

PUMA Program for Unifying Microbiome Analysis

Christopher Dao

MIMG 109L



Topics

1. UCLA MIMG 109L class

- **Research immersion in microbial ecology**

2. Overview of microbiome analysis

- Data processing
- Data analysis

3. PUMA project

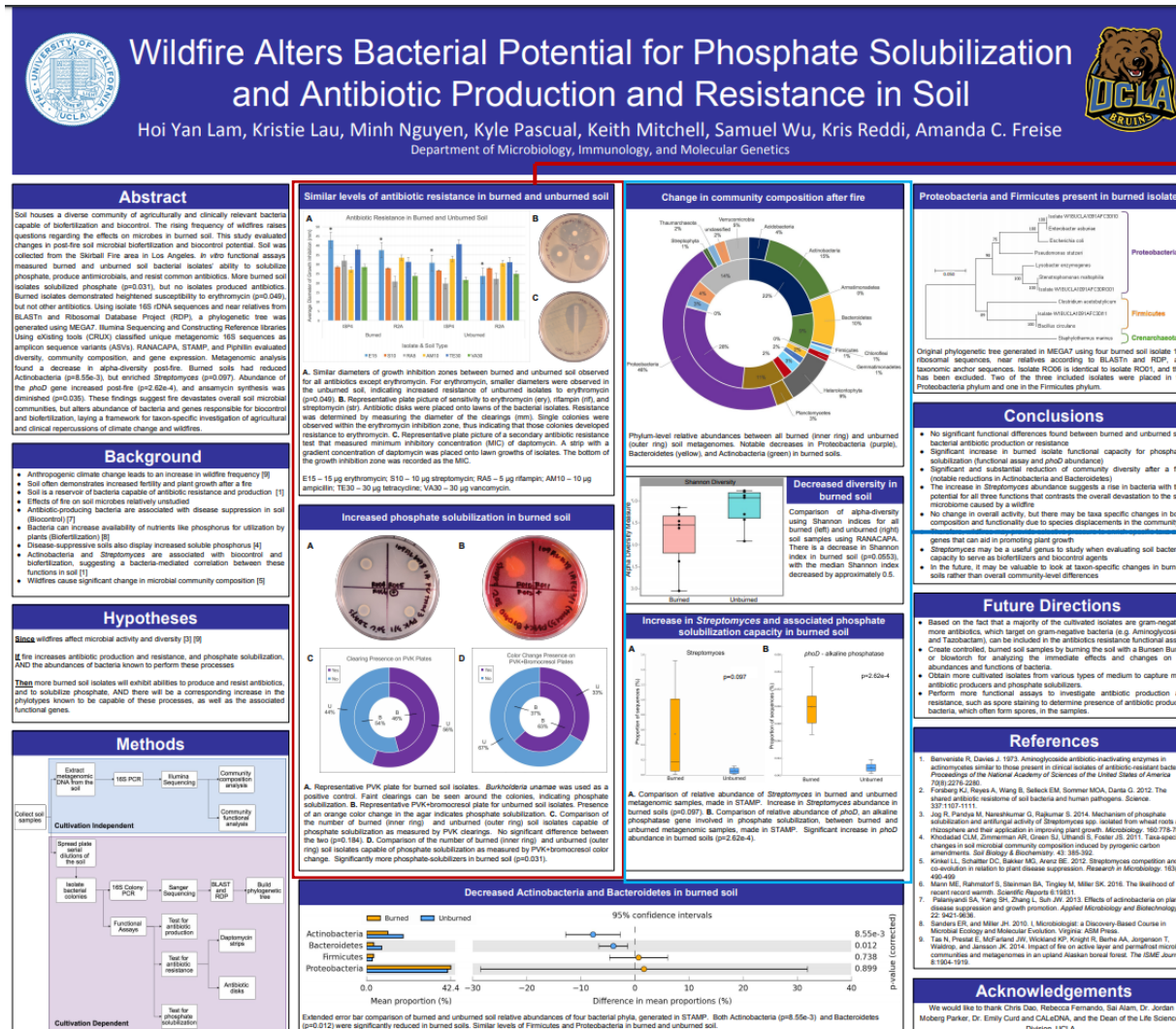
MIMG 109AL/BL – What's it about?

Microbial ecology: how microorganisms interact with each other and their environment

- **Biodiversity** – isolate, identify, and quantify microorganisms in various habitats
“WHO'S THERE?”
- **Microbial activity** – measure what microorganisms are doing in their habitats
“WHAT ARE THEY DOING THERE?”



MIMG 109L Posters



Cultivation dependent analysis
on cultured bacterial isolates

Cultivation independent analysis
of microbial community profiles

Topics

1. UCLA MIMG 109L class
 - Research immersion in microbial ecology

2. **Overview of microbiome analysis**
 - **Data processing**
 - **Data analysis**

3. PUMA project

What is the microbiome?

- **Microbiome** – community of microorganisms (mainly bacteria) in any environment
- The microbiome has been studied in:
 - human samples
 - animal samples
 - soil samples
 - indoor surfaces etc., etc.



How can DNA sequencing techniques be used to study the microbiome?

Amplicon sequencing of the 16S gene

Use an molecular evolutionary
clock!



16S rRNA gene

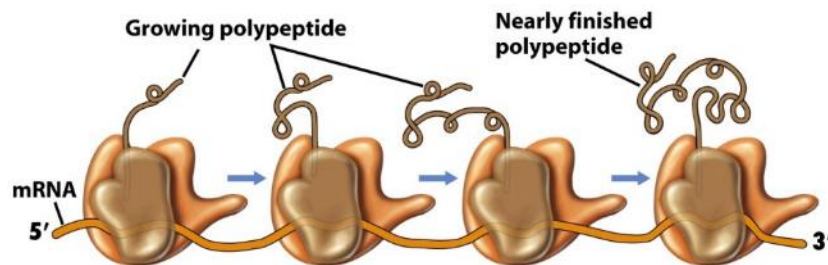
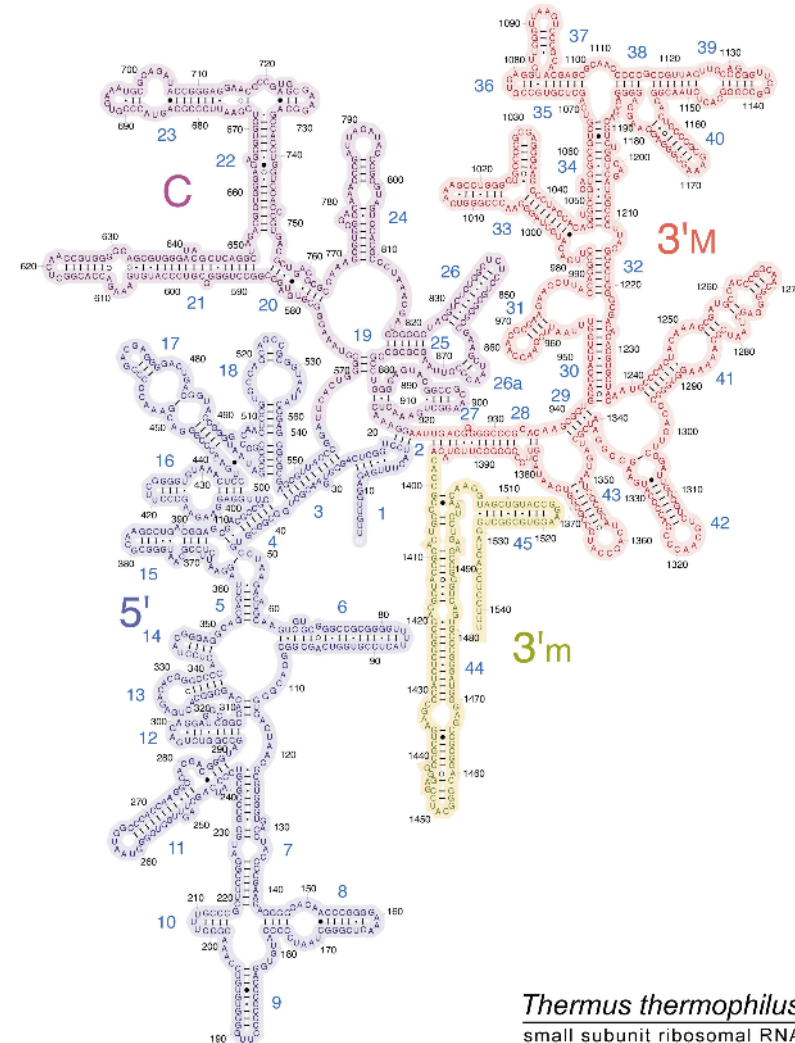
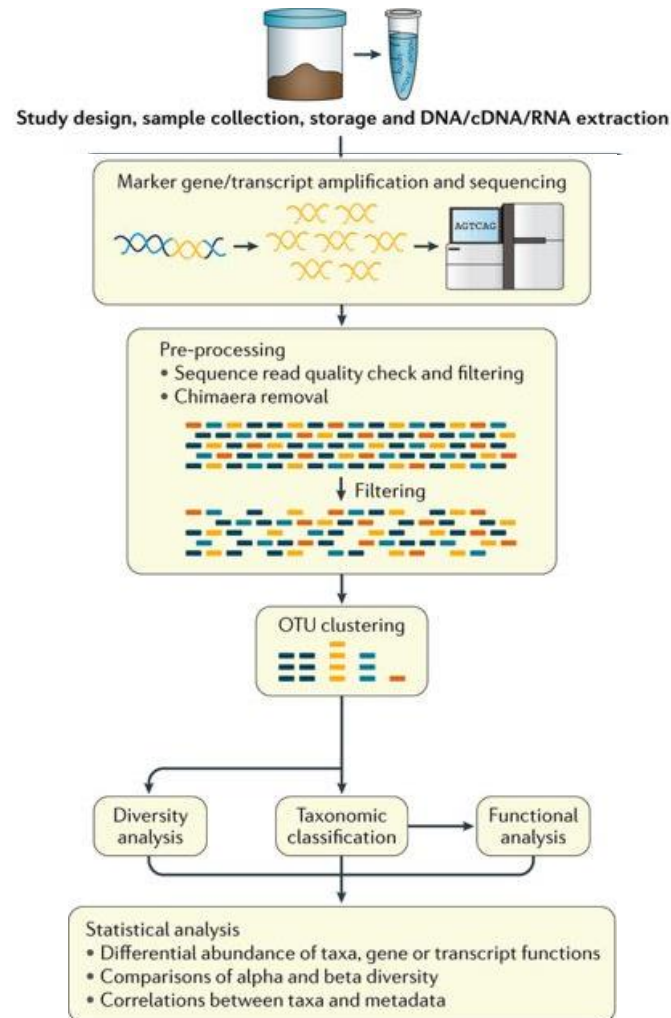


Figure 7-39 Brock Biology of Microorganisms 11/e
© 2006 Pearson Prentice Hall, Inc.



Thermus thermophilus
small subunit ribosomal RNA

Microbiome Analysis Flowchart



Topics

1. UCLA MIMG 109L class
 - Research immersion in microbial ecology






2. Overview of microbiome analysis
 - Data processing
 - Data analysis

3. **PUMA project**

PUMA: Program for Unifying Microbiome Analysis



PUMA: A tool for processing 16S rRNA taxonomy data for analysis and visualization

 Keith Mitchell,  Chris Dao,  Amanda Freise,  Serghei Mangul,  Jordan Moberg Parker
doi: <https://doi.org/10.1101/482380>

Abstract Info/History Metrics  Preview PDF

Abstract

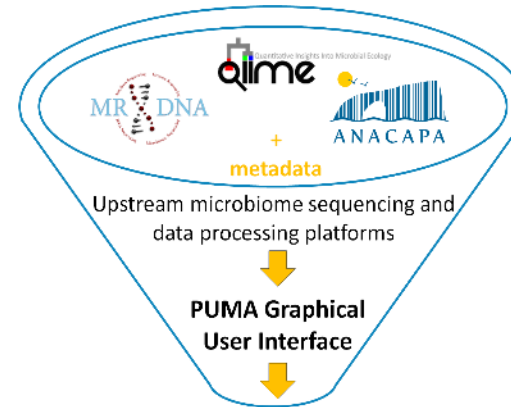
Microbial community profiling and functional inference via 16S rRNA analysis is quickly expanding across various areas of microbiology due to improvements to technology. There are numerous platforms for producing 16S rRNA taxonomic data which often vary in file and sequence formatting, creating a common barrier in microbiome studies. Additionally, many of the methods for analyzing and visualizing this sequencing data each require their own specific formatting. As a result, efficient and reproducible comparative analysis of taxonomic data and corresponding metadata in multiple programs remains a challenge in the investigation of microbial communities. PUMA, the Program for Unifying Microbiome Analysis, alleviates this problem in microbiome studies by allowing users to take advantage of numerous 16S rRNA

<https://www.biorxiv.org/content/early/2018/11/29/482380>

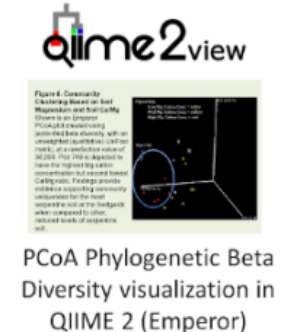
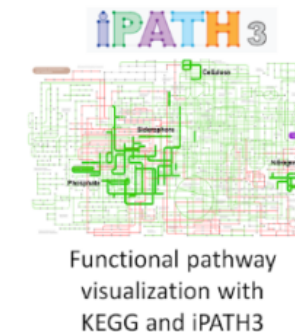
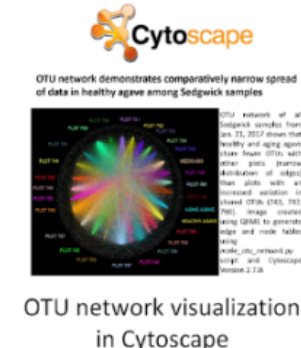
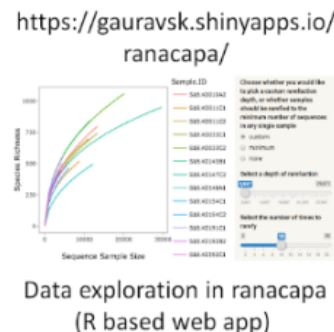
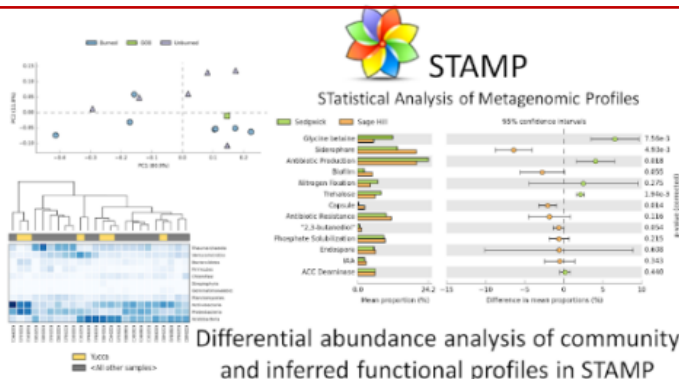
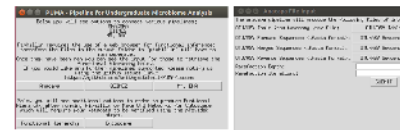
Mitchell 2018

PUMA: Program for Unifying Microbiome Analysis

Sequence processing/
taxonomic assignment platforms



PUMA



Statistical analysis and data visualization tools

PUMA: Program for Unifying Microbiome Analysis

PUMA is a Python GUI tool that formats data from different sequence processing and taxonomic assignment platforms to use in statistical analysis and data visualization tools

Goal: allow researchers to conduct microbiome analyses without needing to use command-line interface or R/Python scripting

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	#OTU ID	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	sum taxonomy	
2	forward_1	158	256	19	12	2	0	58	1066	455	79	7	2	5	74	Proteobacteria;Bet	
3	forward_1	39	80	2	3	0	0	17	307	130	17	3	0	1	23	Proteobacteria;Bet	
4	forward_1	71	4	280	12	0	0	100	0	55	4	30	4	56	5	Acidobacteria;Blast	
5	forward_1	36	54	2	1	1	0	11	279	69	15	2	0	1	17	Proteobacteria;Bet	
6	forward_1	84	1	202	1	3	4	0	0	4	10	33	28	14	0	Thaumarchaeota;Ni	
7	forward_1	24	3	7	3	171	21	19	0	4	1	3	50	5	2	Thaumarchaeota;Ni	
8	forward_1	22	36	6	2	0	1	10	158	53	9	2	0	0	5	Proteobacteria;Bet	
9	forward_1	15	42	1	2	0	0	7	163	53	6	0	1	0	11	Proteobacteria;Bet	
10	forward_1	14	0	219	9	7	0	3	0	12	4	2	3	7	4	Acidobacteria;Blast	
11	forward_1	31	1	1	5	0	0	19	104	4	25	30	6	3	10	Proteobacteria;Bet	
12	forward_1	30	1	96	2	1	0	39	0	20	0	15	3	17	9	Acidobacteria;Blast	
13	forward_1	20	18	11	23	0	0	45	29	19	6	12	0	6	14	Actinobacteria;Acti	
14	forward_1	9	9	1	1	1	4	0	58	13	1	2	1	3	92	Proteobacteria;Gan	
15	forward_1	13	93	1	3	0	0	6	41	23	8	1	0	1	2	Proteobacteria;Bet	
16	forward_1	21	7	22	20	0	0	59	1	3	5	13	0	31	3	Proteobacteria;Alpl	
17	forward_1	0	10	0	1	2	0	0	0	0	27	0	0	0	122	;Oomycetes;Pythial	
18	forward_1	5	2	0	7	0	0	147	6	2	2	0	0	7	2	Firmicutes;Bacilli;B	
19	forward_1	15	17	14	41	0	0	64	0	6	6	1	0	4	7	Proteobacteria;Alpl	

Anacapa data format

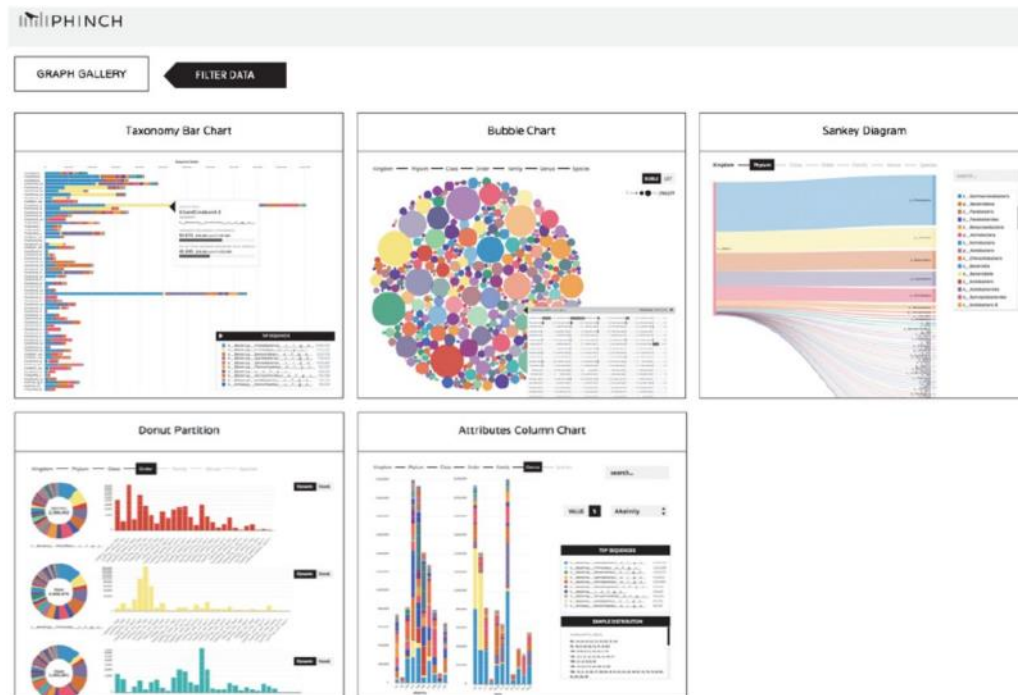
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	otu name	16rRNA	Percent H	evalue	bitscore	homology	total leng	BAF15	MWPF15	RKF15	BSGF15	WRF15	GNF15	MMF15
2	OTU_1	u69636.1	99.25926	1.78E-133	479.179			169	131	109	149	47	263	32
3	OTU_2	unculture	99.26199	6.22E-133	477.3757			2	149	28	156	260	97	20
4	OTU_3	unculture	98.89299	2.65E-131	471.9656			1	94	1	273	197	39	0
5	OTU_4	actinoph	98.89299	2.18E-132	475.5723			353	5	0	0	0	8	1
6	OTU_5	unculture	99.26199	6.22E-133	477.3757			44	11	109	0	0	11	326
7	OTU_6	chitinoph	98.52399	2.66E-131	471.9656			24	486	0	1	0	0	0
8	OTU_7	skermane	98.89299	2.65E-131	471.9656			2	113	3	292	260	71	1
9	OTU_8	unculture	99.26199	5.13E-134	480.9824			0	81	183	11	0	150	48
10	OTU_9	af529109.1	100	6.27E-133	477.3757			0	2	6	30	314	3	0
11	OTU_10	unculture	99.631	4.21E-135	484.5891			23	126	88	57	168	36	13
12	OTU_11	unculture	99.26199	5.13E-134	480.9824			71	45	64	0	0	145	139
13	OTU_12	ab537170.1	99.26199	5.13E-134	480.9824			70	39	21	25	3	196	5
14	OTU_13	pseudom	99.63235	1.47E-134	482.7858			411	7	16	3	1	16	14
15	OTU_14	unculture	98.89299	2.65E-131	471.9656			57	67	116	2	15	50	241
16	OTU_15	unculture	98.89299	2.65E-131	471.9656			130	94	56	49	15	128	10
17	OTU_16	unculture	98.15498	1.37E-128	462.9487			5	46	169	8	56	44	37
18	OTU_17	unculture	99.26199	6.22E-133	477.3757			161	17	75	0	3	111	5
19	OTU_18	deinococc	98.89299	2.65E-131	471.9656			11	29	11	45	116	15	9
20	OTU_19	piscinibac	99.26199	5.13E-134	480.9824			14	78	78	118	39	30	52
21	OTU_20	dq768114.1	99.26199	5.13E-134	480.9824			9	69	13	311	117	34	7

MrDNA data format

Possible PUMA Projects

- Incorporate additional analysis tools as outputs of the pipeline:

Phinch



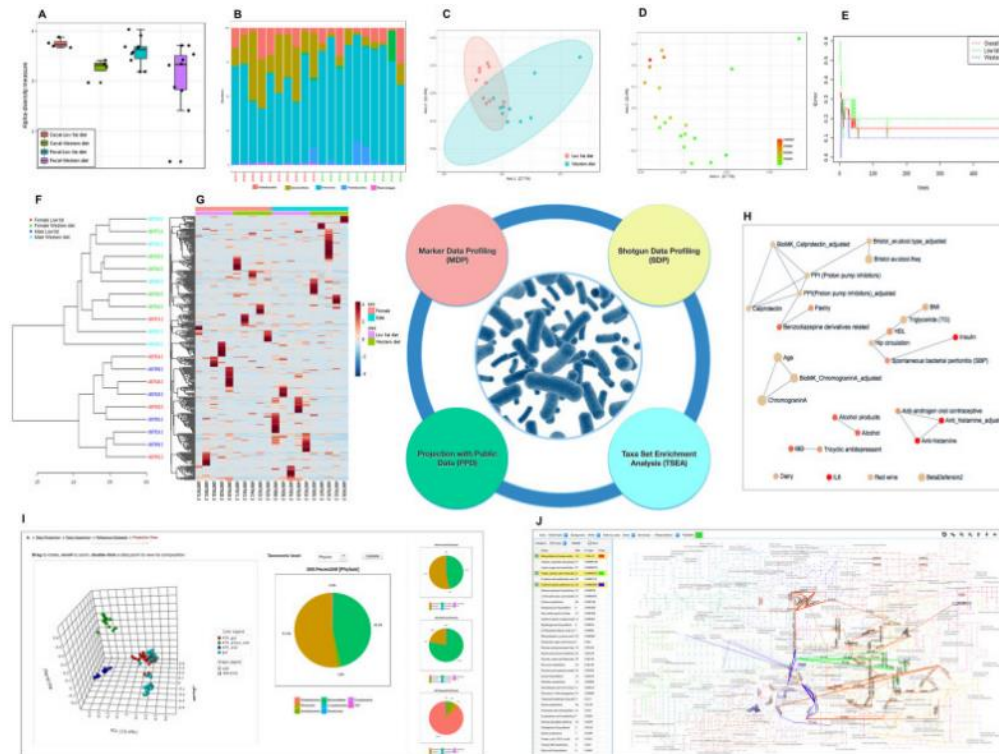
Exploratory data visualization

Bik 2014

Possible PUMA Projects

- Incorporate additional analysis tools as outputs of the pipeline:

MicrobiomeAnalyst



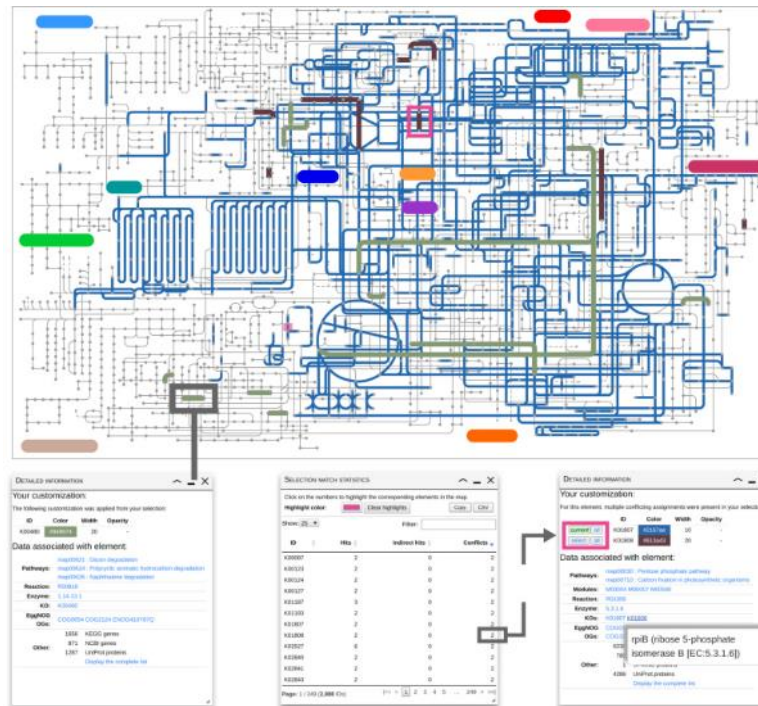
Diversity analysis
Differential abundance analysis

Dhariwal 2017

Possible PUMA Projects

- Incorporate additional analysis tools as outputs of the pipeline:

iPathway



Functional pathway visualization

Darzi 2018

Questions?

Links to tools cited

- <https://pathways.embl.de/>
- <https://www.microbiomeanalyst.ca/>
- <http://phinch.org/>