Lecture 20

Metagenomic data analysis

MCB 416A/516A Statistical Bioinformatics and Genomic Analysis

Prof. Lingling An Univ of Arizona

Metagenomic data analysis pipeline

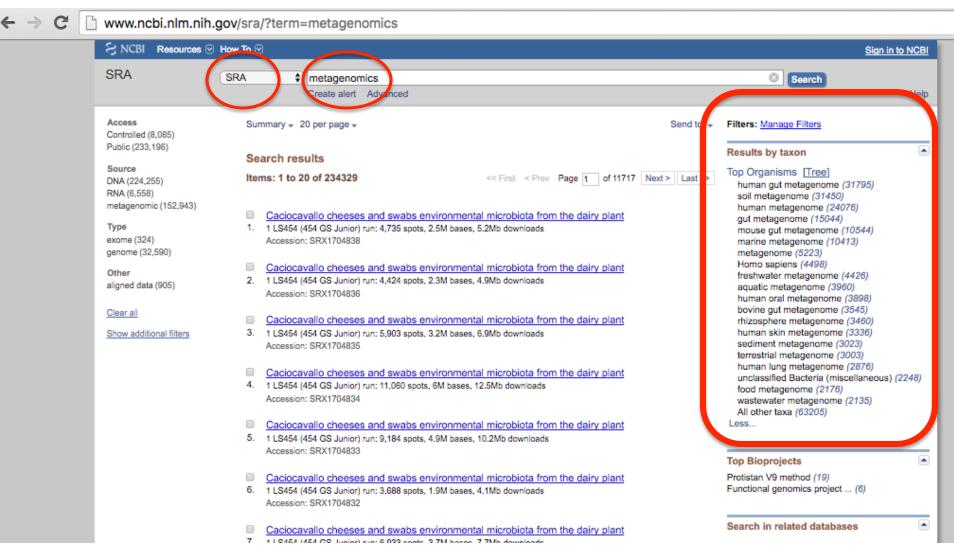
- Sequence data to count data
- Downstream analysis
 - Visualization (among samples)
 - Differential abundant analysis
 - Network analysis (among features)

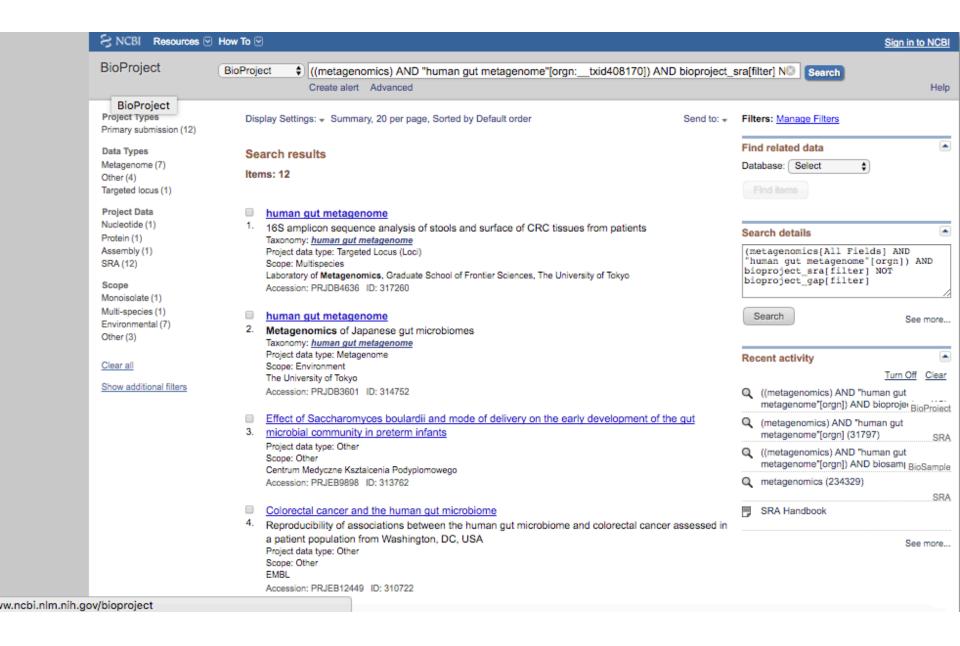
— ...

From sequence data to count data

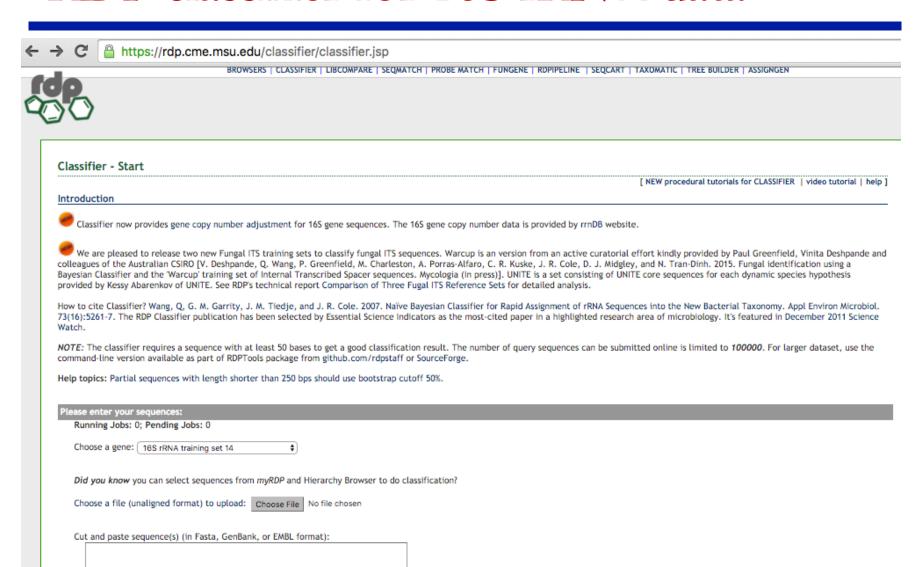
- Metagenomic sequence data
 - whole genome
 - marker genes (e.g., 16s rRNA, or 18S rRNA)
- Download datasets
 - NCBI -SRA
 - MG-RAST
 - Papers
 - ...
- Mapping sequence to reference database
 - blast or blastx (for whole genome data)
 - Use RDP classifier (for 16S data)
- In R
 - Read alignment results into R and make count table

Download 16S rRNA dataset

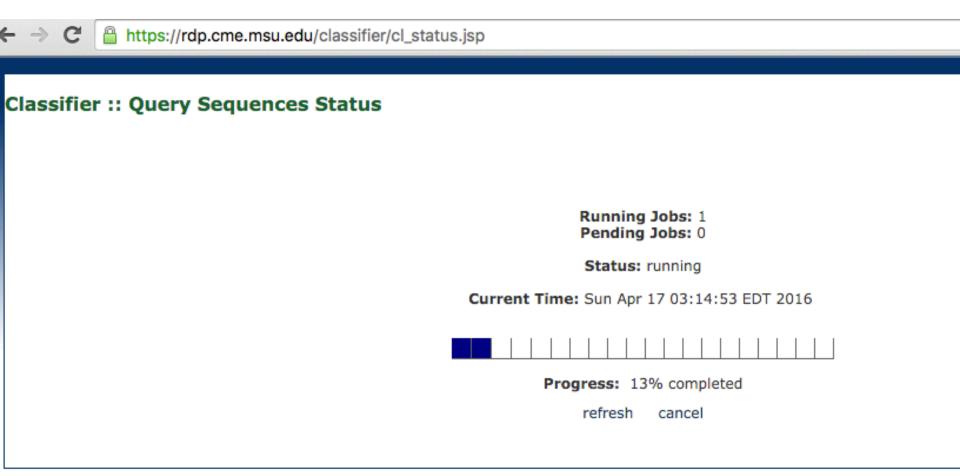




RDP classifier for 16S rRNA data



Submit a .fastq dataset and it's running...



Download the alignment result

BROWSERS | CLASSIFIER | LIBCOMPARE | SEQMATCH | PROBE MATCH | FUNGENE | RDPIPELINE | SEQCART | TAXOMATIC | TREE BUILDER | ASSIGNGEN



Classifier :: Assignment Detail

[start over | hierarchy view | help]

Classifier: RDP Naive Bayesian rRNA Classifier Version 2.10, October 2014

Taxonomical Hierarchy: RDP 16S rRNA training set 14

Query File: ERR345422.fastq

Query Submit Date: Sun Apr 17 03:16:39 EDT 2016

Lineage (click to return to particular node):

Root (30442)		
Assignment Det	etail (for Root with Confidence threshold: 80%): download allrank resu	xrank result
ERR345422.1	Root[100%] Bacteria[99%] "Bacteroidetes"[96%] "Bacteroidia"[92%] "Bacteroidales"[92%] "Porphyromonadaceae"[88%] Butyricimonas[46%]	
ERR345422.2	Root[100%] Bacteria[100%] Firmicutes[100%] Bacilli[100%] Lactobacillales[100%] Lactobacillaceae[96%] Lactobacillus[71%]	
ERR345422.3	Root[100%] Bacteria[100%] "Bacteroidetes"[91%] "Bacteroidia"[59%] "Bacteroidales"[59%] "Porphyromonadaceae"[37%] Barnesiella[9%]	
ERR345422.4	Root[100%] Bacteria[95%] "Bacteroidetes"[89%] "Bacteroidia"[67%] "Bacteroidales"[67%] "Porphyromonadaceae"[48%] Dysgonomonas[11%]	
ERR345422.5	Root[100%] Bacteria[99%] "Bacteroidetes"[95%] "Bacteroidia"[61%] "Bacteroidales"[61%] "Porphyromonadaceae"[57%] Barnesiella[8%]	
ERR345422.6	Root[100%] Bacteria[100%] "Bacteroidetes"[91%] "Bacteroidia"[60%] "Bacteroidales"[60%] "Porphyromonadaceae"[38%] Barnesiella[9%]	
ERR345422.7	Root[100%] Bacteria[100%] "Bacteroidetes"[88%] "Bacteroidia"[69%] "Bacteroidales"[69%] "Porphyromonadaceae"[50%] Barnesiella[12%]	
ERR345422.8	Root[100%] Bacteria[100%] Firmicutes[98%] Bacilli[98%] Lactobacillales[98%] Lactobacillaceae[87%] Lactobacillus[84%]	
ERR345422.9	Root[100%] Bacteria[97%] "Bacteroidetes"[82%] "Bacteroidia"[70%] "Bacteroidales"[70%] "Porphyromonadaceae"[54%] Barnesiella[16%]	
ERR345422.10	Root[100%] Bacteria[99%] "Bacteroidetes"[94%] "Bacteroidia"[88%] "Bacteroidales"[88%] "Porphyromonadaceae"[67%] Dysgonomonas[19%]	
ERR345422.11	Root[100%] Bacteria[100%] "Bacteroidetes"[99%] "Bacteroidia"[81%] "Bacteroidales"[81%] "Porphyromonadaceae"[72%] Dysgonomonas[32%]	
ERR345422.12	Root[100%] Bacteria[100%] Firmicutes[100%] Clostridia[100%] Clostridiales[100%] Lachnospiraceae[100%] Lactonifactor[54%]	
ERR345422.13	Root[100%] Bacteria[100%] "Bacteroidetes"[87%] "Bacteroidia"[74%] "Bacteroidales"[74%] "Porphyromonadaceae"[54%] Barnesiella[14%]	



Classifier: RDP Naive Bayesian rRNA Classifier Version 2.10, October 2014
Taxonomical Hierarchy: RDP 16S rRNA training set 14
Query File: ERR345422.fastq
Submit Date: Thu Oct 29 17:56:13 EDT 2015
Confidence threshold (for classification to Root ONLY): 80%
Symbol +/- indicates predicted sequence orientation

ERR345422.2:+:Bacteria:100%;Firmicutes:100%;Bacilli:100%;Lactobacillales:100%;Lactobacillaceae:96%;Lactobacillus:71% ERR345422.3;+;Bacteria;100%;"Bacteroidetes";91%;"Bacteroidia";59%;"Bacteroidales";59%;"Porphyromonadaceae";36%;Barnesiella;8% ERR345422.4;+;Bacteria;95%;"Bacteroidetes";89%;"Bacteroidia";67%;"Bacteroidales";67%;"Porphyromonadaceae";48%;Dysgonomonas;11% ERR345422.5;+;Bacteria;99%;"Bacteroidetes";95%;"Bacteroidia";61%;"Bacteroidales";61%;"Porphyromonadaceae";57%;Barnesiella;8% ERR345422.6;+;Bacteria;100%;"Bacteroidetes";91%;"Bacteroidia";59%;"Bacteroidales";59%;"Porphyromonadaceae";37%;Barnesiella;9% ERR345422.7;+;Bacteria;100%;"Bacteroidetes";88%;"Bacteroidia";69%;"Bacteroidales";69%;"Porphyromonadaceae";50%;Barnesiella;12% ERR345422.8;+;Bacteria;100%;Firmicutes;98%;Bacilli;98%;Lactobacillales;98%;Lactobacillaceae;87%;Lactobacillus;84% ERR345422.9;+;Bacteria;97%;"Bacteroidetes";82%;"Bacteroidia";71%;"Bacteroidales";71%;"Porphyromonadaceae";55%;Barnesiella;15% ERR345422.10;+;Bacteria;99%;"Bacteroidetes";94%;"Bacteroidia";87%;"Bacteroidales";87%;"Porphyromonadaceae";64%;Dysgonomonas;18% ERR345422.11;+;Bacteria;100%;"Bacteroidetes";99%;"Bacteroidia";81%;"Bacteroidales";81%;"Porphyromonadaceae";72%;Dysgonomonas;32% ERR345422.12;+;Bacteria;100%;Firmicutes;100%;Clostridia;100%;Clostridiales;100%;Lachnospiraceae;100%;Lactonifactor;53% ERR345422.13;+;Bacteria;100%;"Bacteroidetes";87%;"Bacteroidia";74%;"Bacteroidales";74%;"Porphyromonadaceae";53%;Barnesiella;14% ERR345422.14;+;Bacteria;100%;Firmicutes;100%;Bacilli;100%;Lactobacillales;100%;Lactobacillaceae;96%;Lactobacillus;71% ERR345422.15;+;Bacteria;95%;Firmicutes;92%;Clostridia;92%;Clostridiales;92%;Lachnospiraceae;91%;Dorea;15% ERR345422.16;+;Bacteria;100%;Firmicutes;94%;Clostridia;94%;Clostridiales;94%;Lachnospiraceae;92%;Dorea;5닣 ERR345422.17;+;Bacteria;100%;Firmicutes;76%;Clostridia;75%;Clostridiales;75%;Lachnospiraceae;73%;Johnsonella;39% ERR345422.18;+;Bacteria;100%;Firmicutes;94%;Clostridia;94%;Clostridiales;94%;Lachnospiraceae;92%;Dorea;5% ERR345422.19;+;Bacteria;100%;Firmicutes;38%;Bacilli;38%;Lactobacillales;35%;Lactobacillaceae;18%;Paralactobacillus;3% ERR345422.20;+;Bacteria;100%;Firmicutes;98%;Clostridia;98%;Clostridiales;97%;Lachnospiraceae;86%;Lachnobacterium;36% ERR345422.21;+;Bacteria;100%;"Bacteroidetes";97%;"Bacteroidia";86%;"Bacteroidales";86%;"Porphyromonadaceae";79%;Butyricimonas;39% ERR345422.22;+;Bacteria;98%;"Bacteroidetes";92%;"Bacteroidia";69%;"Bacteroidales";69%;"Porphyromonadaceae";53%;Parabacteroides;15% ERR345422.23;+;Bacteria;100%;Firmicutes;93%;Bacilli;91%;Lactobacillales;71%;Lactobacillaceae;51%;Lactobacillus;47% ERR345422.24;+;Bacteria;98%;"Bacteroidetes";63%;"Bacteroidia";36%;"Bacteroidales";36%;Marinilabiliaceae;8%;Thermophagus;4% ERR345422.25;+;Bacteria;100%;Firmicutes;100%;Bacilli;100%;Lactobacillales;100%;Lactobacillaceae;99%;Lactobacillus;85% ERR345422.26;+;Bacteria;100%;Firmicutes;100%;Bacilli;100%;Lactobacillales;98%;Lactobacillaceae;91%;Lactobacillus;85% ERR345422.27;+;Bacteria;100%;"Bacteroidetes";91%;"Bacteroidia";59%;"Bacteroidales";59%;"Porphyromonadaceae";38%;Barnesiella;9% ERR345422.28;+;Bacteria;100%;"Bacteroidetes";95%;"Bacteroidia";69%;"Bacteroidales";69%;"Porphyromonadaceae";49%;Dysgonomonas;16% ERR345422.29;+;Bacteria;100%;"Bacteroidetes";97%;"Bacteroidia";55%;"Bacteroidales";55%;"Porphyromonadaceae";47%;Barnesiella;8% ERR345422.30;+;Bacteria;100%;"Bacteroidetes";94%;"Bacteroidia";81%;"Bacteroidales";81%;"Porphyromonadaceae";62%;Dysgonomonas;40% ERR345422.31;+;Bacteria;100%;"Bacteroidetes";89%;"Bacteroidia";71%;"Bacteroidales";71%;"Porphyromonadaceae";67%;Dysgonomonas;41% ERR345422.32;+;Bacteria;99%;Firmicutes;87%;Clostridia;87%;Clostridiales;87%;Lachnospiraceae;84%;Clostridium XlVa;39% ERR345422.33;+;Bacteria;99%;"Bacteroidetes";96%;"Bacteroidia";81%;"Bacteroidales";81%;"Porphyromonadaceae";67%;Dysgonomonas;30% ERR345422.34;+;Bacteria;100%;"Bacteroidetes";95%;"Bacteroidia";89%;"Bacteroidales";89%;"Prevotellaceae";86%;Prevotella;69%

ERR345422.1;+;Bacteria;99%;"Bacteroidetes";96%;"Bacteroidia";92%;"Bacteroidales";92%;"Porphyromonadaceae";88%;Butyricimonas;46%