

Metagenome analysis

Fundamental metagenomic analysis

- 16S amplicon analysis
 - Demultiplexing, clustering OTUs and taxonomic profiling
 - Calculating and visualizing alpha- and beta-diversity
- Shotgun metagenomic analysis
 - Taxonomic profiling
 - Calculating and visualizing alpha- and beta-diversity

Ecological diversity for microbiome

- Meaning
 - The complexity and variation in ecosystems
- α -diversity and β -diversity



Shannon

$$H = - \sum_{i=1}^M P_i \log_2 P_i$$

Simpson

$$D = \sum (n / N)^2$$

- Commonly used metrics
 - α -diversity: Shannon and Simpson diversity, Faith's Phylogenetic Diversity
 - β -diversity: UniFrac distance, Bray–Curtis dissimilarity

Richness of the species (function unit)

- Estimate the number of species present
 - Rarefaction and counting
 - Count the observed data
 - non-parametric species estimators, Chao1, Chao2...
 - Correct the richness for the unobserved low frequent ones

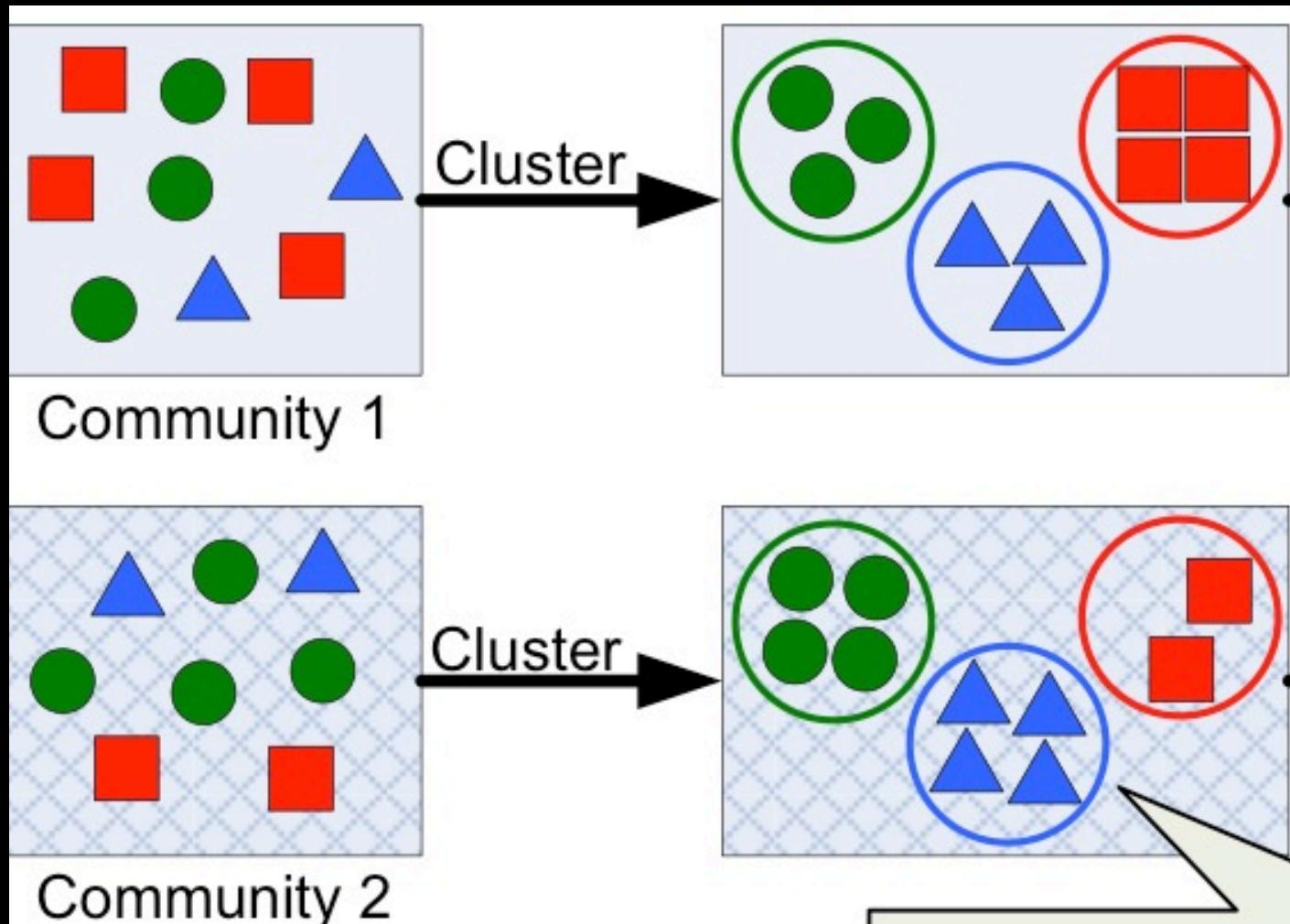
$$S_1 = S_{obs} + \frac{F_1^2}{2F_2}$$

Chao1

$$S_2 = S_{obs} + \frac{Q_1^2}{2Q_2}$$

Chao2

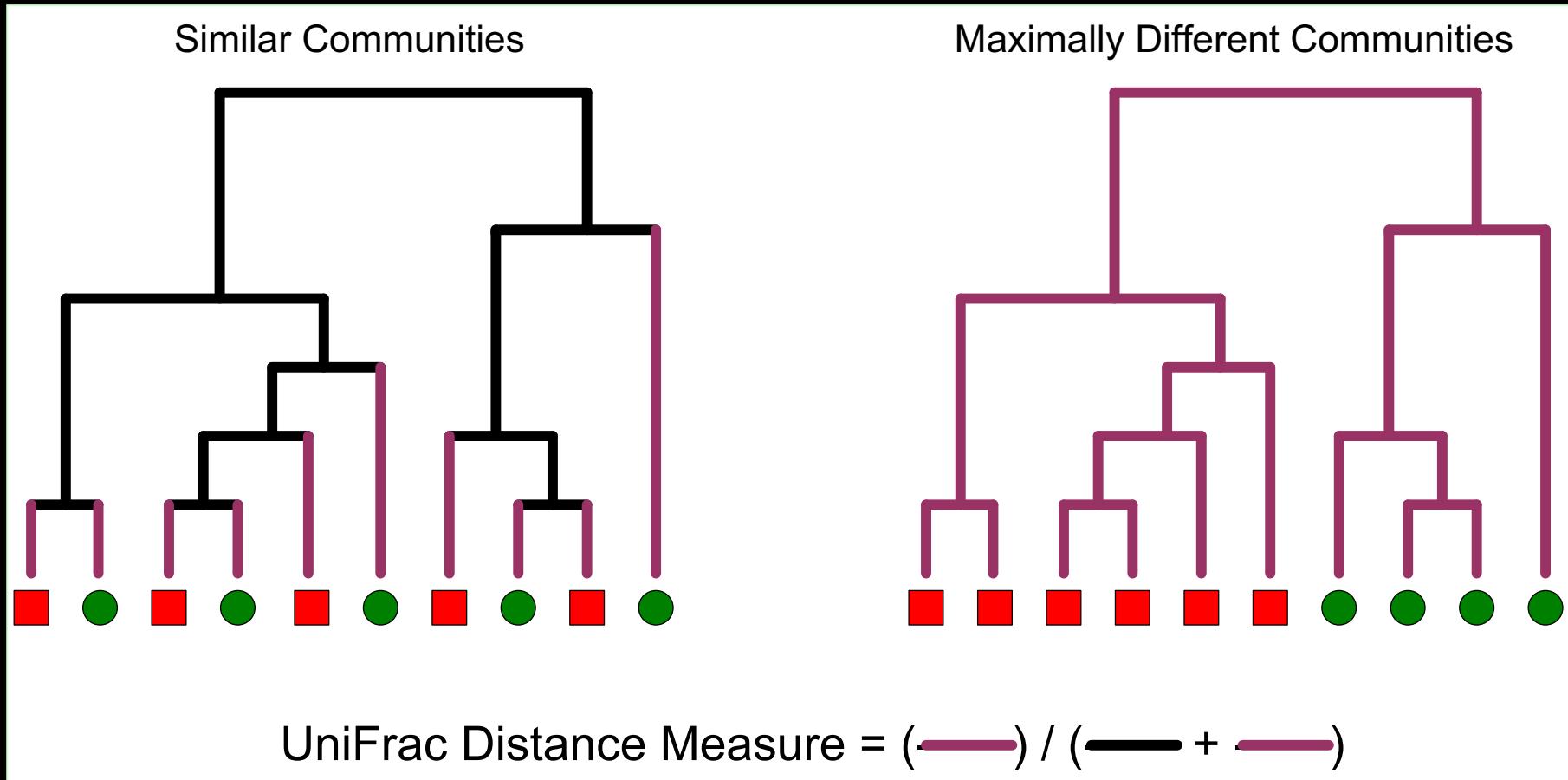
Bray–Curtis dissimilarity



$$d^{BCD}(i, j) = \frac{\sum_{k=0}^{n-1} |y_{i,k} - y_{j,k}|}{\sum_{k=0}^{n-1} (y_{i,k} + y_{j,k})}$$

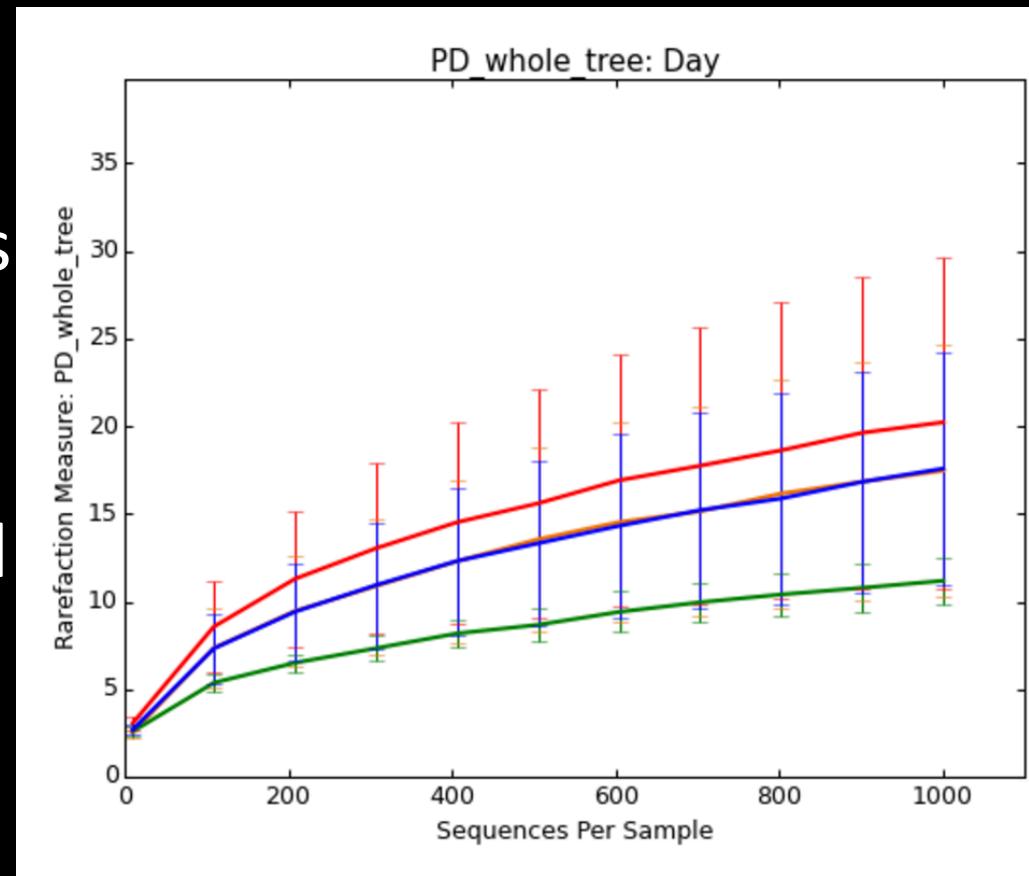
UniFrac distance

- Definition
 - β -diversity between organismal communities using phylogenetic information

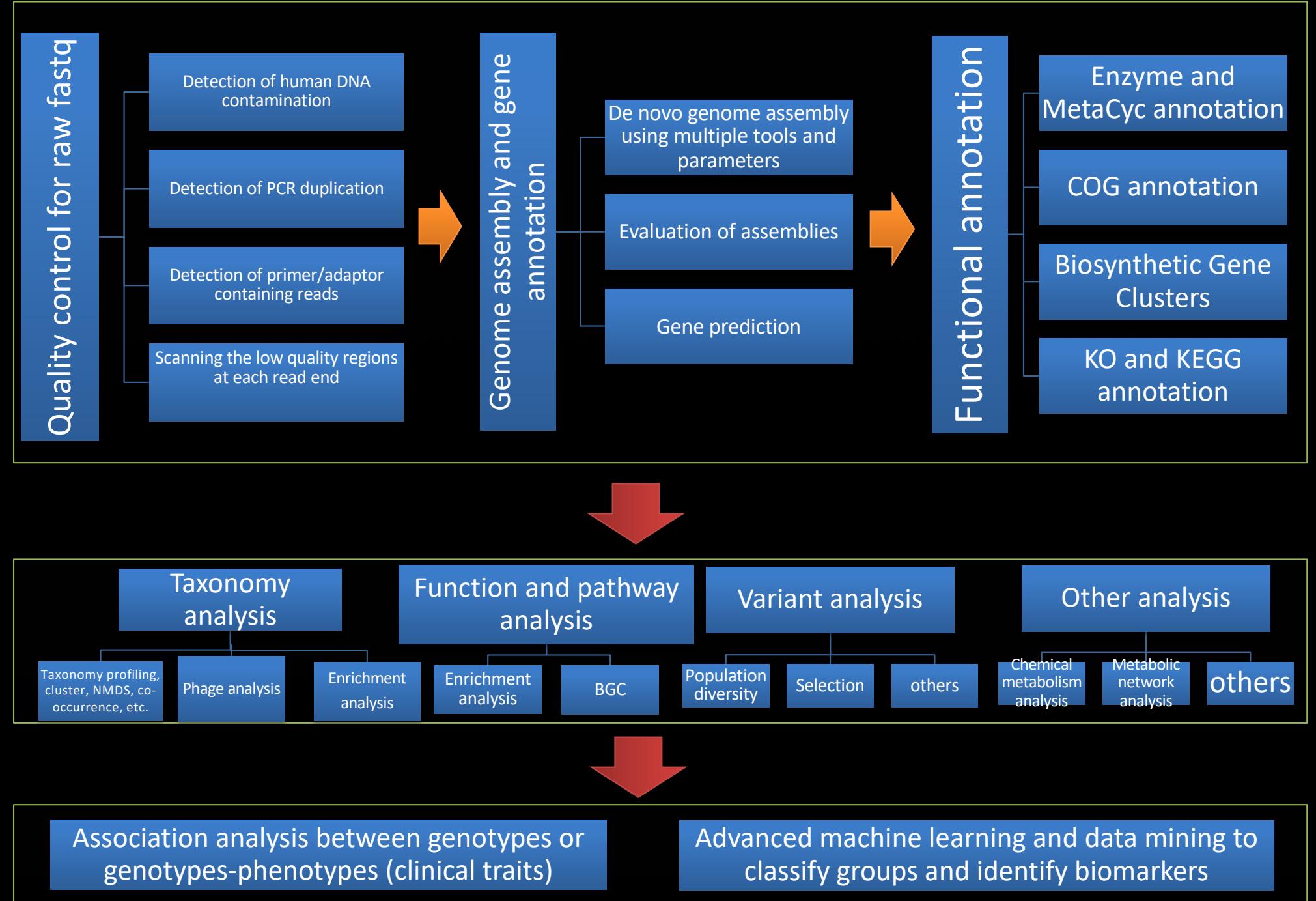


Rarefaction curve

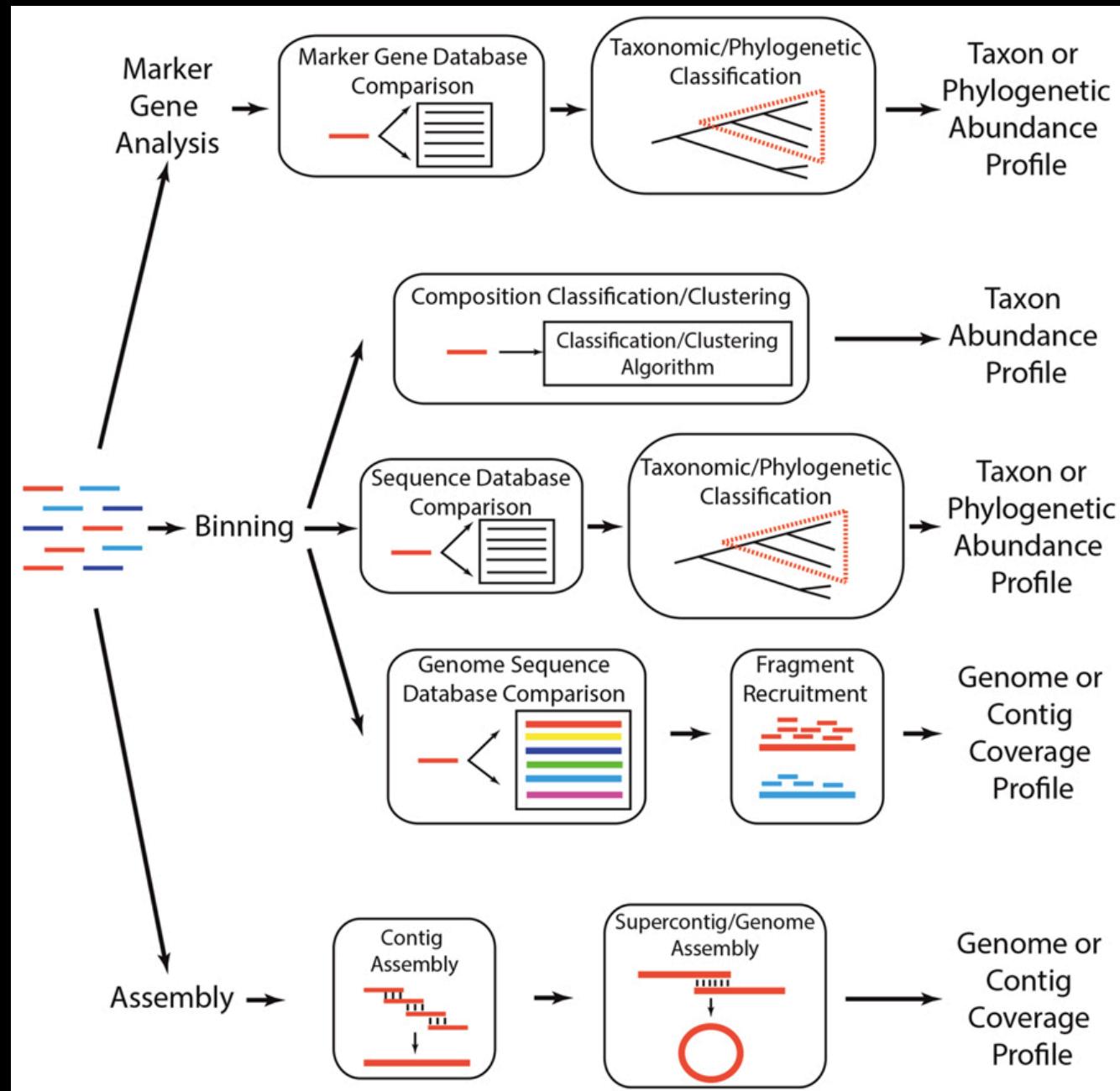
- Rarefaction
 - A subsampling procedure frequently used in ecology studies
- Rarefaction curve
 - A series of subsampled points (with different depth) measuring alpha-diversity or richness,etc.



Workflow for metagenomic analysis



Strategies for taxonomy profiling in shotgun metagenomics



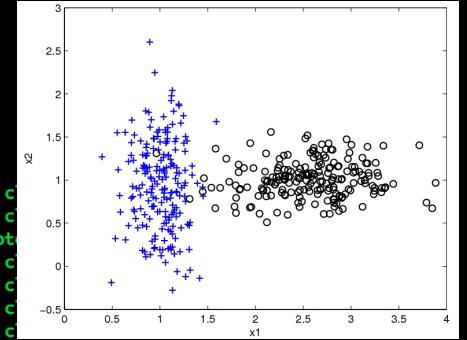
Which software should be used in taxonomy profiling for shotgun data

- Softwares
 - RDP classifier
 - Megan
 - Metaphlan
 - MetaPhyler
 - Phymm
 - Kraken
 - Kaiju
 -
- 16S, shotgun reads or conserved markers reads?
 - For Human microbiome, conserved markers
 - Other microbiome, 16S or all shotgun reads

Taxonomy profiling based on 16S using RDP classifier

- Principle
 - Naïve Bayes classifier based on percentage of shared seven-character words between a query and database sequences
- Results

HWI-C00135:94:C4KWJACXX:8:2315:7395:85490	Bacteria	domain	0.99	Firmicutes	phylum	0.4	Clostridia	class	0.2	Methanopyrales	order	
HWI-C00135:94:C4KWJACXX:8:1211:7456:9986	Bacteria	domain	1.0	Firmicutes	phylum	0.58	Clostridia	class	0.96	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:2312:4964:92073	-	Bacteria	domain	0.99	"Proteobacteria"	phylum	0.71	Alphaproteobacteriales	class	0.97	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:1312:2681:63438	-	Bacteria	domain	0.99	Firmicutes	phylum	0.86	Clostridia	class	0.81	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:2106:8432:14019	Bacteria	domain	1.0	Firmicutes	phylum	0.7	Clostridia	class	0.98	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2102:4859:8966	Bacteria	domain	1.0	Firmicutes	phylum	0.82	Clostridia	class	0.99	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2314:15770:46870	Bacteria	domain	1.0	Firmicutes	phylum	0.64	Clostridia	class	0.99	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2105:19690:44073	-	Archaea	domain	0.36	"Euryarchaeota"	phylum	0.29	Methanopyri	class	0.2	Methanopyrales	order
HWI-C00135:94:C4KWJACXX:8:1306:20670:85769	Bacteria	domain	0.99	Firmicutes	phylum	0.96	Clostridia	class	0.96	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:2104:18856:16320	Bacteria	domain	1.0	"Bacteroidetes"	phylum	1.0	"Bacteroidia"	class	0.97	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:1214:9691:78233	Bacteria	domain	0.99	"Bacteroidetes"	phylum	0.91	"Bacteroidia"	class	0.81	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2104:20072:26972	-	Bacteria	domain	1.0	"Bacteroidetes"	phylum	0.99	"Bacteroidia"	class	0.99	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:1102:10897:74181	Bacteria	domain	0.99	Firmicutes	phylum	0.78	Clostridia	class	0.73	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:1215:7135:2658	Bacteria	domain	0.99	Firmicutes	phylum	0.83	Clostridia	class	0.81	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:1213:17603:84829	Bacteria	domain	1.0	Firmicutes	phylum	0.97	Clostridia	class	0.97	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:1106:5230:75525	Bacteria	domain	1.0	"Bacteroidetes"	phylum	1.0	"Bacteroidia"	class	0.94	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:1303:14277:10624	-	Bacteria	domain	1.0	Firmicutes	phylum	0.97	Clostridia	class	0.93	Clostridiale	family
HWI-C00135:94:C4KWJACXX:8:1204:3521:62719	-	Bacteria	domain	1.0	"Proteobacteria"	phylum	0.35	Betaproteobacteria	class	0.18		
HWI-C00135:94:C4KWJACXX:8:1207:2131:3545	Bacteria	domain	1.0	Firmicutes	phylum	0.99	Clostridia	class	0.97	Clostridiale	family	
HWI-C00135:94:C4KWJACXX:8:2215:8729:54569	-	Bacteria	domain	1.0	Firmicutes	phylum	0.99	Clostridia	class	0.98	Clostridiale	family
HWI-C00135:94:C4KWJACXX:8:1211:12329:40223	Bacteria	domain	0.99	"Bacteroidetes"	phylum	0.55	"Bacteroidia"	class	0.24	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2308:8490:39424	-	Bacteria	domain	0.99	"Bacteroidetes"	phylum	0.75	"Bacteroidia"	class	0.45	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:2112:1827:41149	-	Bacteria	domain	0.96	"Bacteroidetes"	phylum	0.67	"Bacteroidia"	class	0.58	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:1314:5314:45072	-	Bacteria	domain	1.0	Firmicutes	phylum	0.87	Clostridia	class	0.87	Clostridiale	family
HWI-C00135:94:C4KWJACXX:8:1106:14103:28551	-	Bacteria	domain	1.0	"Bacteroidetes"	phylum	0.98	"Bacteroidia"	class	0.97	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:2304:20196:96499	-	Bacteria	domain	0.93	"Bacteroidetes"	phylum	0.51	Sphingobacteriia	class	0.49	"Sph	
HWI-C00135:94:C4KWJACXX:8:2316:12608:79481	-	Bacteria	domain	0.97	Firmicutes	phylum	0.92	Clostridia	class	0.92	Clostridiale	family
HWI-C00135:94:C4KWJACXX:8:2313:6488:34071	-	Bacteria	domain	1.0	"Bacteroidetes"	phylum	0.93	"Bacteroidia"	class	0.81	"Bacteroidal"	class
HWI-C00135:94:C4KWJACXX:8:2116:11946:91068	Bacteria	domain	1.0	"Bacteroidetes"	phylum	0.99	"Bacteroidia"	class	0.98	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2115:1487:37341	Bacteria	domain	0.98	"Bacteroidetes"	phylum	0.53	"Bacteroidia"	class	0.48	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:1311:8060:3573	Bacteria	domain	0.98	"Bacteroidetes"	phylum	0.86	"Bacteroidia"	class	0.79	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:2105:18517:96615	Bacteria	domain	0.99	"Bacteroidetes"	phylum	0.92	"Bacteroidia"	class	0.72	"Bacteroidal"	class	
HWI-C00135:94:C4KWJACXX:8:1213:13063:59198	-	Bacteria	domain	0.99	Firmicutes	phylum	0.4	Clostridia	class	0.28	Halanaerobia	class
HWI-C00135:94:C4KWJACXX:8:1106:8094:68206	Bacteria	domain	0.92	"Bacteroidetes"	phylum	0.39	Sphingobacteriia	class	0.38	"Sph		



Taxonomy profiling based on conserved markers using Metaphlan

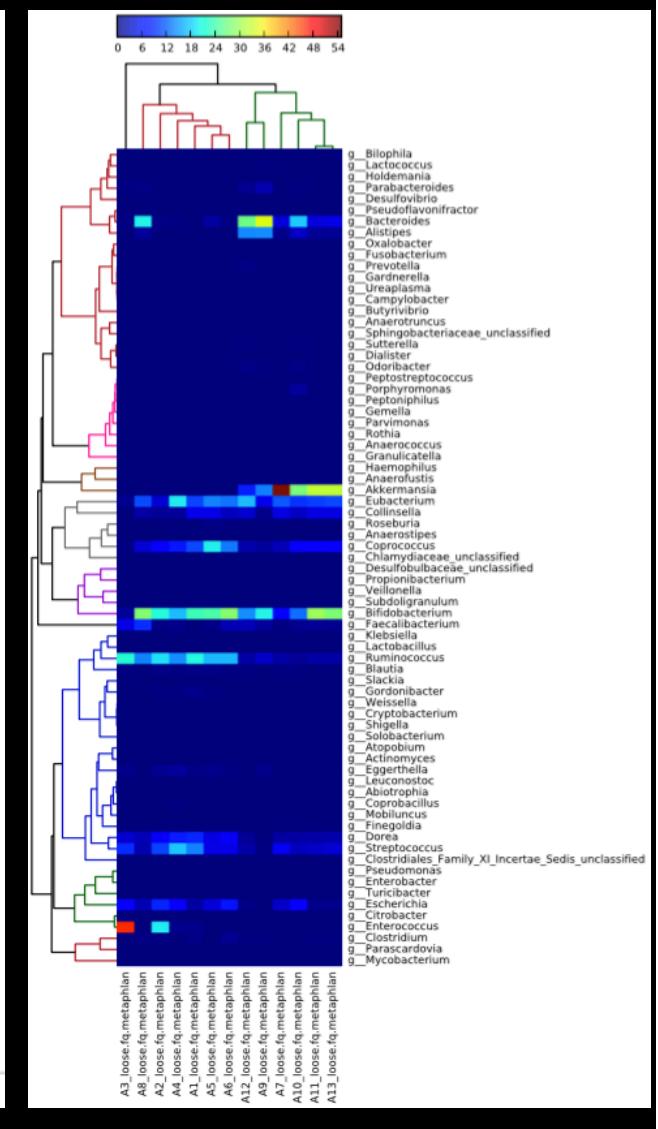
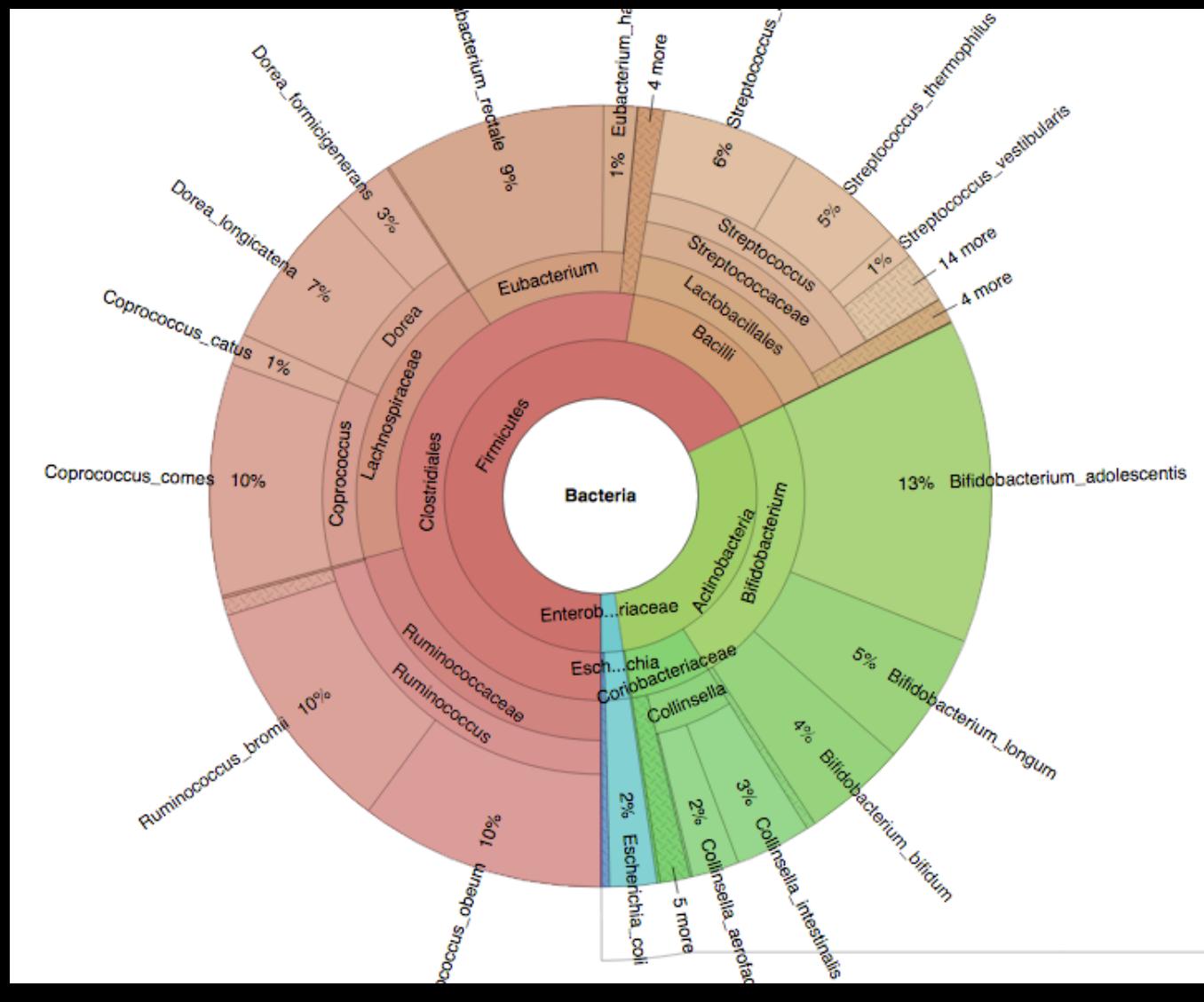
- Principle
 - Mapping reads to clade-specific marker genes
- Results

```
k_Bacteria      100.0
k_Bacterialp_Proteobacteria   87.41688
k_Bacterialp_Bacteroidetes    12.58312
k_Bacterialp_Proteobacterialc_Epsilonproteobacteria 87.41688
k_Bacterialp_Bacteroideteslc_Bacteroidia    12.58312
k_Bacterialp_Proteobacterialc_Epsilonproteobacterialo_Campylobacterales    87.41688
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidales    12.58312
k_Bacterialp_Proteobacterialc_Epsilonproteobacterialo_Campylobacteraleslf_Helicobacteraceae      87.41688
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Bacteroidaceae 10.90048
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Rikenellaceae  1.68264
k_Bacterialp_Proteobacterialc_Epsilonproteobacterialo_Campylobacteraleslf_Helicobacteraceaelg_Helicobacter      87.41688
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Bacteroides    10.90048
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Rikenellaceaelg_Alistipes    1.68264
k_Bacterialp_Proteobacterialc_Epsilonproteobacterialo_Campylobacteraleslf_Helicobacteraceaelg_Helicobacter_hepticus    87.4:
88
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Bacteroidaceaelg_Bacteroidesls_Bacteroides_unclassified    10.90048
k_Bacterialp_Bacteroideteslc_Bacteroidialo_Bacteroidaleslf_Rikenellaceaelg_Alistipesls_Alistipes_unclassified    1.68264
ControlD0.sub.fa.metaphlan.out (END)
```

Tools used for visualization

- Krona
 - Visualize hierarchical data with zoomable pie charts
- MetaPhlAn
 - Profile bacterial community
 - Visualize the community composition using heatmap
- In-house scripts
 - Manipulate/format the results

Visualize taxonomy profile and cluster multiple samples according to the profiles



Practice

- Taxonomy profiling based on 16S markers
- Taxonomy profiling using evolutionary conserved markers using Metaphlan
- Visualize the taxonomy profile using krona for single sample
- Visualize the taxonomy profile from multiple samples and cluster the samples according to the profile

Practice

- Calculate taxonomic alpha-diversity (Shannon and Simpson) within each sample (mouse liver cancer)
- Calculate Unifrac distance (both unweighted and weighted) between samples
- Calculate functional Bray–Curtis distance between samples and visualize it
- Calculate Chao1 richness
- Calculate the gene abundance in metagenomic samples