

Lab

Data Information Protocol

20250301 – v1

基本資料格式定義

Long Data

- 在 column 中 ID 不是唯一
- 較常用來分組畫圖使用格式
- 進階統計視覺化使用格式

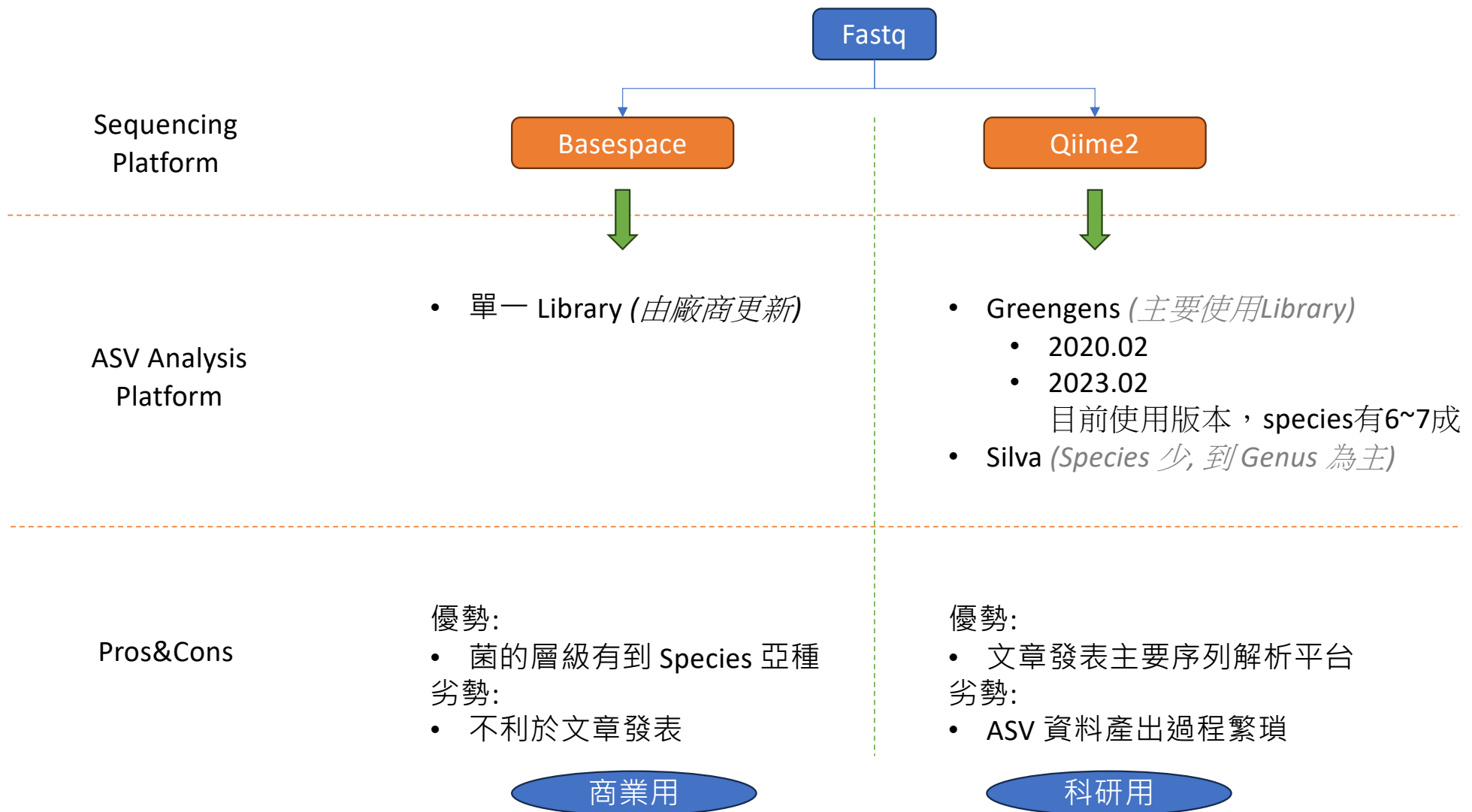
Patient ID	Age	Measurement	Value
P001	45	Glucose	120
P001	45	Cholesterol	200
P001	45	Blood Pressure	130
P002	60	Glucose	150
P002	60	Cholesterol	240
P002	60	Blood Pressure	140
P003	30	Glucose	90
P003	30	Cholesterol	180
P003	30	Blood Pressure	115

Tabular Data Format / Wide Data Format

- 在 column 中 ID 是唯一
- 較常用來分析、報表使用格式
- 機器學習常使用格式

Patient ID	Age	Glucose (mg/dL)	Cholesterol (mg/dL)	Blood Pressure (mmHg)
P001	45	120	200	130
P002	60	150	240	140
P003	30	90	180	115

文庫比較



標準資料型態格式

Summary Table

- CSV由產出的產物Basespace

CH4504-Trim_S76.summary

Kingdom	Phylum	Class	Order	Family	Genus	Species	num_hits	%_hits
Unclassified							143	0.085
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	vulgatus	19814	11.711
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides		18294	10.812
Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae			18183	10.747
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri	15817	9.348
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Sutterella		10995	6.498
Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megasphaera	elsdenii	7999	4.728
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	stercoris	7547	4.461
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	coprocola	6141	3.630
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Ruminococcus		5884	3.478
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	finnegoldii	4969	2.937
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia		4128	2.440
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides	merdae	3887	2.297
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Alkaliphilus	crotonatoxidans	3357	1.984
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	producta	3110	1.838
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	massiliensis	2530	1.495
Bacteria	Firmicutes	Clostridia	Clostridiales				2478	1.465
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	coccoides	2407	1.423
Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Fusobacterium	gonidiaformans	2339	1.382
Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megasphaera	hominis	2274	1.344
Bacteria	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	Megamonas		2269	1.341
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Odoribacteraceae	Butyricimonas		1367	0.808
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium		1291	0.763
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Peptoniphilus	methioninivorax	1197	0.707
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Natronincola		1183	0.699
Bacteria	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Cetobacterium	ceti	1176	0.695
Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Sutterella	sanguinus	1061	0.627
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	xylanisolvans	873	0.516
Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae			859	0.508
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Ruminococcus	gnavus	841	0.497

*Sample Table

- SummaryTable的型態延伸
- 每個檔案為一個sample
- 按照Taxonomy階級分column
- 相對豐度為**小數型態**
 - 為對應文獻使用規則
- 空白處保持不填入資料



Taxonomy 後續轉換:

- 間格為'_'
- s_Prevotella_copri
- Genus後面帶spp.
 - g_Bacteroides spp.

3179C

Kingdom	Phylum	Class	Order	Family	Genus	Species	num_hits	%_hits
Bacteria							355	0.012726751272675100
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium		0	0.0
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	llatzerense	0	0.0
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionibacterium	acnes	0	0.0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae			0	0.0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Adlercreutzia		6	0.00021510002151000200
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales				5	0.00017925001792500200
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides		125	0.004481250448125050
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	acidifaciens	44	0.0015774001577400200
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	plebeius	0	0.0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides		0	0.0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides	distasonis	6	0.00021510002151000200
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella		98	0.003513300351330040
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri	0	0.0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae			662	0.02373270237327020
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	AF12		6	0.00021510002151000200
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Rikenella		17	0.0006094500609450060
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	S24-7			1031	0.03696135369613540
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	Odoribacter		280	0.010038001003800100
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]	[Prevotella]		13	0.00046605004660500500
Bacteria	Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	Sediminibacterium		0	0.0
Bacteria	Cyanobacteria	4C0d-2	MLE1-12				0	0.0
Bacteria	Cyanobacteria	Chloroplast	Streptophyta				10	0.00035850003585000400
Bacteria	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae	Mucispirillum	schaedleri	1616	0.05793360579336060
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Jeotgalicoccus	psychrophilus	0	0.0
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus		0	0.0

豐度要是小數型態

已清理過的sample ID命名



Population Table-reads

- 多個sampleID
- 按照Taxonomy階級分column
- 數值為reads
 - 用於計算alpha diversity
 - Chao1 / Faith's PD用reads計算
- 空白處保持不填入資料
- SampleID 需已清理過



Taxonomy 後續轉換:

- 間格為'_'
- s_Prevotella_copri
- Genus後面帶spp.
 - g_Bacteroides spp.

Kingdom	Phylum	Class	Order	Family	Genus	Species	3179C	3179D	3179F	3179P	3182C	3182D	3182F	3182P
Bacteria							355	1072	300	426	121	116	41	116
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium		0	6	0	6	0	0	0	0
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	llatzerense	0	0	0	0	0	0	0	2
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionibacterium	acnes	0	11	0	4	0	0	0	0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae			0	0	0	0	0	0	0	0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Adlercreutzia		6	0	28	34	33	17	4	16
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales				5	5	0	26	26	54	27	42
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides		125	103	57	417	322	978	472	831
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	acidifaciens	44	16	42	120	102	197	77	178
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	plebeius	0	5	0	0	0	0	0	0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides		0	0	0	19	0	31	0	19
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides	distasonis	6	0	0	12	10	36	15	30
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella		98	83	120	219	146	409	176	353
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri	0	5	0	14	0	0	0	0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae			662	263	1306	1146	904	1732	745	1434
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	AF12		6	0	17	4	4	9	0	3
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Rikenella		17	8	10	17	17	34	16	47
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	S24-7			1031	1603	589	2731	2600	5559	1460	4068
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	Odoribacter		280	32	821	407	82	200	90	203
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]	[Prevotella]		13	9	19	24	18	41	12	30
Bacteria	Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	Sediminibacterium		0	0	0	0	0	4	0	0
Bacteria	Cyanobacteria	4C0d-2	MLE1-12				0	0	0	0	3	0	0	0
Bacteria	Cyanobacteria	Chloroplast	Streptophyta				10	15	16	11	6	6	4	8
Bacteria	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae	Mucispirillum	schaedleri	1616	458	3149	2368	1461	3582	1287	2715
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Jeotgalicoccus	psychrophilus	0	6	0	0	0	0	0	0
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus		0	18	0	6	7	14	0	0
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	sciuri	0	0	0	0	0	8	0	8
Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Aerococcus		0	12	0	0	0	10	0	8

已清理過的sample ID命名

Population Table-abundance

- 多個sampleID
- 按照Taxonomy階級分column
- 數值為豐度，並為**小數型態**
 - 用於一般計算
- 空白處保持不填入資料
- SampleID 需已清理過



Taxonomy 後續轉換:

- 間格為'_'
- s_Prevotella_copri
- Genus後面帶spp.
 - g_Bacteroides spp.

Kingdom	Phylum	Class	Order	Family	Genus	Species	3179C	3179D	3179F	3179P	3182C	3182D	3182F	3182P
Bacteria							355	1072	300	426	121	116	41	116
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium		0	6	0	6	0	0	0	0
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	llatzerense	0	0	0	0	0	0	0	2
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionibacterium	acnes	0	11	0	4	0	0	0	0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae			0	0	0	0	0	0	0	0
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Adlercreutzia		6	0	28	34	33	17	4	16
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales				5	5	0	26	26	54	27	42
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides		125	103	57	417	322	978	472	831
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	acidifaciens	44	16	42	120	102	197	77	178
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	plebeius	0	5	0	0	0	0	0	0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides		0	0	0	19	0	31	0	19
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides	distasonis	6	0	0	12	10	36	15	30
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella		98	83	120	219	146	409	176	353
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	copri	0	5	0	14	0	0	0	0
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae			662	263	1306	1146	904	1732	745	1434
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	AF12		6	0	17	4	4	9	0	3
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Rikenella		17	8	10	17	17	34	16	47
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	S24-7			1031	1603	589	2731	2600	5559	1460	4068
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	Odoribacter		280	32	821	407	82	200	90	203
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	[Paraprevotellaceae]	[Prevotella]		13	9	19	24	18	41	12	30
Bacteria	Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	Sediminibacterium		0	0	0	0	0	4	0	0
Bacteria	Cyanobacteria	4C0d-2	MLE1-12				0	0	0	0	3	0	0	0
Bacteria	Cyanobacteria	Chloroplast	Streptophyta				10	15	16	11	6	6	4	8
Bacteria	Deferribacteres	Deferribacteres	Deferribacterales	Deferribacteraceae	Mucispirillum	schaedleri	1616	458	3149	2368	1461	3582	1287	2715
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Jeotgalicoccus	psychrophilus	0	6	0	0	0	0	0	0
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus		0	18	0	6	7	14	0	0
Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	sciuri	0	0	0	0	0	8	0	8
Bacteria	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Aerococcus		0	12	0	0	0	10	0	8

豐度要是小數型態

每個column為sample ID

Population Table-metabolite

- SampleID 需已清理過
- 使用Qiime2後加上KEGG產生

pathway	description	3691C	3691D	3691F	3691PA	3692C	3692D
ko00010	Glycolysis / Gluconeogenesis	2227.32025	5367.253000000000	4085.349	3614.30625	436.13575	5985.078500000000
ko00020	Citrate cycle (TCA cycle)	1840.6785185185200	4525.23037037037	2277.264444444444	2339.387777777778	198.89222222222200	2059.6081481481500
ko00030	Pentose phosphate pathway	2368.004545454550	5672.241818181820	4060.104545454550	3585.339090909090	556.2554545454550	5890.955909090910
ko00040	Pentose and glucuronate interconversions	971.3761538461540	2837.42	1348.965	1415.1311538461500	152.48384615384600	1574.4719230769200
ko00051	Fructose and mannose metabolism	1503.7978787878800	4065.166363636360	3353.500303030300	2683.3372727272700	413.0812121212120	5204.084545454550
ko00052	Galactose metabolism	1313.862413793100	3843.5617241379300	3172.215862068970	2413.4896551724100	395.786551724138	5077.553793103450
ko00053	Ascorbate and aldarate metabolism	699.0388888888890	1820.290555555560	773.7716666666670	974.6327777777780	49.39277777777780	496.3594444444440
ko00061	Fatty acid biosynthesis	3017.1338461538500	6971.484615384620	3674.1153846153800	4275.593076923080	355.25230769230800	4741.966153846150
ko00071	Fatty acid metabolism	823.7008333333330	2337.057083333330	1247.652916666670	1281.2020833333300	130.34916666666700	1379.88375
ko00072	Synthesis and degradation of ketone bodies	1196.5525	3339.720000000000	2653.7325	2109.867500000000	313.9675000000000	2860.657500000000
ko00100	Steroid biosynthesis	4.346875	6.440625	5.0625	5.19375	0.0	1.25
ko00120	Primary bile acid biosynthesis	58.208888888888900	36.45	532.6477777777780	181.7888888888890	101.5	1127.2955555555600
ko00121	Secondary bile acid biosynthesis	523.88	328.05	4793.83	1636.1	913.5	10145.66
ko00130	Ubiquinone and other terpenoid-quinone biosynthesis	1028.1125	2544.3315	1157.59	1182.507500000000	118.5260000000000	981.324
ko00140	Steroid hormone biosynthesis	100.61111111111100	279.1111111111110	100.5	114.61111111111100	0.208333333333330	11.416666666666700
ko00190	Oxidative phosphorylation	679.618834951456	1720.580873786410	971.7148543689320	911.0479611650490	104.06864077669900	1032.5994174757300
ko00195	Photosynthesis	0.0	0.0	0.0	0.0	0.0	1883.3003125
ko00196	Photosynthesis - antenna proteins	0.0	0.0	0.0	0.0	0.0	5.714285714285710
ko00230	Purine metabolism	1275.8018487395000	2774.512184873950	2158.028319327730	1968.3436974789900	257.0491596638660	3307.90411764706
ko00240	Pyrimidine metabolism	1566.9928	3350.576266666670	2929.946133333330	2432.916666666670	365.9750666666670	4492.404933333330
ko00250	Alanine, aspartate and glutamate metabolism	1912.1903703703700	4707.207407407410	3320.1118518518500	2871.7922222222200	408.7644444444450	4876.71
ko00260	Glycine, serine and threonine metabolism	1830.5473529411800	4742.582352941180	2215.6729411764700	2455.1176470588200	238.80882352941200	2284.57
ko00270	Cysteine and methionine metabolism	1501.2057575757600	3536.0545454545500	2304.738484848490	2217.451515151520	323.3833333333330	3562.83787878788
ko00280	Valine, leucine and isoleucine degradation	1061.613448275860	2885.942413793100	1341.5920689655200	1452.664827586210	89.58517241379310	827.6965517241380
ko00281	Geraniol degradation	176.71875	564.89125	184.23375	231.43375	31.121250000000000	86.74375

資料保存規格

資料儲存結構規則

* 為必要檔案

```
.
├── 科研案A
│   ├── human
│   │   ├── Stool
│   │   ├── YF
│   │   ├── Stomach
│   │   └── YF
│   └── metadata
├── 科研案B
│   ├── metadata
│   ├── human
│   │   ├── Stool
│   │   │   ├── Fastq
│   │   │   ├── Basespace
│   │   │   │   ├── *BasespaceFiles
│   │   │   │   └── *SampleTable
│   │   │   │       ├── SampleA.csv
│   │   │   │       └── SampleB.csv ...
│   │   │   └── PopulationTable
│   │   │       ├── PopulationTable_abundance.csv
│   │   │       └── PopulationTable_reads.csv
│   │   └── Qiime2
│   │       ├── *Qiime_ver
│   │       │   ├── *SampleTable
│   │       │   └── TabularTable
│   │       │       ├── microbes_tabular_level_type_Qiimer_ver.csv
│   │       │       ├── pathway_tabular_Qiimer_ver.csv
│   │       │       └── enzyme_tabular_Qiimer_ver.csv
│   │       └── PopulationTable
│   │           ├── PopulationTable_abundance.csv
│   │           ├── PopulationTable_reads.csv
│   │           ├── PopulationTable_pathway.csv
│   │           └── PopulationTable_enzyme.csv
```