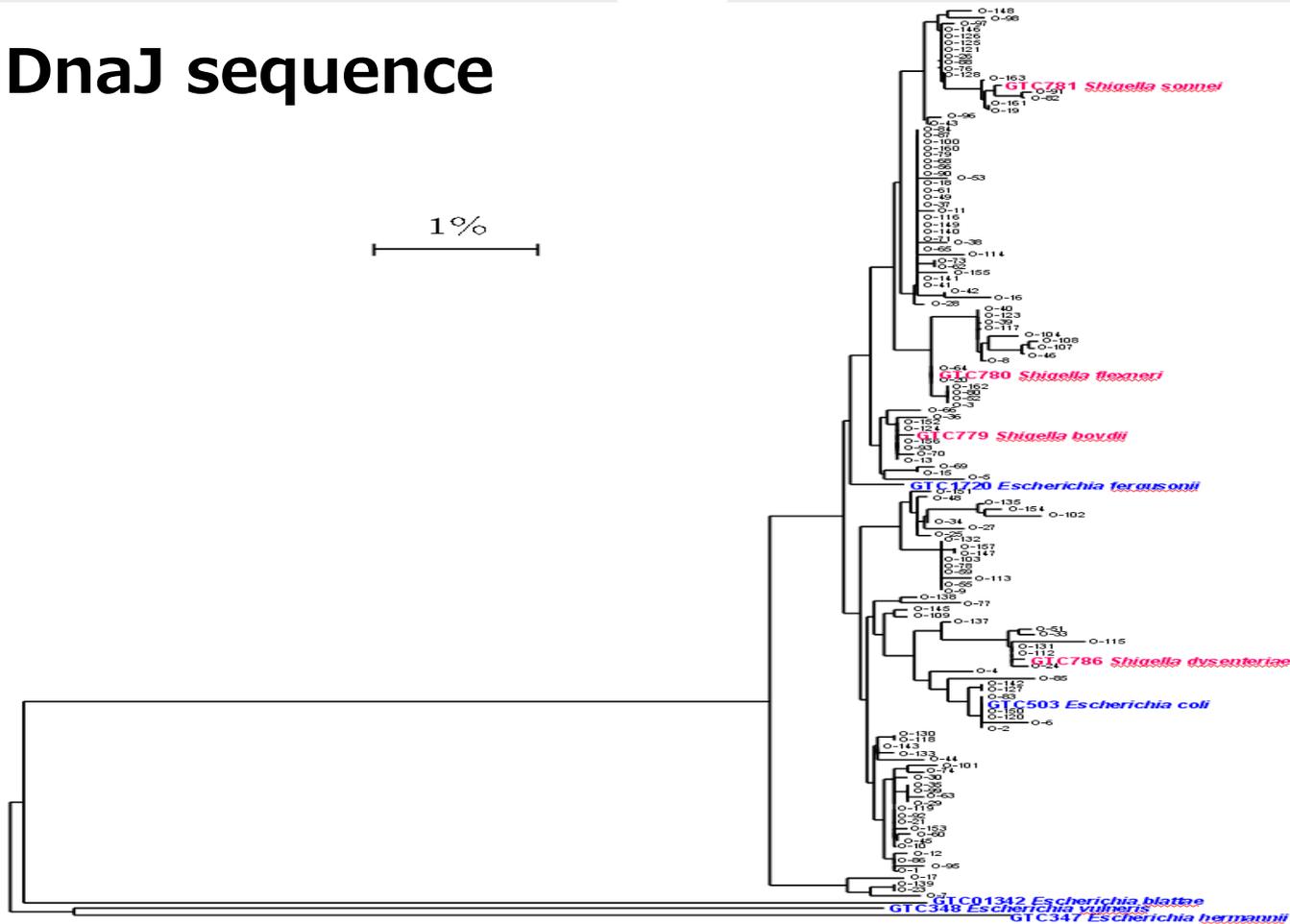
E. coli
E. fergusonii
E. arbertii
Shigella spp.

E. vulneri
E. hermannii
E. blattae

16S rRNA phylogeny

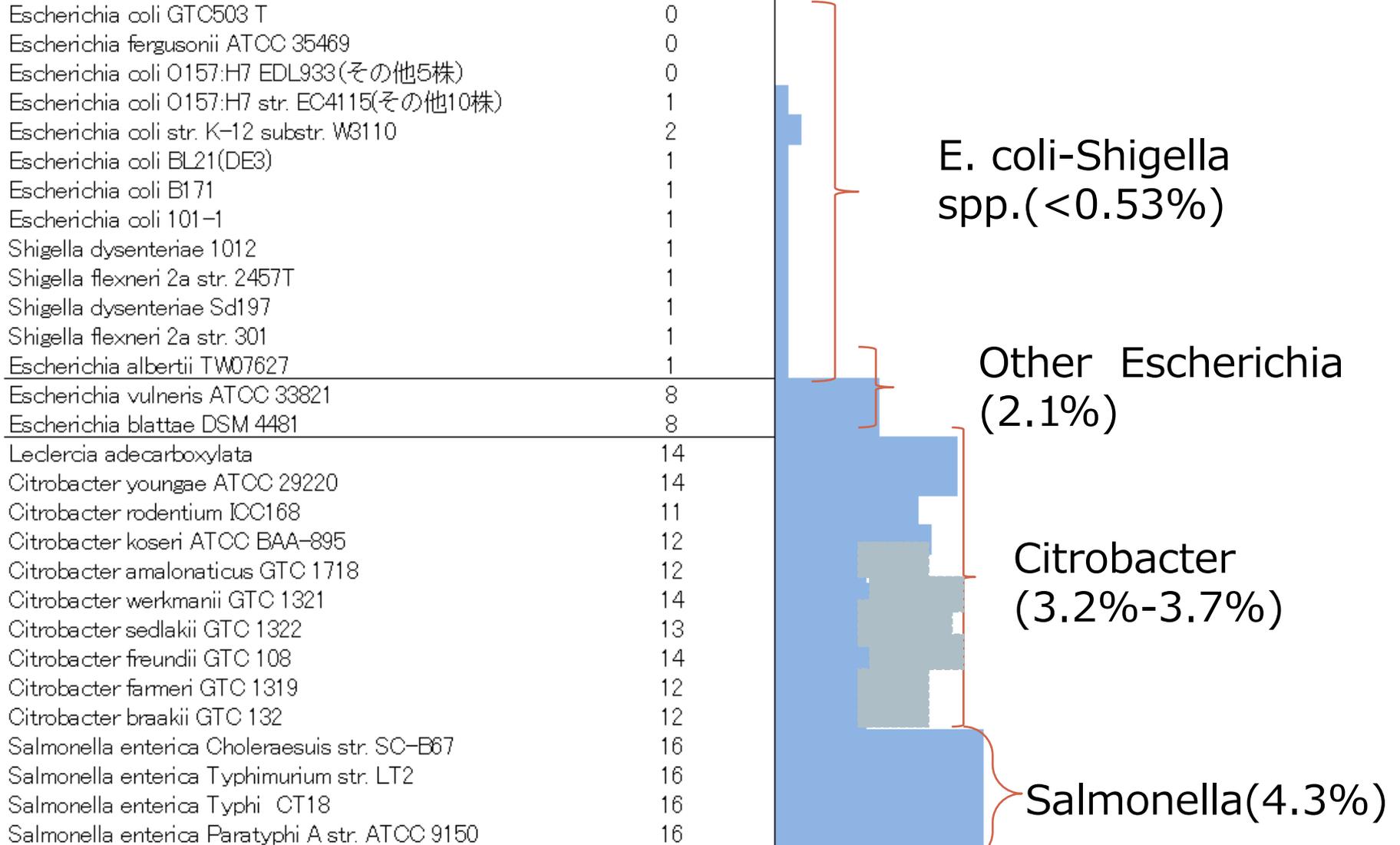
E.coli serovars (O1 to 180) and *Shigella* sp.

DnaJ sequence

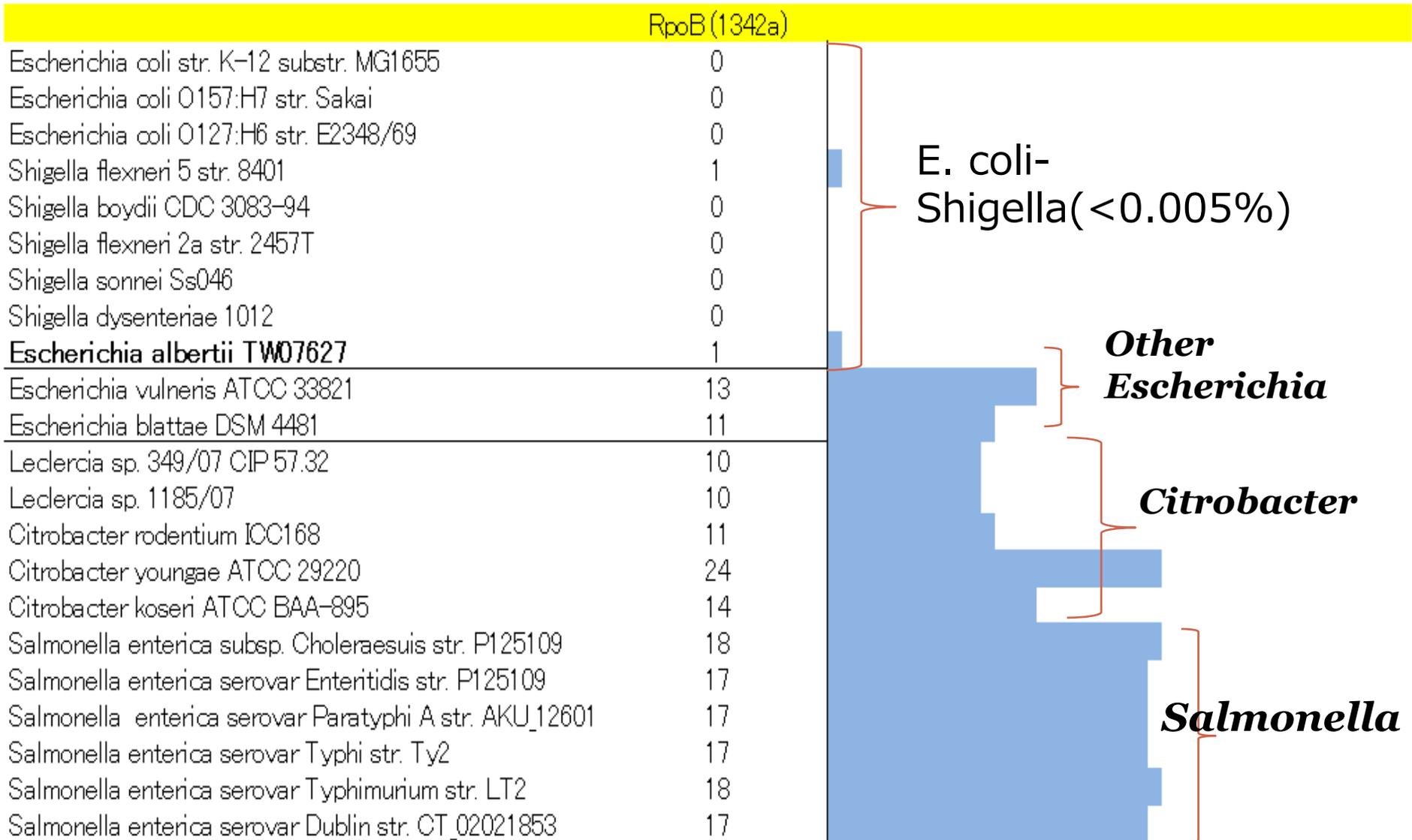


DnaJ Protein SAP(*E.coli*)

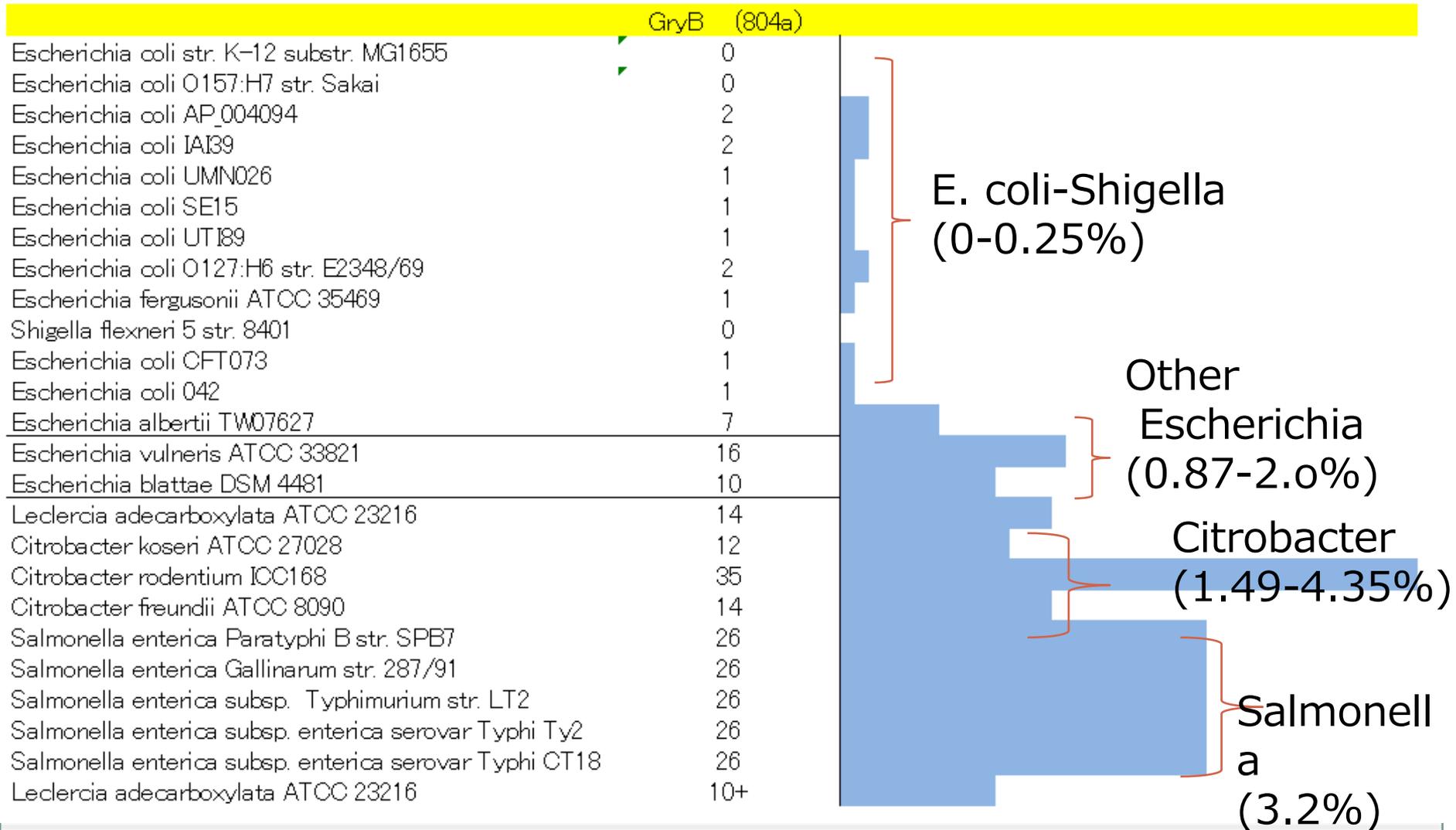
DnaJ (376a)



RpoB Protein SAP



GryB Protein SAP



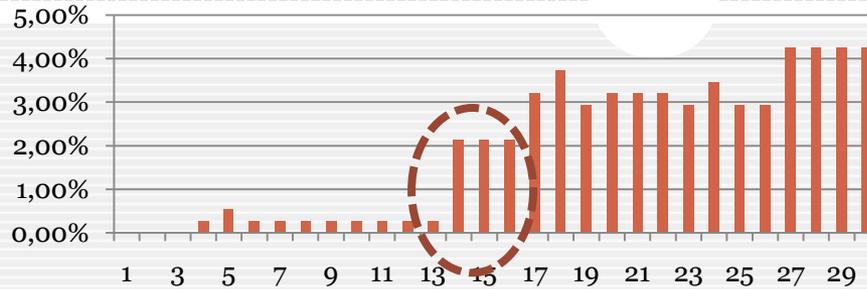
House keeping Protein SAP

Escherichia

Citorbacter

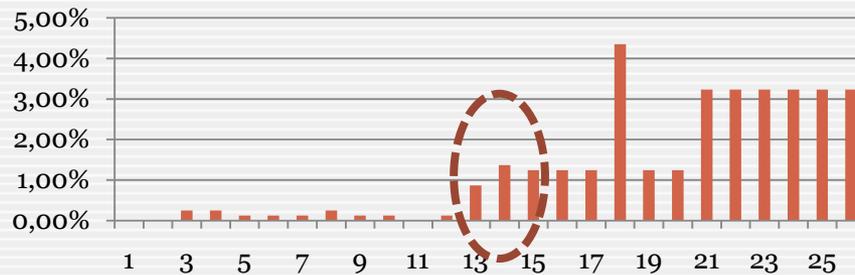
Salmonella

376 AA



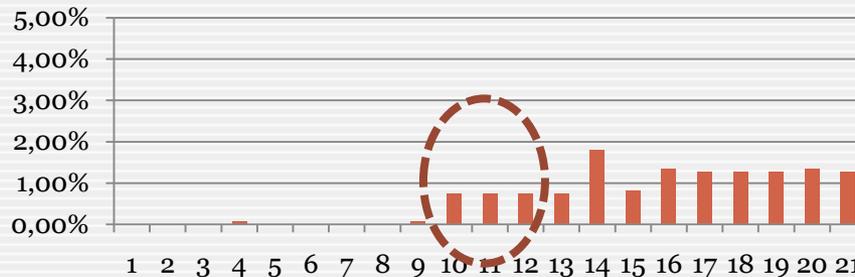
DnaJ < 0.53%

1342 AA



RopB < 0.005%

804 AA



GryB < 0.25%

E. vulneri, *E. hurmanii*, *E. blattae* should be reclassified as different genera

Genus level DnaJ Amino acid diversities

| Genus | <i>dnaJ</i> (%) | | | 16S rRNA (%) | | |
|----------------------------------------|-----------------|------|------|--------------|-----|------|
| | min | ave | max | min | ave | max |
| <i>Salmonella</i> | 2.7 | 4.8 | 10.1 | 0.1 | 1.0 | 2.3 |
| <i>Escherichia</i> | 2.2 | 12.3 | 18.4 | 0.7 | 2.1 | 4.2 |
| <i>Shigella</i> | 0.7 | 1.2 | 2.4 | 0.1 | 0.5 | 1.2 |
| <i>E.coli</i> and <i>Shigella spp.</i> | 0.7 | 1.4 | 2.6 | 0.3 | 0.5 | 1.2 |
| <i>Citrobacter</i> | 2.1 | 9.2 | 19.8 | 0 | 1.8 | 4.1 |
| <i>Enterobacter</i> | 3.8 | 11.6 | 17.4 | 0 | 1.9 | 3.8 |
| <i>Klebsiella</i> | 0.5 | 6.1 | 13.5 | 0.3 | 1.2 | 2.5 |
| <i>Yersinia</i> | 9.5 | 11.6 | 16.6 | 0.1 | 1.1 | 2.2 |
| <i>Vibrio</i> | 49.6 | 52.2 | 57.7 | 7.9 | 9.1 | 10.1 |

Reclassification of *Shigella* and *E.coli*

Escherichia coli

Enterohaemorrhagic

Enteroinvasive

Enterotoxigenic

Enteroadherent

Enteropathogenic

E. fergusonii -> *nonpathogenic E.coli*

E. arbutii -> **independent species in the genus *Escherichia***

E. vulneris

E. blattae

E. hurmanii

***Citrobacter* or new genera**

Shigella dysenteriae

Serovar O1

Serovar all others

Shigella flexneri

Shigella sonnei

Shigella boydii

* *E. blattae* is already proposed to transfer to a new genus

Bacillus cereus group

- *B.anthraxis*
- *B.cereus*
- *B.thuringiensis*

These shares more than 99.5% 16S rRNA similarities with *B.anthraxis*

- *B.mycoides*
- *B.pseudomycooides*
- *B.weihenstephanensis*

99.3%-95% similarities to *B.anthraxis*

DnaJ sequence of Out break strains of *B.cereus* was more than 99% related to *B.anthraxis*

(DnaJ variation within *B.cereus* (9.6%) is bigger than

Table 3 16SrRNA and DnaJ sequence similarity matrix among type strains and *B.cereus* groups

| Species | Sequence Similarity% (16S rRNA(left lower), DnaJ(right upper)) | | | | | | | | | |
|-------------------------------------------------------|----------------------------------------------------------------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| <i>B. anthracis</i> str.Ames Ancestor ^T | 1 | * | 99.5 | 95.8 | 94.5 | 95.6 | 94.4 | 91.6 | 87.8 | 91.7 |
| <i>B. cereus</i> GTC 02916 outbreak group | 2 | 99.9 | * | 95.9 | 94.6 | 95.7 | 94.5 | 91.8 | 87.6 | 91.8 |
| <i>B. cereus</i> GTC 02927 group1 | 3 | 99.9 | 99.9 | * | 93.9 | 95.4 | 94.1 | 92.7 | 86.6 | 92.9 |
| <i>B. cereus</i> GTC 02826 ^T group2 | 4 | 99.9 | 99.9 | 99.9 | * | 94.3 | 99.1 | 91.4 | 86.9 | 91.7 |
| <i>B. cereus</i> GTC 02892 group 3 | 5 | 99.9 | 99.9 | 99.9 | 99.9 | * | 94.5 | 92.4 | 86.8 | 92.6 |
| <i>B. thuringensis</i> GTC 02829 ^T | 6 | 99.6 | 99.7 | 99.7 | 99.7 | 99.7 | * | 91.8 | 87.1 | 91.9 |
| <i>B. mycoides</i> GTC 01994 ^T | 7 | 99.3 | 99.5 | 99.5 | 99.5 | 99.5 | 99.3 | * | 86.8 | 98.1 |
| <i>B. pseudomycoides</i> GTC 02831 ^T | 8 | 99.6 | 99.6 | 99.6 | 99.6 | 99.6 | 99.3 | 99.3 | * | 87.2 |
| <i>B. weihenstephanensis</i> GTC 02834 ^T | 9 | 99.3 | 99.5 | 99.3 | 99.3 | 99.3 | 99.3 | 100 | 99 | * |

dnaJ gene similarity values were shown with boldface on the upper right; 16SrDNA gene similarity value, located at the lower left.

Pathogenic factors to differentiate B.cereus group

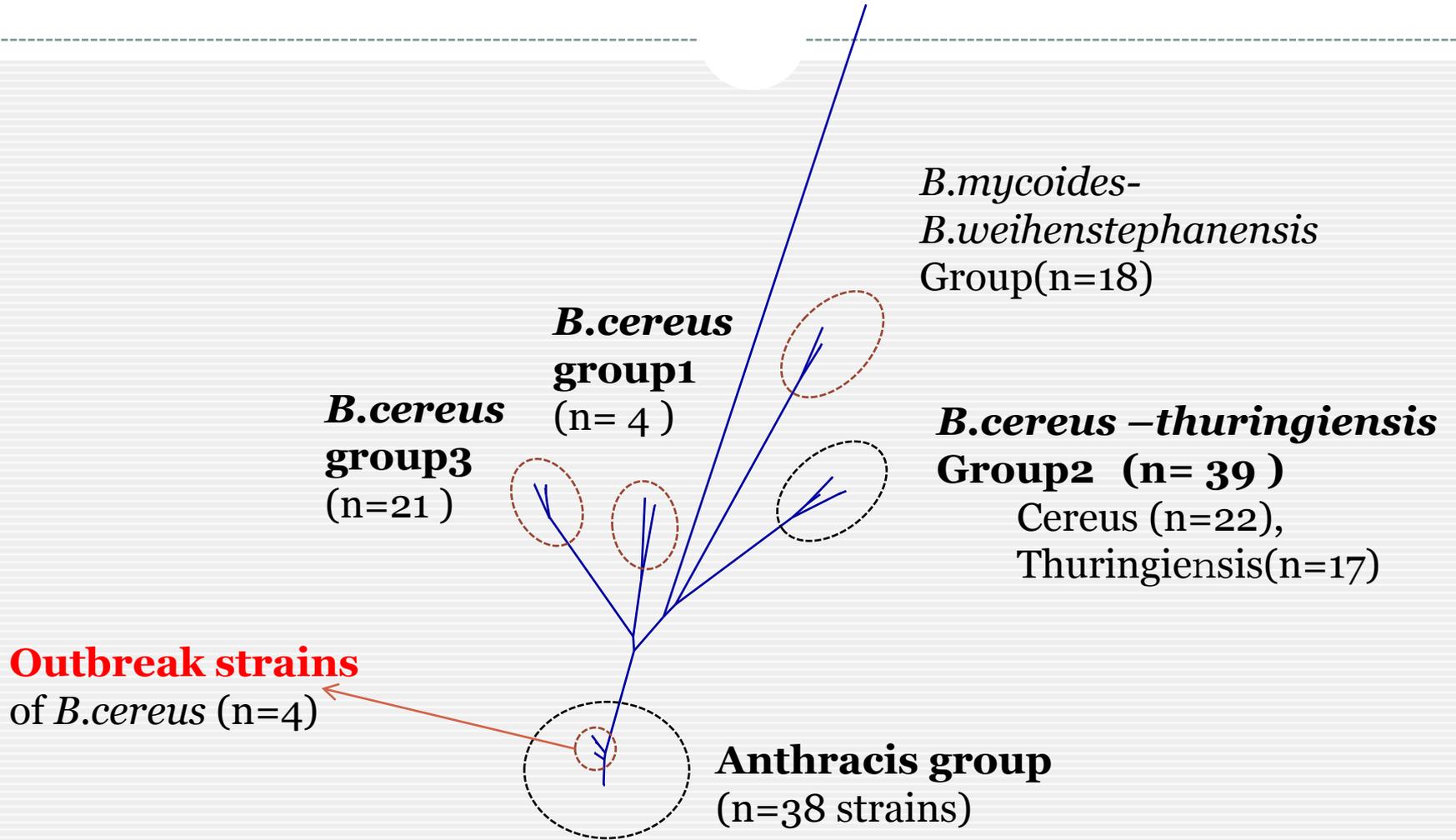
| Species | Strain | Source | gene amplification for: | | | | | | | | | | | | | |
|------------------------------|------------------------|--------------------------------------|-------------------------|-------------|-------------|-------------|---------|------------|------------|------------|------------|------------|------------|------------|-------|-------|
| | | | <i>nhe</i> | <i>hblC</i> | <i>hblD</i> | <i>hblA</i> | pCER270 | <i>ceB</i> | <i>cry</i> | <i>pag</i> | <i>lef</i> | <i>cya</i> | <i>sap</i> | <i>cap</i> | Ba813 | |
| <i>B.cereus</i> | N=47 strains | Sporadic Isolates from Blood culture | 47/47* | 10/47 | 20/47 | 20/47 | 0/47 | 0/47 | 0/47 | 0/47 | 0/47 | 0/47 | 1/47 | 0/47 | 0/47 | |
| <i>B.cereus</i> | N=18 strains | Food poisoning Isolates | 18/18 | 0/18 | 0/18 | 0/18 | 18/18 | 18/18 | 0/18 | 0/18 | 0/18 | 0/18 | 0/18 | 0/18 | 0/18 | |
| <i>B.cereus</i> | N=4 strains | Out break strains from Blood culture | 4/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 0/4 | 1/4 | 4/4 | |
| | GTC 02891 | Blood culture | + | - | - | - | - | - | - | - | - | - | - | + | + | |
| | GTC 02896 | Blood culture | + | - | - | - | - | - | - | - | - | - | - | - | + | |
| | GTC 02916 | Blood culture | + | - | - | - | - | - | - | - | - | - | - | - | + | |
| | GTC 02917 | Blood culture | + | - | - | - | - | - | - | - | - | - | - | - | + | |
| <i>B.anthraxis</i> | N=20 strains | Wild animal Isolates | 19/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 17/20 | 17/20 | 17/20 | 20/20 | 17/20 | 20/20 |
| <i>B.thuringiensis</i> | GTC 02829 ^T | IAM 12077 | + | - | + | + | - | - | + | - | - | - | - | - | - | |
| <i>B.mycoides</i> | GTC 01994 ^T | DSM 2048 | + | - | + | + | - | - | - | - | - | - | - | - | - | |
| <i>B.pseudomycoides</i> | GTC 02831 ^T | NRRL B-617 | - | - | - | - | - | - | - | - | - | - | - | - | - | |
| <i>B.walthenstephanensis</i> | GTC 02834 ^T | WSBC10204 | + | + | - | + | - | - | - | - | - | - | - | - | - | |

(+): positive; (-):negative.

* number of positive strains

DnaJ sequence SNP

B.pseudomycooides
GTC 02831T



Outbreak strains make typical *B.cereus* colonies



Suspect
B.cereus , *B.thuringiensis*



Suspect
B.subtilis , *B.anthraxis*



Suspect
B.mycooides



When does a strain spin out from a species ?

House keeping Protein variation (SAP)

Use complete amino acid sequence to discuss Species border or genus border. Until amino acid of a protein changes, the function is same among strains with different codon usage.

House keeping Gene Sequence variation (SNP)

Analyze strain evolution within a species.

Different codon usage among strains in a species is useful to analyze whether a strain is ready to spin out from a species.