XX. TÜRK KLİNİK MİKROBİYOLOJİ VE İNFEKSİYON HASTALIKLARI KONGRESİ

Mikrobiyom Çalışmalarının Analiz ve Değerlendirilmesinde Kullanılan Biyoinformatik Yöntemleri ve Kısıtlılıkları

Dr. Mert Ahmet Kuşkucu

İstanbul Üniversitesi-Cerrahpaşa

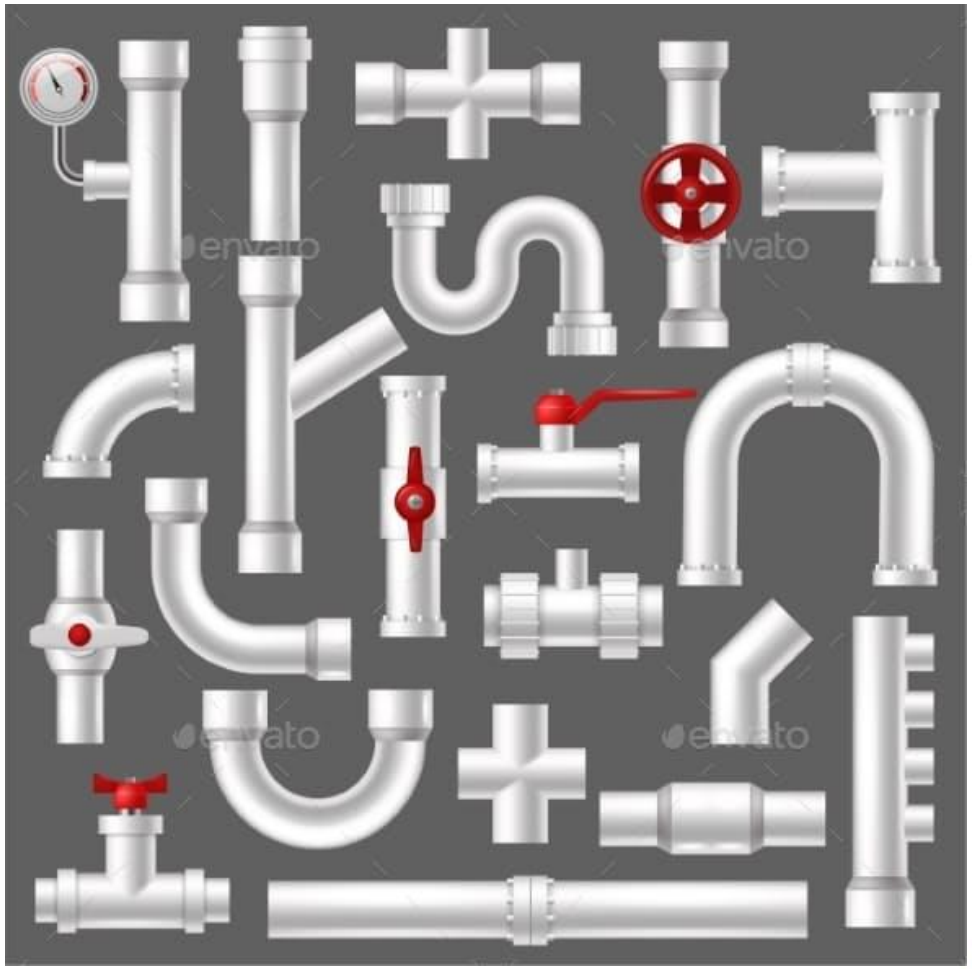
Cerrahpaşa Tıp Fakültesi

Tıbbi Mikrobiyoloji Anabilim Dalı

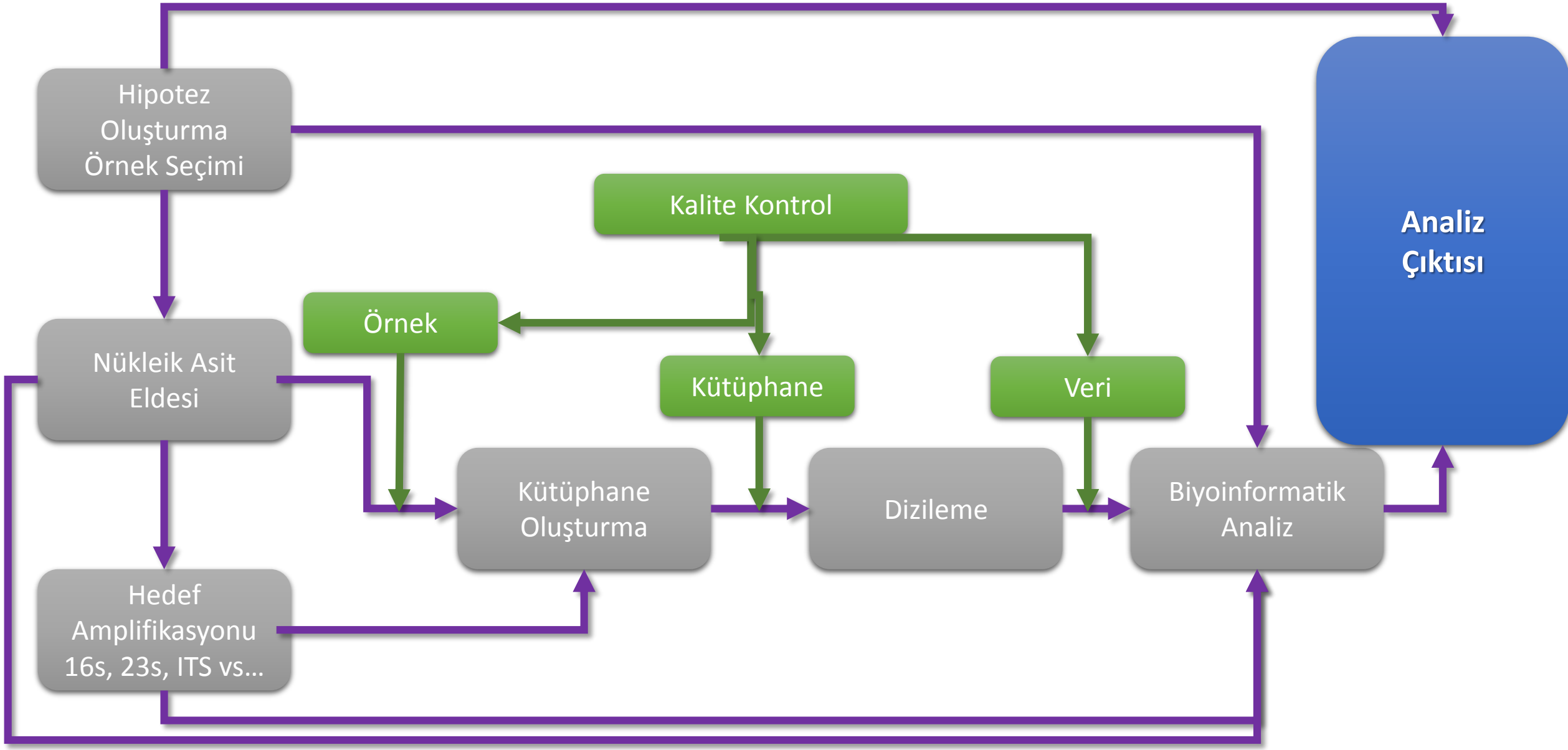
KLİMİK
TÜRK KLİNİK MİKROBİYOLOJİ VE İNFEKSİYON HASTALIKLARI DERNEĞİ

MİBİTÇG
KLİMİK DERNEĞİ MİKROBİYOM VE BİYOLOJİK TEDAVİLER ÇALIŞMA GRUBU

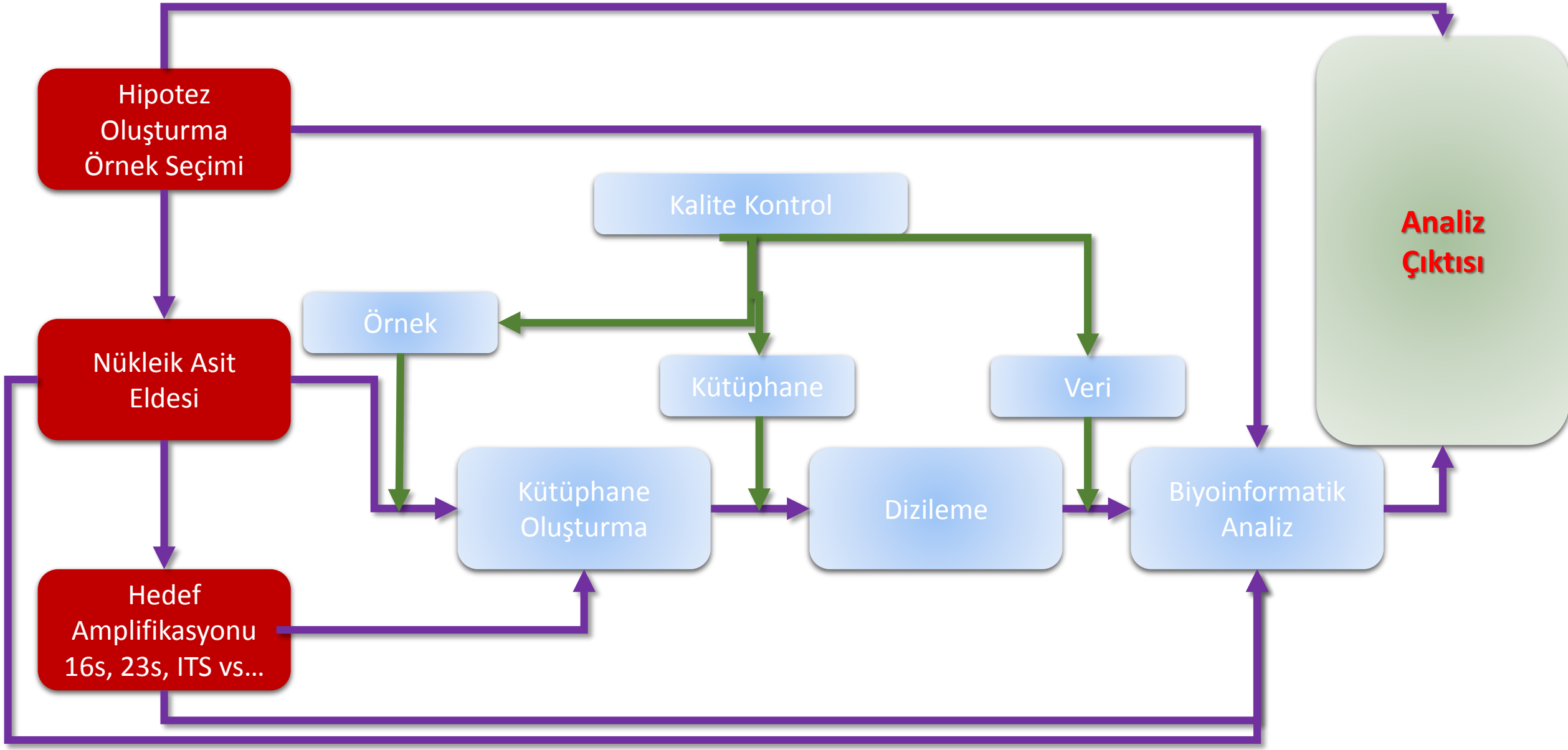
Puzelı çzmek için ‘Pipeline, Kavramı

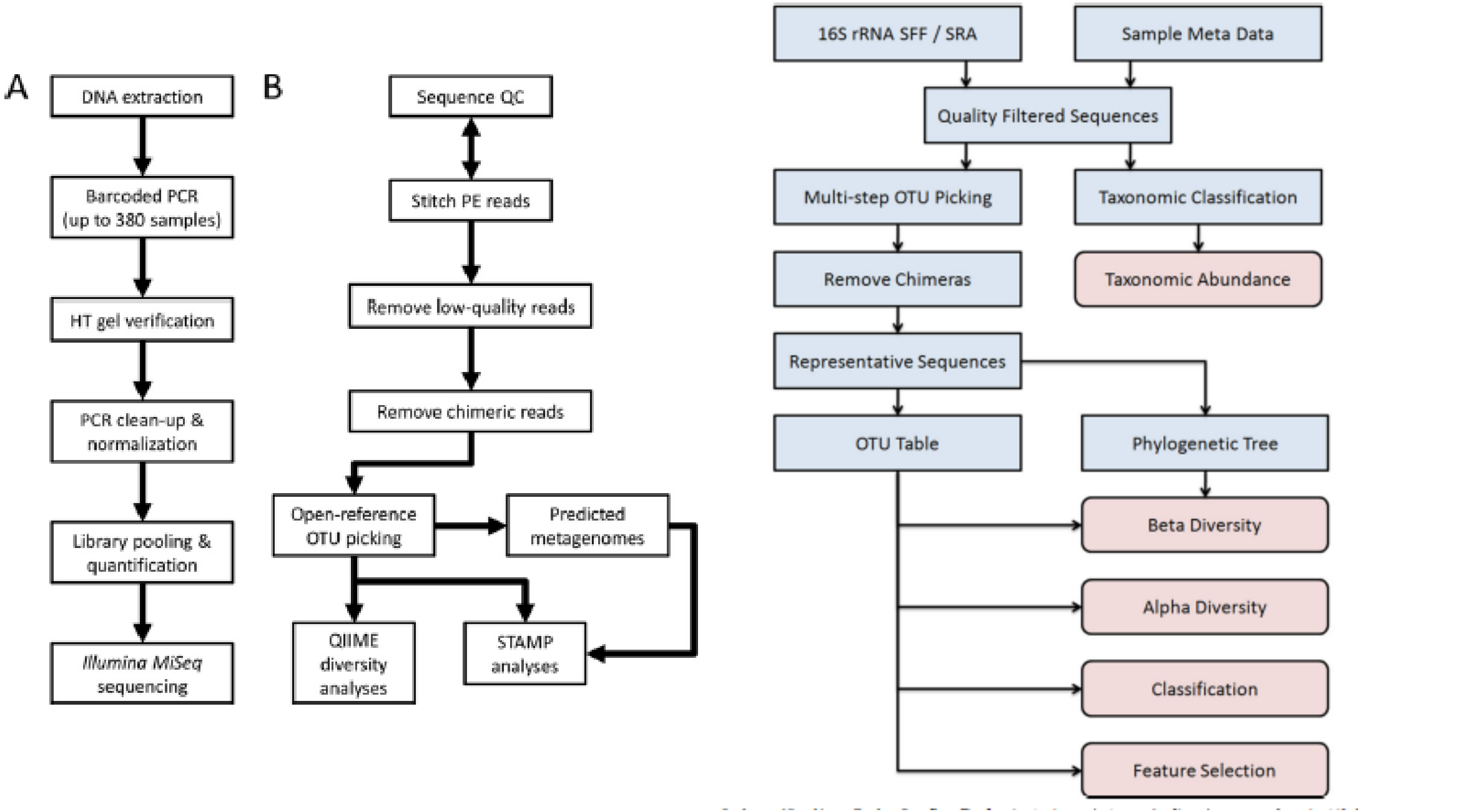


Mikrobiyom Çalışmaları için Temel İş Akışı; Ana ‘Pipeline’



Mikrobiyom➔ Metagenom Çalışmaları için Temel İş Akışı; Ana ‘Pipeline’




A
DNA extraction
Barcoded PCR (up to 380 samples)
HT gel verification
PCR clean-up & normalization
Library pooling & quantification
Illumina MiSeq sequencing
B
Sequence QC
Stitch PE reads
Remove low-quality reads
Remove chimeric reads
Open-reference OTU picking
Predicted metagenomes
QIIME diversity analyses
STAMP analyses
16S rRNA SFF / SRA
Sample Meta Data
Quality Filtered Sequences
Multi-step OTU Picking
Taxonomic Classification
Remove Chimeras
Taxonomic Abundance
Representative Sequences
OTU Table
Phylogenetic Tree
Beta Diversity
Alpha Diversity
Classification
Feature Selection

Illumina MiSeq

Illumina HiSeq

IonTorrent PGM

Roche 454

Pacific Biosciences RS

Images (.tiff)
Cluster intensity file (.cif)
Base call file (.bcl)

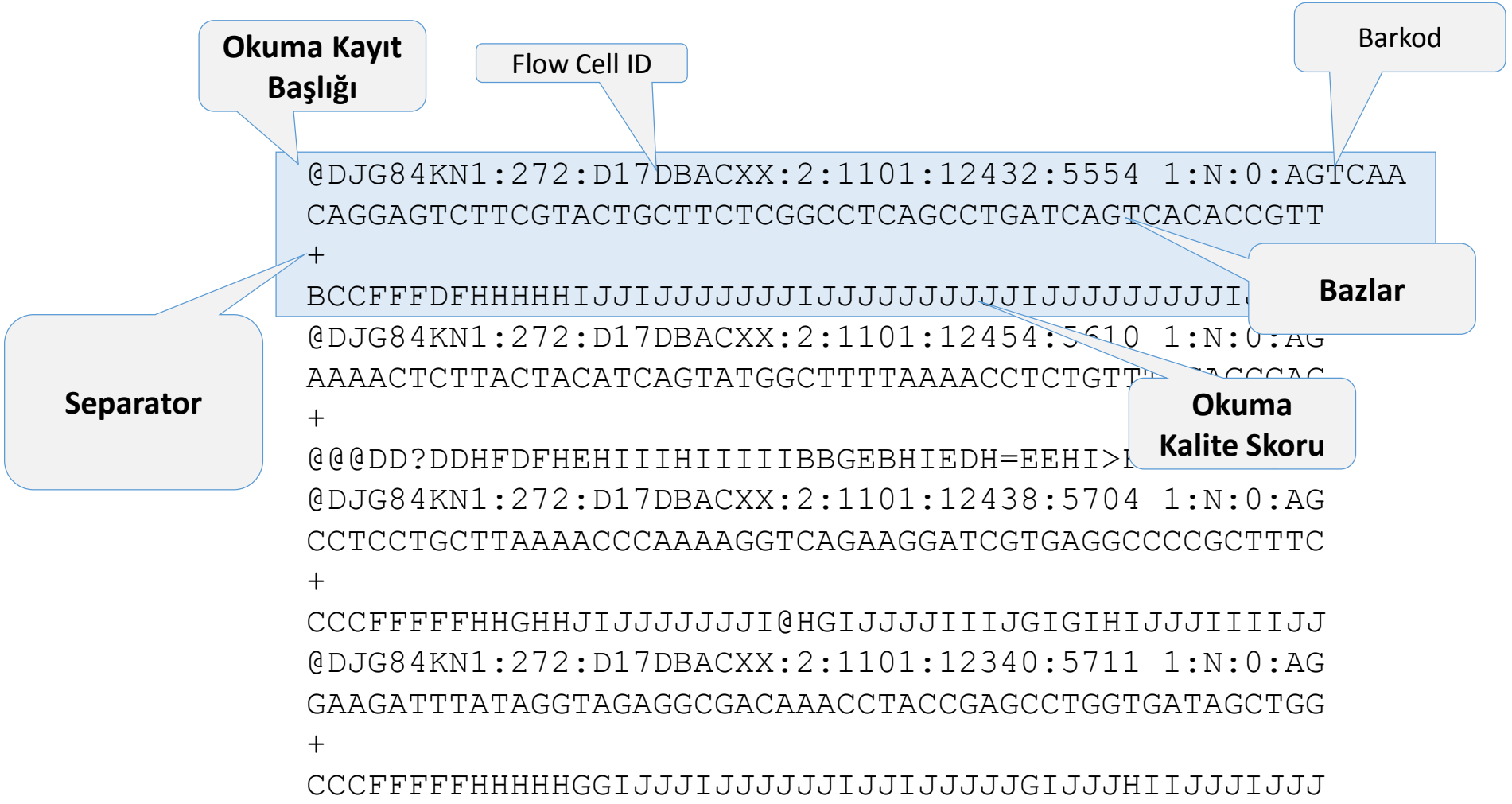
Standard flowgram file (.sff)

Movie
Trace (.trc.h5)
Pulse (.pls.h5)
Base (.bas.h5)

Sequence Data

(FASTQ Format)

FASTQ Format



Base İsimlendirme Kalitesi: Phred Kalite Skoru

Phred* kalite skoru *Q*

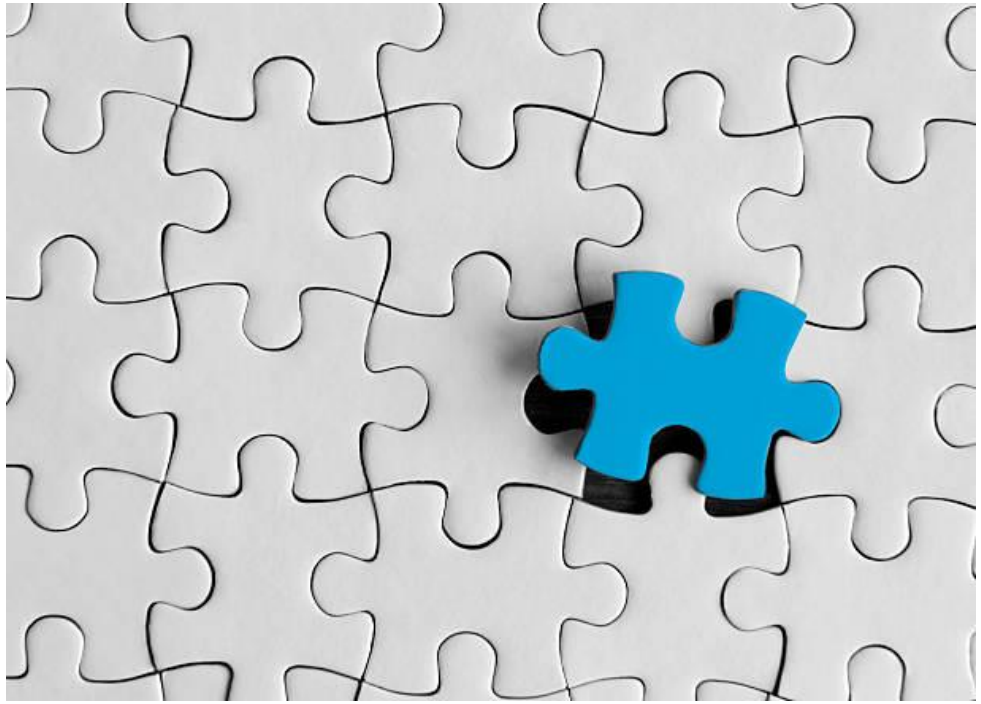
$$Q = -10 \log_{10} P$$

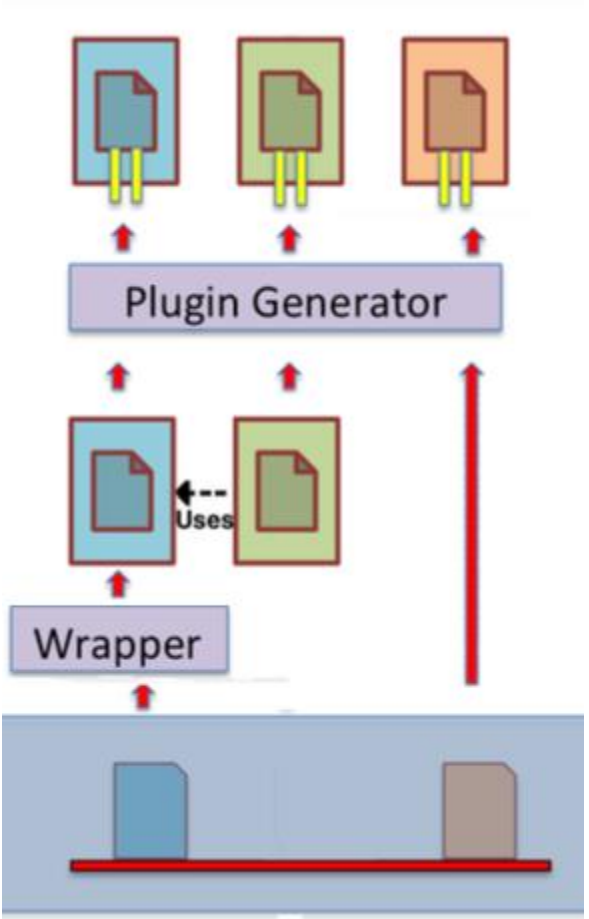
Q score	Olası Hata	Baz Doğruluğu	ASCII Karakteri
10	0.1	90%	“+”
20	0.01	99%	“5”
30	0.001	99.9%	“?”
40	0.0001	99.99%	“I”

```
 SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....................................................
 ..............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....................
 LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....................................................
 !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
 |                         |    |        |                              |                     |
33                        59   64       73                            104                   126

S - Sanger        Phred+33  range: 0 to 40
I - Illumina 1.3+ Phred+64  range: 0 to 40
L - Illumina 1.8+ Phred+33  range: 0 to 41
```

```
[benpass@solexalign]$ ls
Sample_FS53_EPCAM+_CD10-_IL2270-18
Sample_FS53_EPCAM+_CD10+_IL2269-19
Sample_COH77_CD49F-_IL2275-13
Sample_COH77_CD49F+_CD66-_IL2274-14
Sample_COH77_CD49F+_CD66+_IL2273-15
Sample_COH74_EPCAM+_CD10-_IL2272-16
Sample_COH74_EPCAM+_CD10+_IL2271-17
Sample_COH69_EPCAM+_CD10-_IL2268-20
Sample_COH69_EPCAM+_CD10+_IL2267-21

[benpass@solexalign]$ ls Sample_COH77_CD49F-_IL2275-13
COH77_CD49F-_IL2275-13_AGTCAA_L002_R1_001.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R1_002.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R1_003.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R1_004.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R2_001.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R2_002.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R2_003.fastq.gz
COH77_CD49F-_IL2275-13_AGTCAA_L002_R2_004.fastq.gz
```

gzip compressed





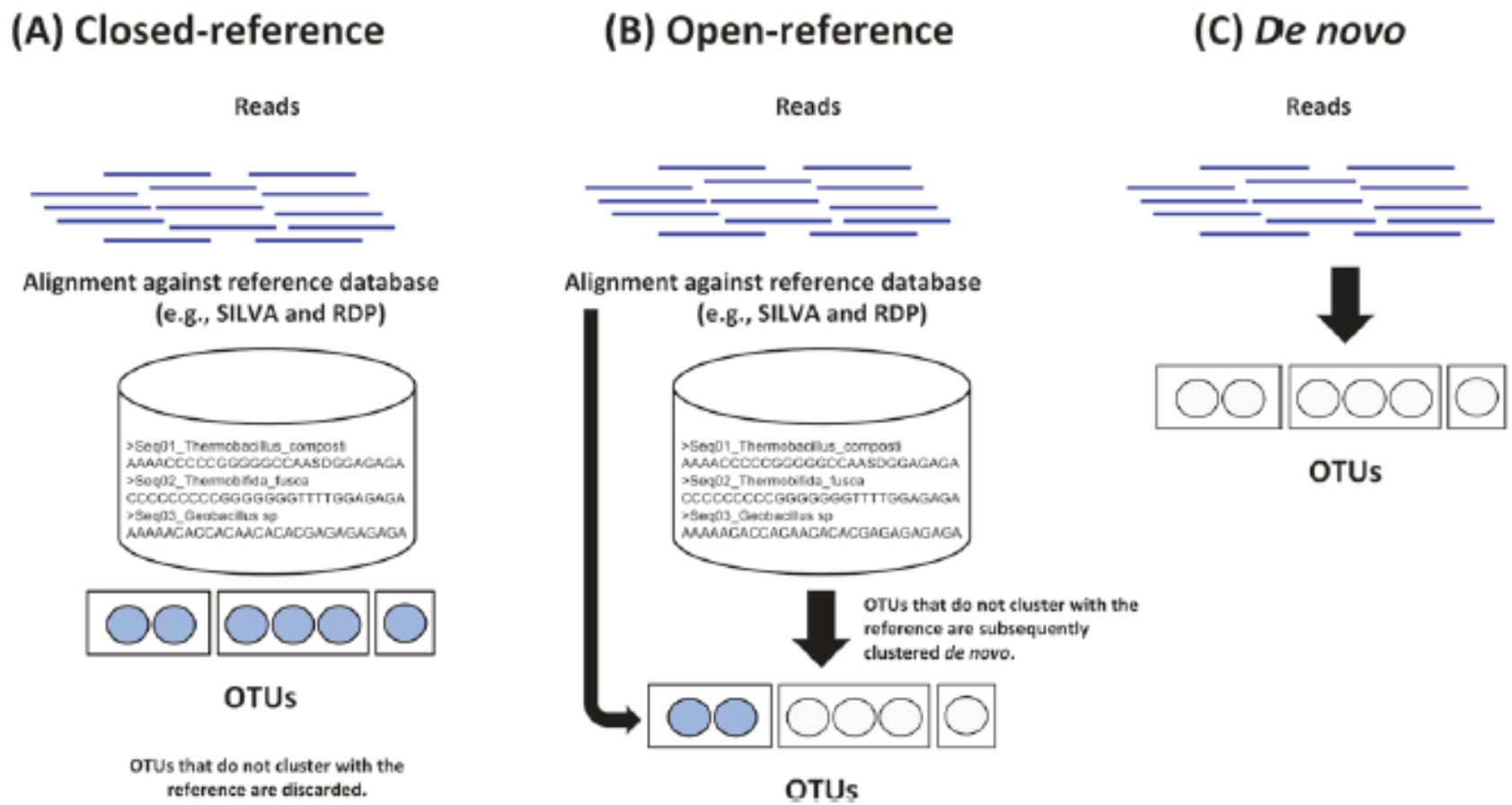


Fig. 1 Flowchart of data analysis depicting different options for picking operational taxonomic units. (**a**) Closed reference-based, (**b**) open-reference-based, and (**c**) de novo

MOCAT

Metagenomics assembly and profiling pipeline

MOCAT is a modular and scalable software pipeline for analyzing shotgun metagenomics datasets generated with Illumina technology. Starting from raw fastQ files, it can quality-filter and remove contaminants from them, assemble metagenomic reads into contigs, predict prokaryotic genes on these, identify phylogenetic marker genes and generate taxonomic abundance profiles by mapping reads to these marker genes.

Homepage **Source code** **Documentation**

Contact: zeller@embl.de
Version: 2.0

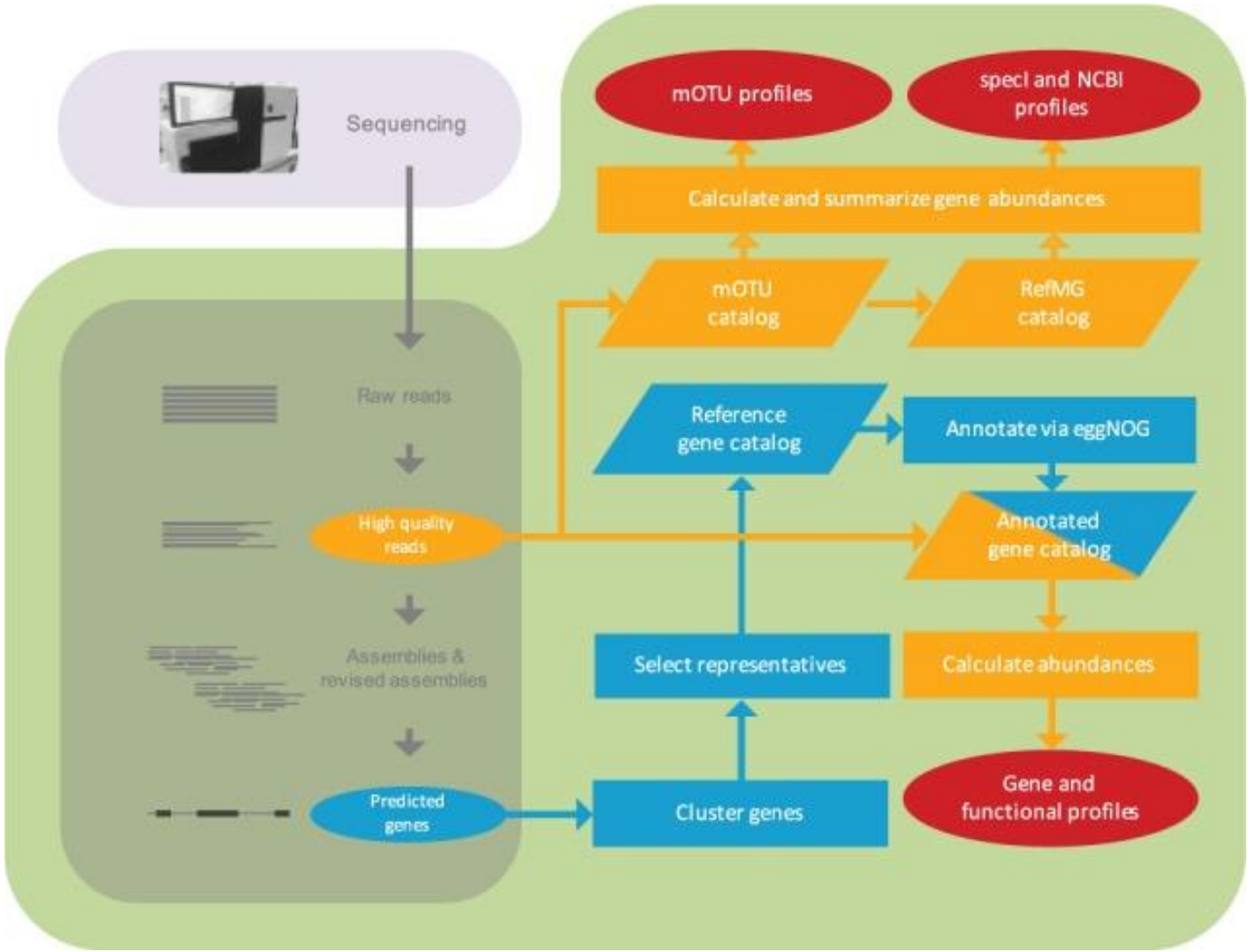
Give feedback

Publications:

Kultima *et al.* MOCAT2: a metagenomic assembly, annotation and profiling framework. *Bioinformatics* 16. 2016. PMID: 27153620

Kultima *et al.* MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS ONE* 10. 2012. PMID: 23082188

Metagenomics



	Proteins	Coverage	Precision	Recall	Reference
Protein domains and families					
eggNOG	7 449 593	100	100	100	Huerta-Cepas et al. (2015)
Pfam	16 230*	87	90	94	Finn et al. (2014)
Superfamily	15 438*	93	89	94	Gough et al. (2001)
(Metabolic) pathways					
KEGG	7 423 864	98	93	93	Kanehisa et al. (2014)
MetaCyc	388 782	100	89	94	Caspi et al. (2014)
SEED	4 247 700	99	94	94	Overbeek et al. (2014)
Complex carbohydrate metabolism					
dbCAN	333*	76	99	99	Yin et al. (2012)
Virulence factors					
MvirDB	29 357	100	95	93	Zhou et al. (2007)
PATRIC	2 194 475	93	93	93	Mao et al. (2015)
vFam	29 655	35	99	86	Skewes-Cox et al. (2014)
VFDB	1 627 380	86	89	91	Chen et al. (2012)
Victors	3 329 893	91	92	94	Mao et al. (2015)
Mobile genetic elements					
ICEberg	13 984	98	79	91	Bi et al. (2012)
Antibiotic resistance					
ARDB	25 360	89	99	88	Liu and Pop (2009)
CARD	2 820	100	81	93	McArthur et al. (2013)
Resfams	123*	80	94	94	Gibson et al. (2014)
Bacterial drug targets and exotoxins					
DBETH	228	100	99	86	Chakraborty et al. (2012)
DrugBank	3 899	99	88	94	Knox et al. (2011)

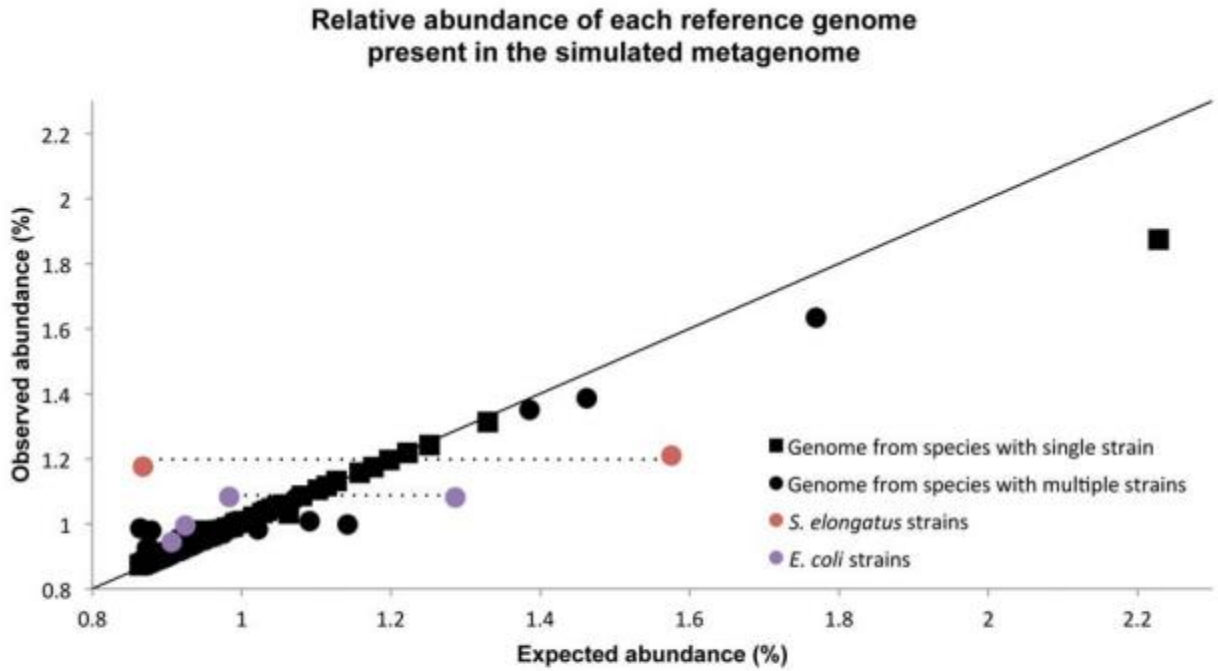


Figure 2. Relative abundance of each reference genome present in the simulated metagenome. The observed abundances by mapping reads to reference genomes and the expected abundance correlate with a Pearson correlation coefficient of 0.95 (base and read counts). Circles represent genomes with multiple strains from one species and squares represent genomes with only one strain within the species. All, but one, of the observations deviating from the diagonal are strains from the same species. These strains are either over- or under represented because reads are mapped to other closely related strains in addition to the strain of origin. Highlighted by dashed lines, are two examples where a high sequence similarity between strains (99.9% and 98.7% for the *Synechococcus elongatus* and *Escherichia coli* strains, respectively) can result in deviations from expected abundances.

doi:10.1371/journal.pone.0047656.g002

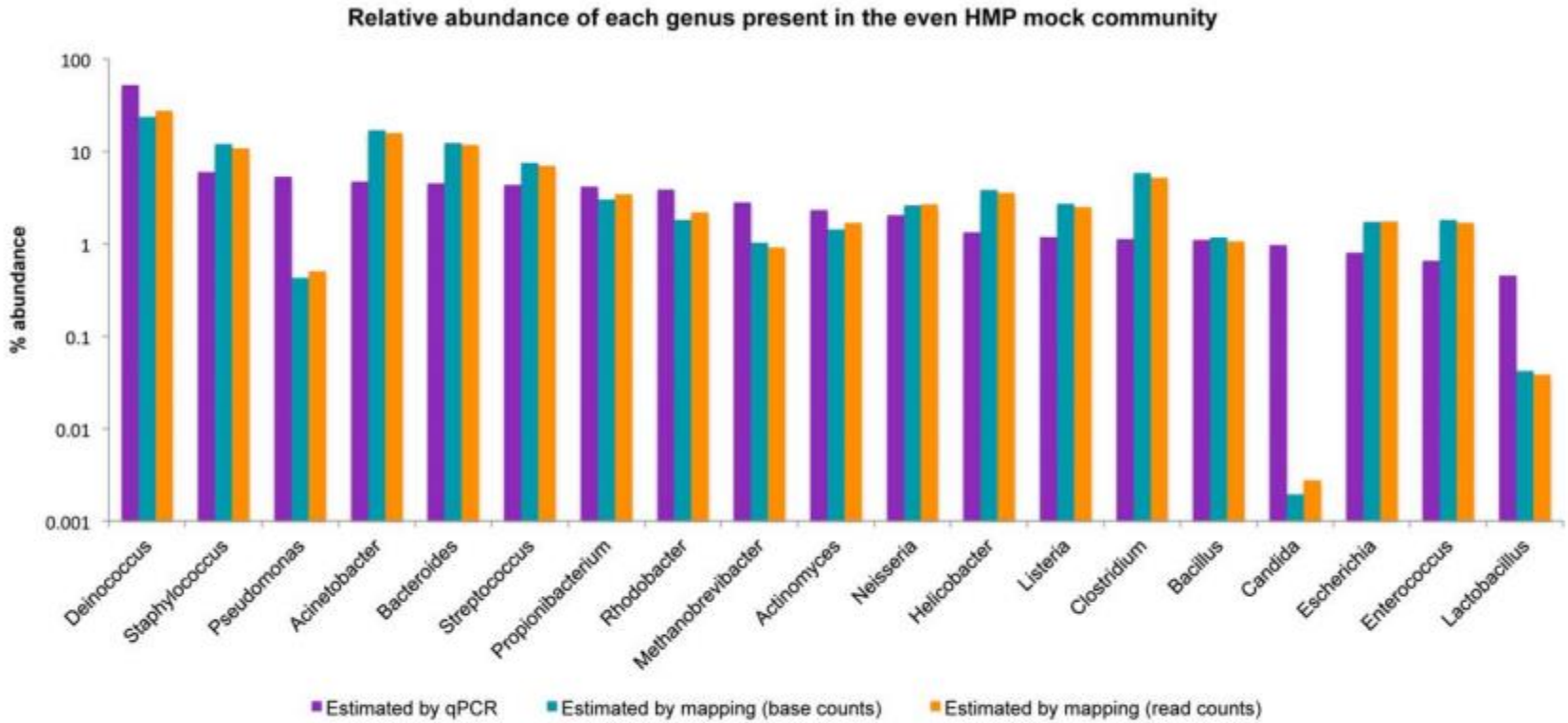


Figure 3. Relative abundance of each genus present in the even HMP mock community. The estimated abundances using qPCR and by mapping reads to reference genomes correlate with a Pearson correlation coefficient of 0.75 (base counts) and 0.83 (read counts).
doi:10.1371/journal.pone.0047656.g003

Specl

Molecular deliniation of prokaryotic species

Specl is a species identification tool using genomic sequences to delineate prokaryotic species. It facilitates fast, accurate and automated taxonomic assignments of newly sequenced genomes based on comparisons of 40 universal, single-copy phylogenetic marker genes extracted from a comprehensive database of sequenced prokaryotic genomes.

Homepage **Source code** **Documentation**

Contact: zeller@embl.de

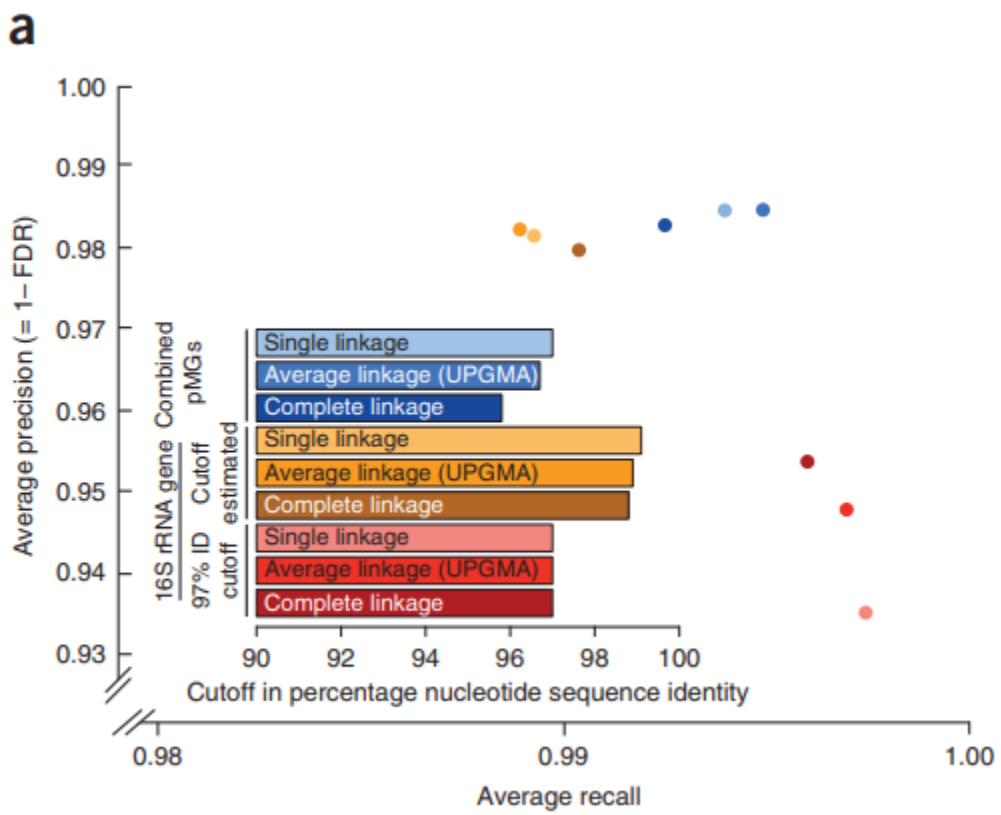
Versio

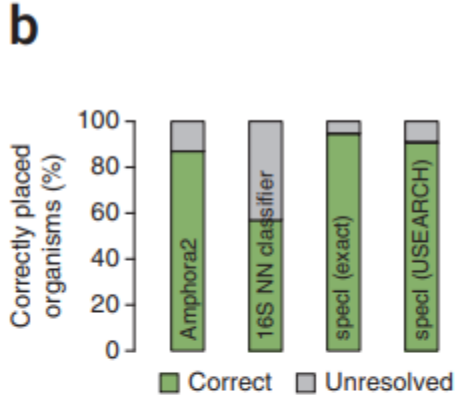
Give fe

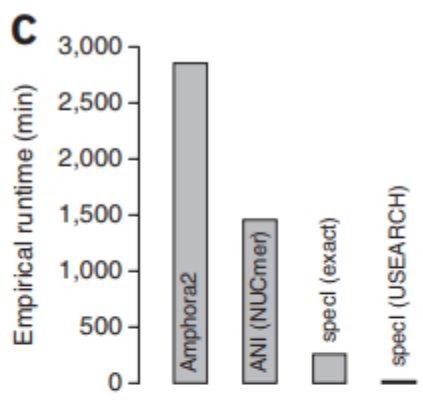
Publica

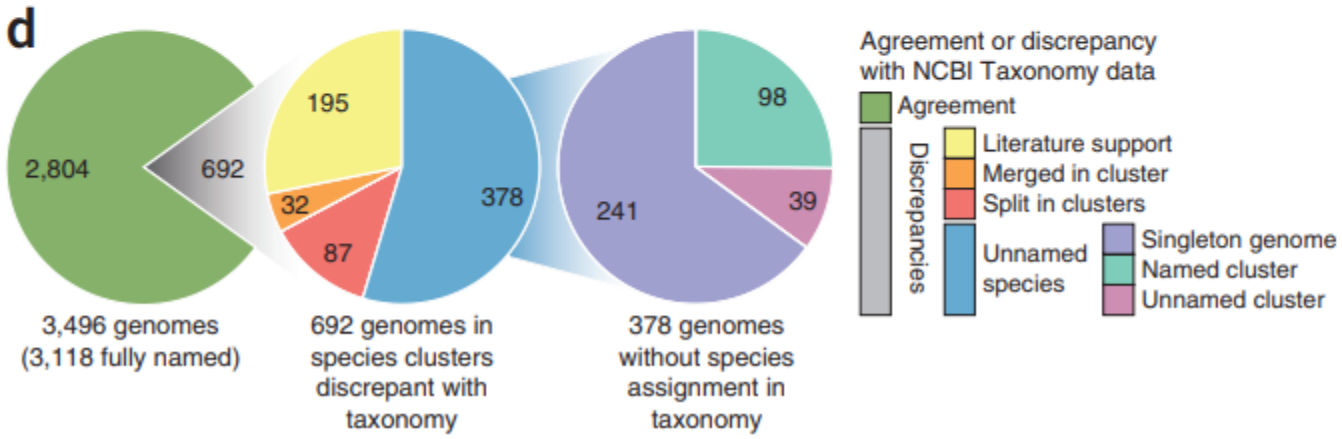
Mende
PMID: 2

Meta









mOTUs

Metagenomic species profiling

Metagenomic operational taxonomic units (mOTUs) allow for the taxonomic profiling of known (sequenced) and unknown microorganisms at species-level resolution from shotgun sequencing data. The method clusters single-copy phylogenetic marker gene sequences from metagenomes and reference genomes into mOTUs to quantify their abundances in shotgun metagenomic samples.

Homepage | Source code | Binaries | Documentation

Contact: zeller@embl.de
Version: 1.0

Give feedback

Publications:

Sunagawa *et al.* Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 12. 2013. PMID: 24141494

Metagenomics

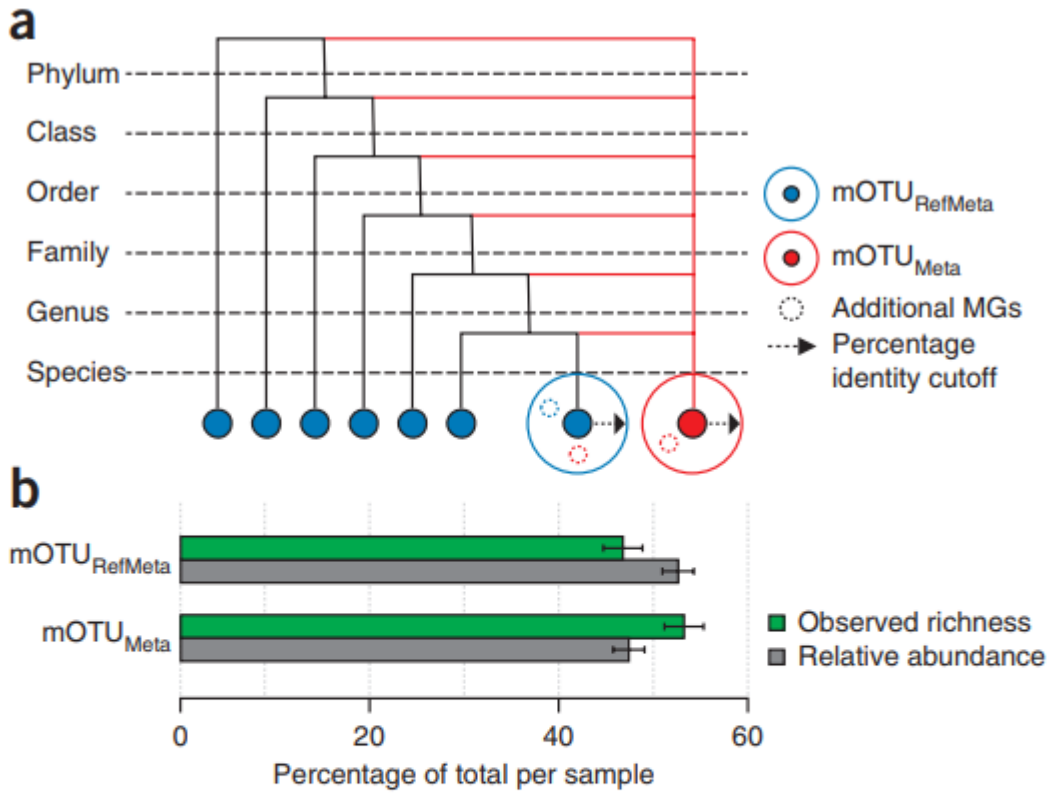


Figure 1 | Phylogenetic marker gene–based mOTUs. (**a**) Schematic showing the mOTUs that contained at least one marker gene (MG) that originated from a sequenced reference genome and at least one metagenomic MG ($mOTU_{RefMeta}$) and mOTUs that contained at least one metagenomic MG but no reference MG ($mOTU_{Meta}$). Black and red lines indicate known and unknown topologies, respectively. (**b**) Mean fractions of $mOTU_{Meta}$ and $mOTU_{RefMeta}$ of the observed mOTU richness per sample, and mean relative abundances based on mOTU abundance profiles of 252 human fecal samples.

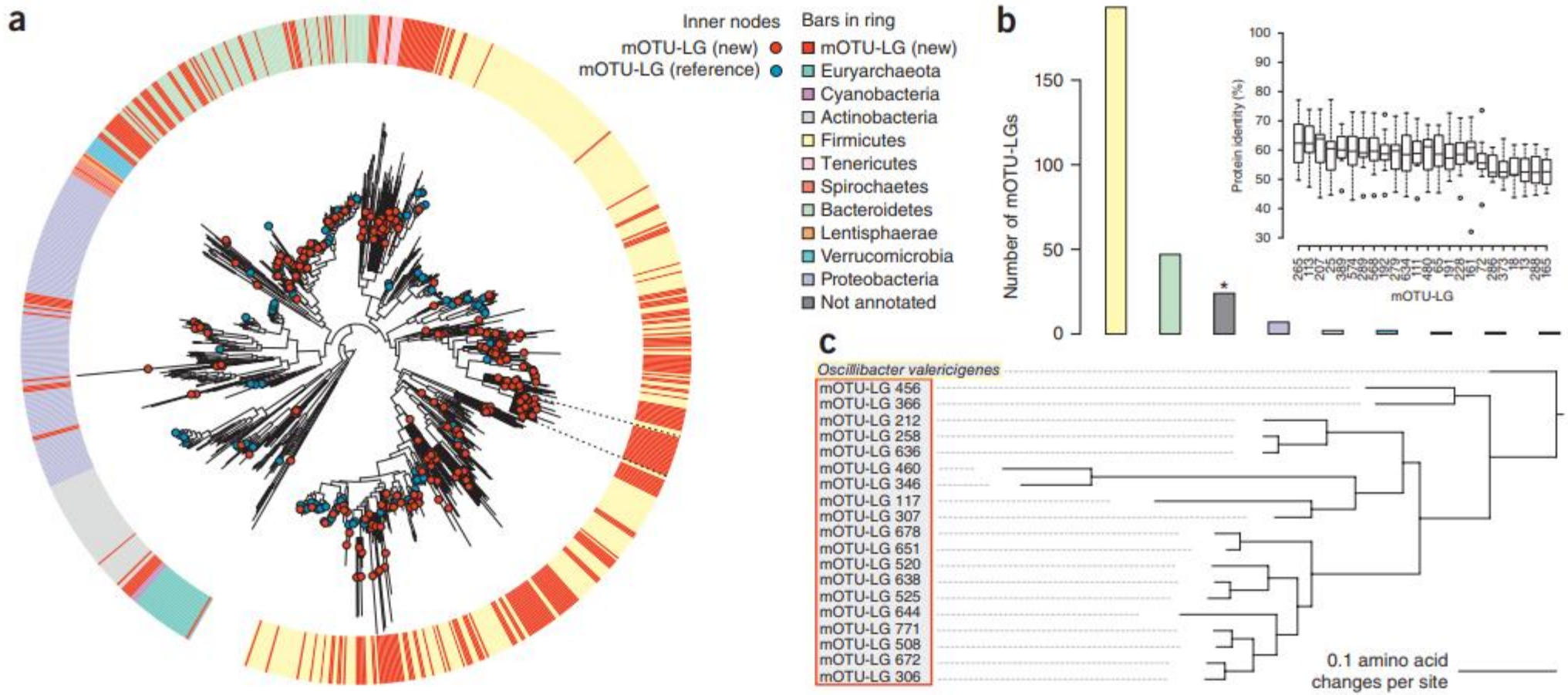


Figure 2 | Phylogenetic analysis of mOTU linkage groups. (**a**) Maximum likelihood phylogenetic tree of prokaryotic species used to infer the topology of mOTU-LGs (Online Methods). US National Center for Biotechnology Information phylum-level taxonomy is color-coded on the outer ring, and placements of mOTU-LGs are shown as circles on tree edges. Dashed lines indicate a clade of *Oscillibacter valericigenes* and related mOTU-LGs that are highlighted in **c**. (**b**) Phylum-level

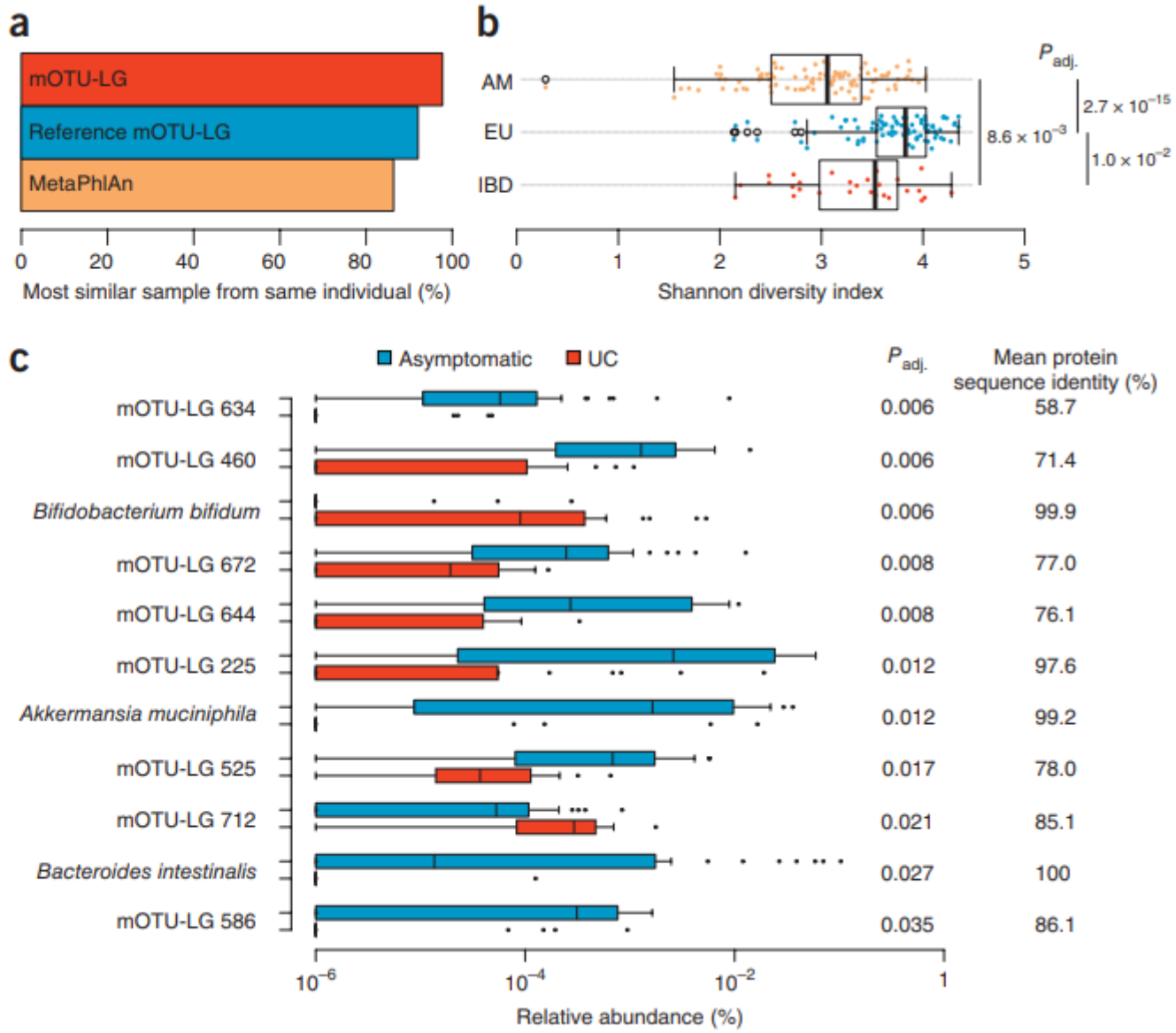


Figure 3 | Performance and application of mOTU linkage groups. (**a**) Fraction of samples originating from 43 individuals that were sampled at least twice (total 88 samples) for which the most similar sample originated from the same individual using mOTU-LG (red), a subset of mOTU-LGs that represent reference species (reference mOTU-LG) and clade-specific genes[7] at species level (MetaPhlAn). (**b**) Shannon diversity index for samples originating from US individuals (AM; $n = 97$), asymptomatic European individuals (EU; $n = 85$) and individuals diagnosed with IBD (IBD; $n = 25$). Individual samples are shown as closed circles and collective data for each group superimposed as box plots. $P_{adj.}$ denotes Bonferroni-adjusted P values of Wilcoxon's rank-sum test results. (**c**) Relative abundances of mOTU-LGs that were significantly different between fecal samples from a cohort of UC patients ($n = 21$) and matched asymptomatic individuals ($n = 35$). The mean (across mOTU-LG members) protein identity for best BLASTp hits is shown as a proxy for phylogenetic distance to the closest organism for which a reference genome sequence was available. $P_{adj.}$ values denote FDR-adjusted P values of Wilcoxon test results.

information per species, but also allow for

eggNOG

Resource for orthologous genes

eggNOG is a database of nested orthologous gene groups (NOGs) infered using unsupervised clustering applied to >2,000 complete genomes followed by comprehensive characterization and analysis of the resulting gene families. eggNOG provides orthologous group assignments at >100 different taxonomic levels as well as multiple sequence alignments, maximum-likelihood trees and broad functional annotations for each group accessible via a web interface or through bulk download.

Homepage | Biological data | Documentation

Contact: zeller@embl.de
Version:

Give feedback

Publications:

Huerta-Cepas *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* D1. 2016. PMID: 26582926

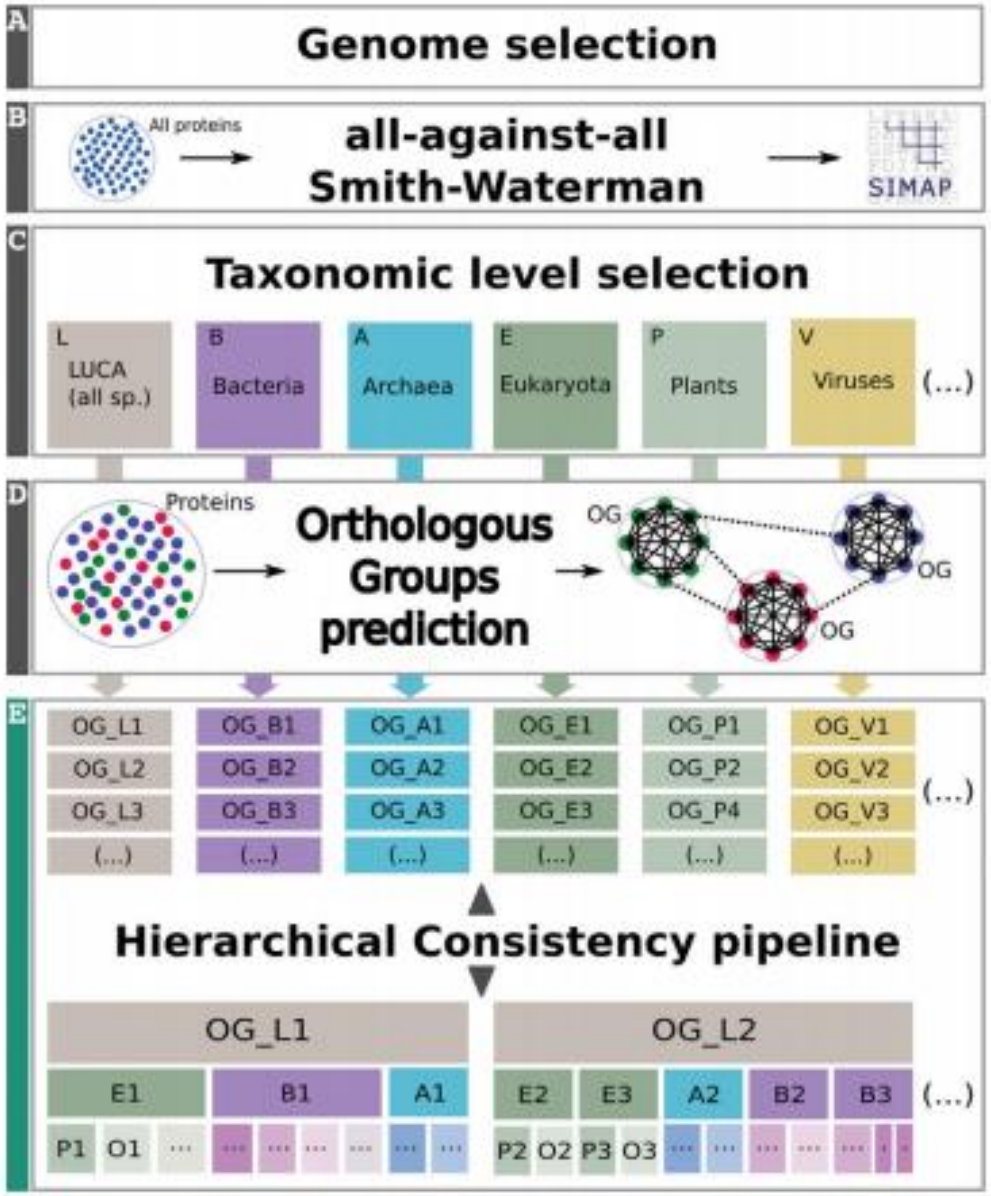
Powell *et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.* Database issue. 2014. PMID: 24297252

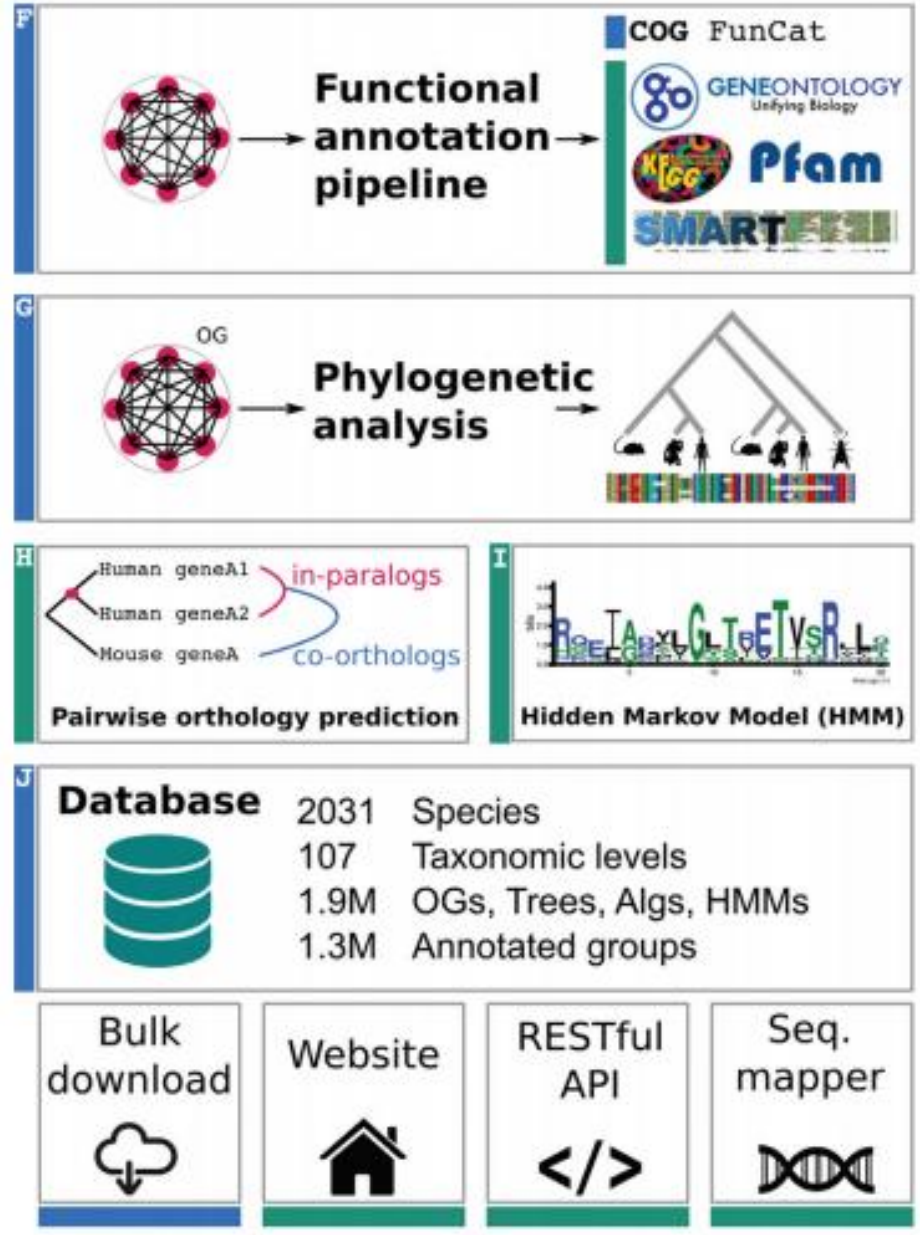
Powell *et al.* eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* Database issue. 2012. PMID: 22096231

Muller *et al.* eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* Database issue. 2010. PMID: 19900971

Jensen *et al.* eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* Database issue. 2008. PMID: 17942413

Metagenomics Genomics


A
Genome selection
B
All proteins
all-against-all
Smith-Waterman
SIMAP
C
Taxonomic level selection
L
LUCA
(all sp.)
B
Bacteria
A
Archaea
E
Eukaryota
P
Plants
V
Viruses
(...)
D
Proteins
Orthologous
Groups
prediction
OG
OG
OG
E
OG_L1
OG_L2
OG_L3
(...)
OG_B1
OG_B2
OG_B3
(...)
OG_A1
OG_A2
OG_A3
(...)
OG_E1
OG_E2
OG_E3
(...)
OG_P1
OG_P2
OG_P4
(...)
OG_V1
OG_V2
OG_V3
(...)
(...)
Hierarchical Consistency pipeline
OG_L1
OG_L2
E1
B1
A1
E2
E3
A2
B2
B3
(...)
P1
O1
P2
O2
P3
O3


F
COG FunCat
Functional
annotation
pipeline
GENEONTOLOGY
Unifying Biology
KEGG
Pfam
SMART
G
OG
Phylogenetic
analysis
H
Human geneA1
Human geneA2
Mouse geneA
in-paralogs
co-orthologs
Pairwise orthology prediction
I
Hidden Markov Model (HMM)
J
Database
2031 Species
107 Taxonomic levels
1.9M OGs, Trees, Algs, HMMs
1.3M Annotated groups
Bulk
download
Website
RESTful
API
</>
Seq.
mapper

iPATH

Cellular pathway mapping tool

iPath is a web-based tool for the visualization and analysis of cellular pathways. Based on functional annotations (such as KEGG), it provides pathway maps for primary cellular metabolism as well as for some additional secondary metabolite synthesis and regulatory pathways. Users can map their own data onto these pathway maps. Due to its navigation and customization functions, iPATH thus allows users to easily explore and analyze the functional and metabolic capabilities of their (meta-)genomic data sets.

Homepage Documentation

Contact: zeller@embl.de
Version: 2.0

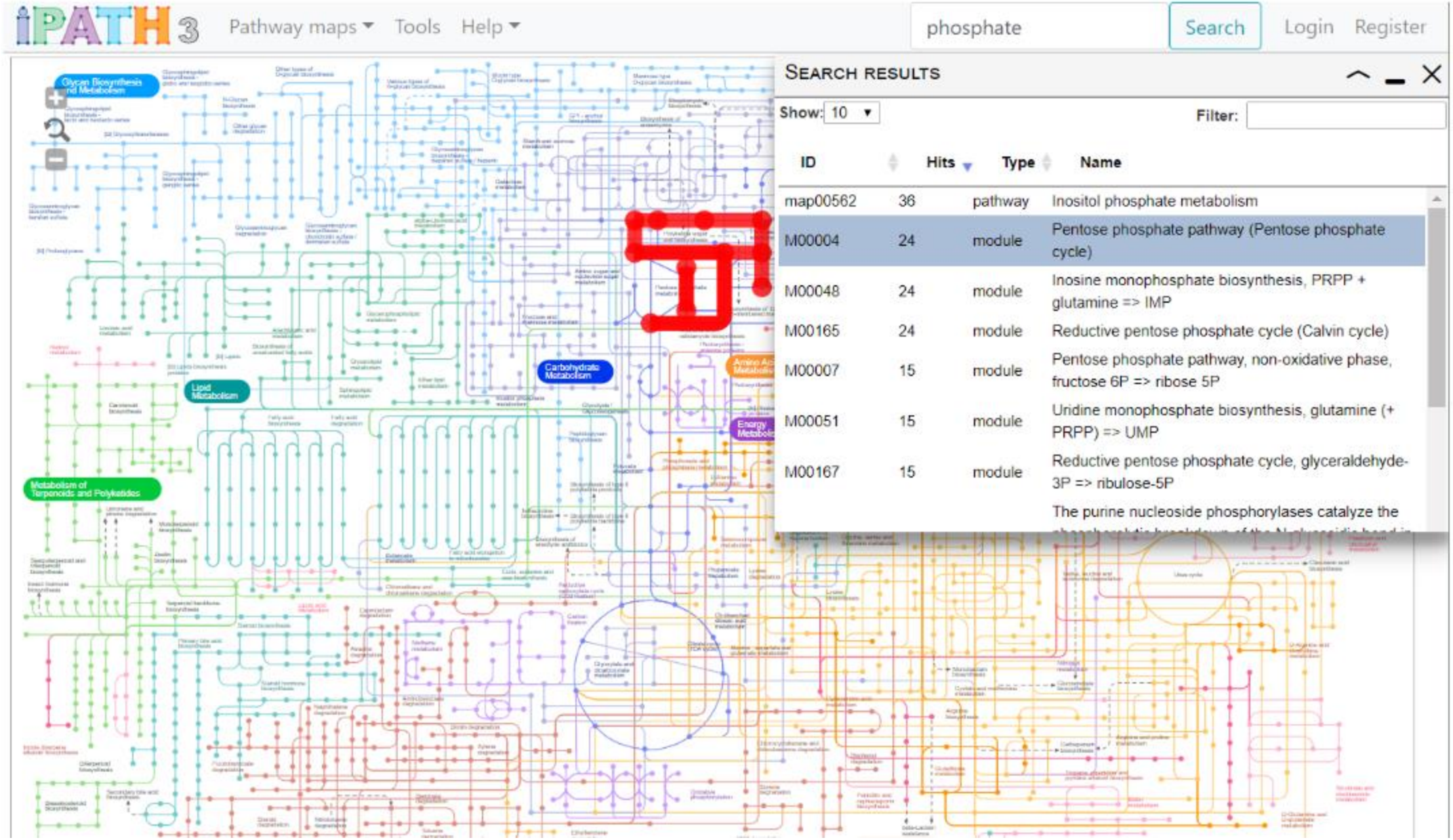
Give feedback

Publications:

Yamada *et al.* iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* Web Server issue. 2011. PMID: 21546551

Letunic *et al.* iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem. Sci.* 3. 2008. PMID: 18276143

Metagenomics Genomics


iPATH 3
Pathway maps
Tools
Help
phosphate
Search
Login
Register
Glycan Biosynthesis and Metabolism
Lipid Metabolism
Carbohydrate Metabolism
Energy Metabolism
Metabolism of Terpenoids and Polyketides
SEARCH RESULTS
Show: 10
Filter:
ID
Hits
Type
Name
map00562 36 pathway Inositol phosphate metabolism
M00004 24 module Pentose phosphate pathway (Pentose phosphate cycle)
M00048 24 module Inosine monophosphate biosynthesis, PRPP + glutamine => IMP
M00165 24 module Reductive pentose phosphate cycle (Calvin cycle)
M00007 15 module Pentose phosphate pathway, non-oxidative phase, fructose 6P => ribose 5P
M00051 15 module Uridine monophosphate biosynthesis, glutamine (+ PRPP) => UMP
M00167 15 module Reductive pentose phosphate cycle, glyceraldehyde-3P => ribulose-5P
The purine nucleoside phosphorylases catalyze the

Interactive Tree Of Life (iTOL)

Interactive tree visualization

Interactive Tree Of Life (iTOL) is an online tool for the display and manipulation of phylogenetic trees. It provides a large variety of tree layouts, drawing and annotation features including circular tree layout, which is well-suited particularly for mid-sized trees (up to several thousand leaves). Tree displays can be exported in several graphical formats, both bitmap and vector based.

Homepage **Documentation**

Contact: zeller@embl.de
Version: 3.0

Give feedback

Publications:

Letunic & Bork Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* W1. 2016. PMID: 27095192

Letunic & Bork Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* Web Server issue. 2011. PMID: 21470960

Letunic & Bork Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 1. 2007. PMID: 17050570

Phylogenetics Metagenomics Genomics

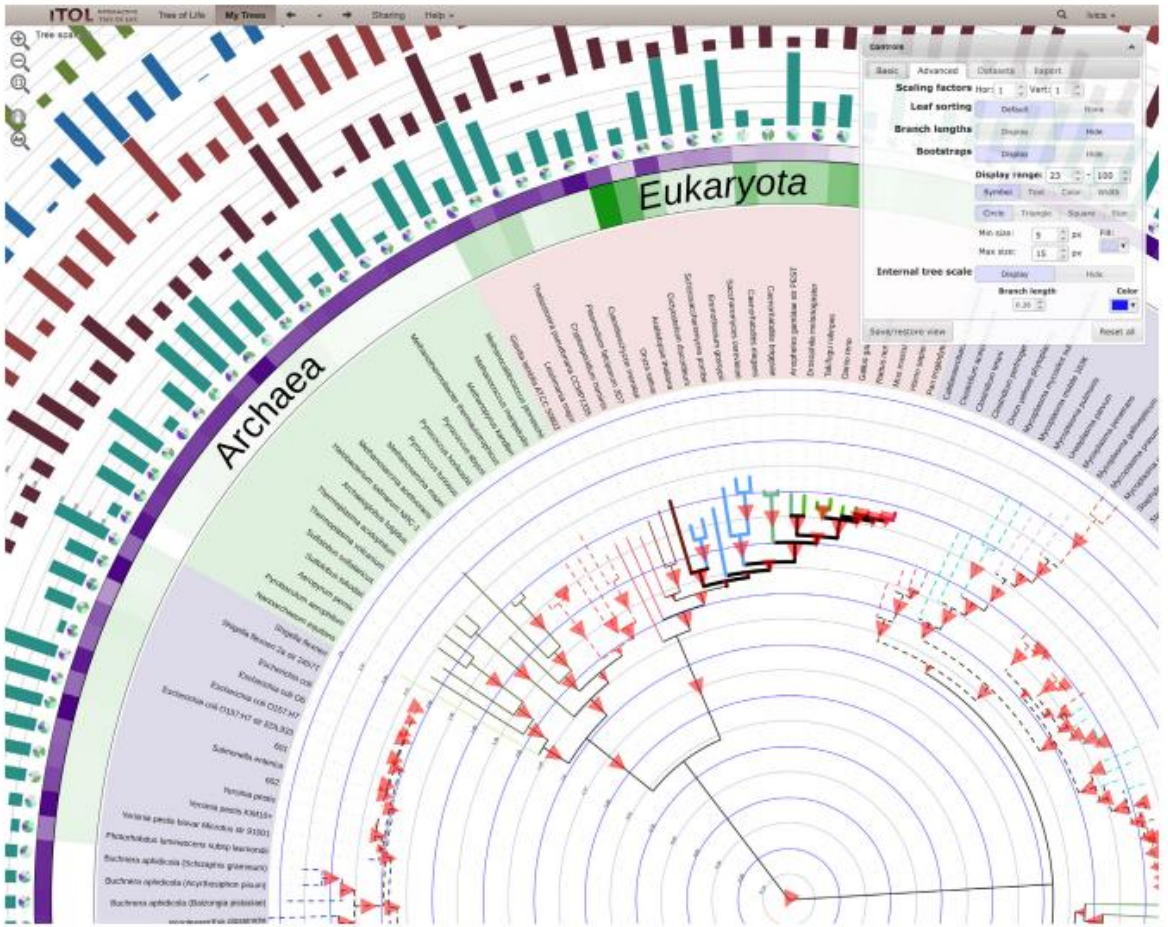


Figure 1. Tree of Life (15) annotated with various datasets, highlighting several novel features, along with iTOL's interactive user interface. Tree branches are colored and displayed with varying styles and widths. Support values are visualized as triangles of varying size. An internal tree scale is displayed, with user-defined colors and level values. A text label dataset was precisely positioned above the colored strips using interactive controls. Several dataset types can include user-defined scales.

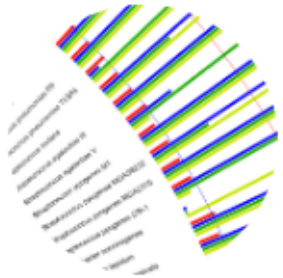
https://itol.embl.de

iTOL Interactive Tree Of Life | Tree of Life | Upload | Sharing data | Help | Login

Manage

Organize your trees into workspaces and projects, and access them from any browser. Simply drag and drop multiple tree files onto a project to upload them all at once.

Create an account »



Annotate

18 dataset types. Full control over branch colors, widths and styles. Individually adjustable label fonts, sizes and styles.

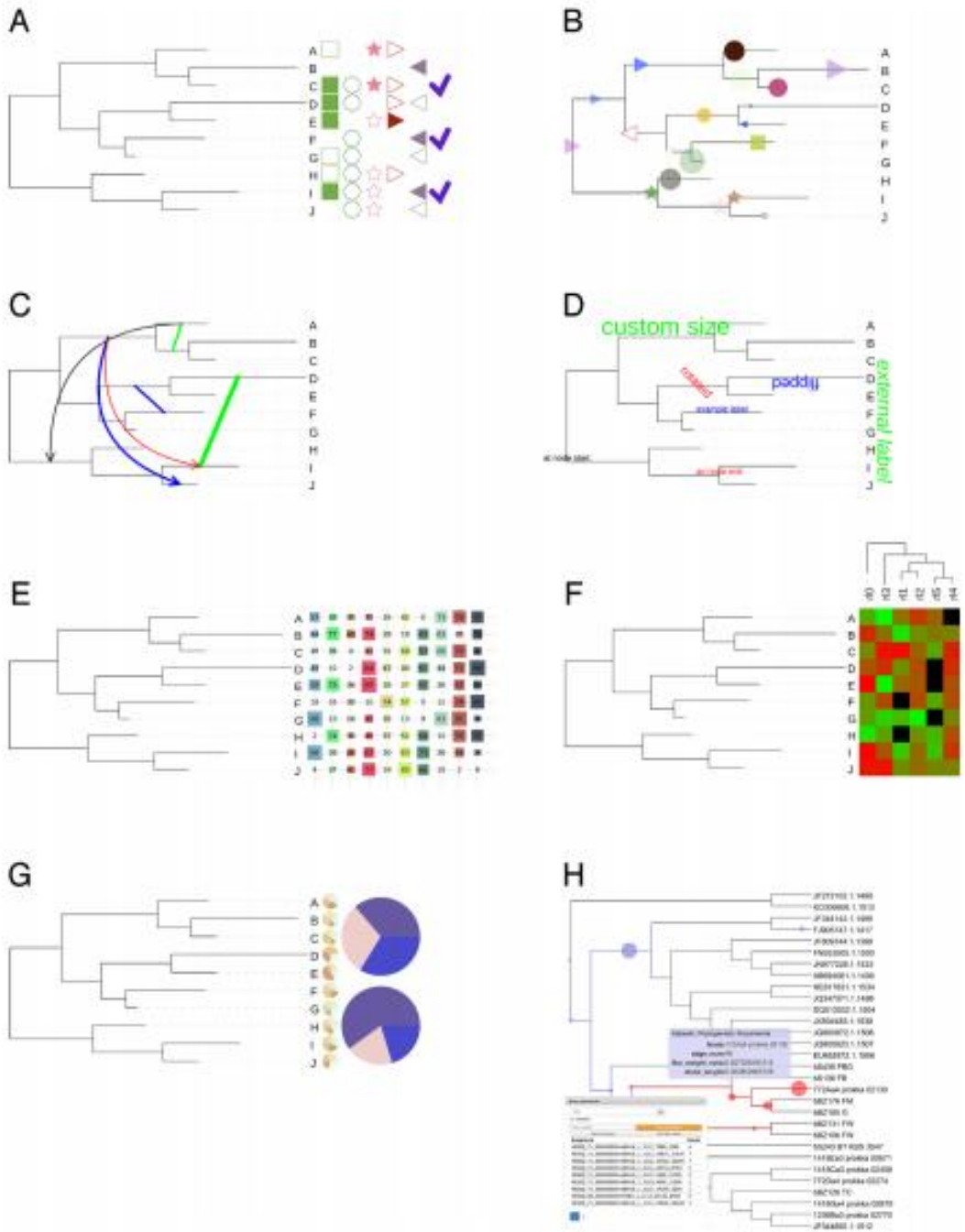
Gallery of user created trees

Upload a tree »

Export

Create high quality vector or bitmap figures for your publications. Direct WYSIWYG export of what is displayed on the screen.

Explore help »


A
B
C
D
E
F
G
H
custom size
paddij
external label

Enterotyping

Gut microbial community typing

Enterotypes are densely populated regions in a high-dimensional space of microbiome community composition, by which human individuals can be stratified (Arumugam, Raes et al. Nature 2011). Computational methods to detect and characterise enterotypes in any dataset, either to reproduce previous reports or determine enterotypes in new studies, are provided and explained.

Homepage **Documentation**

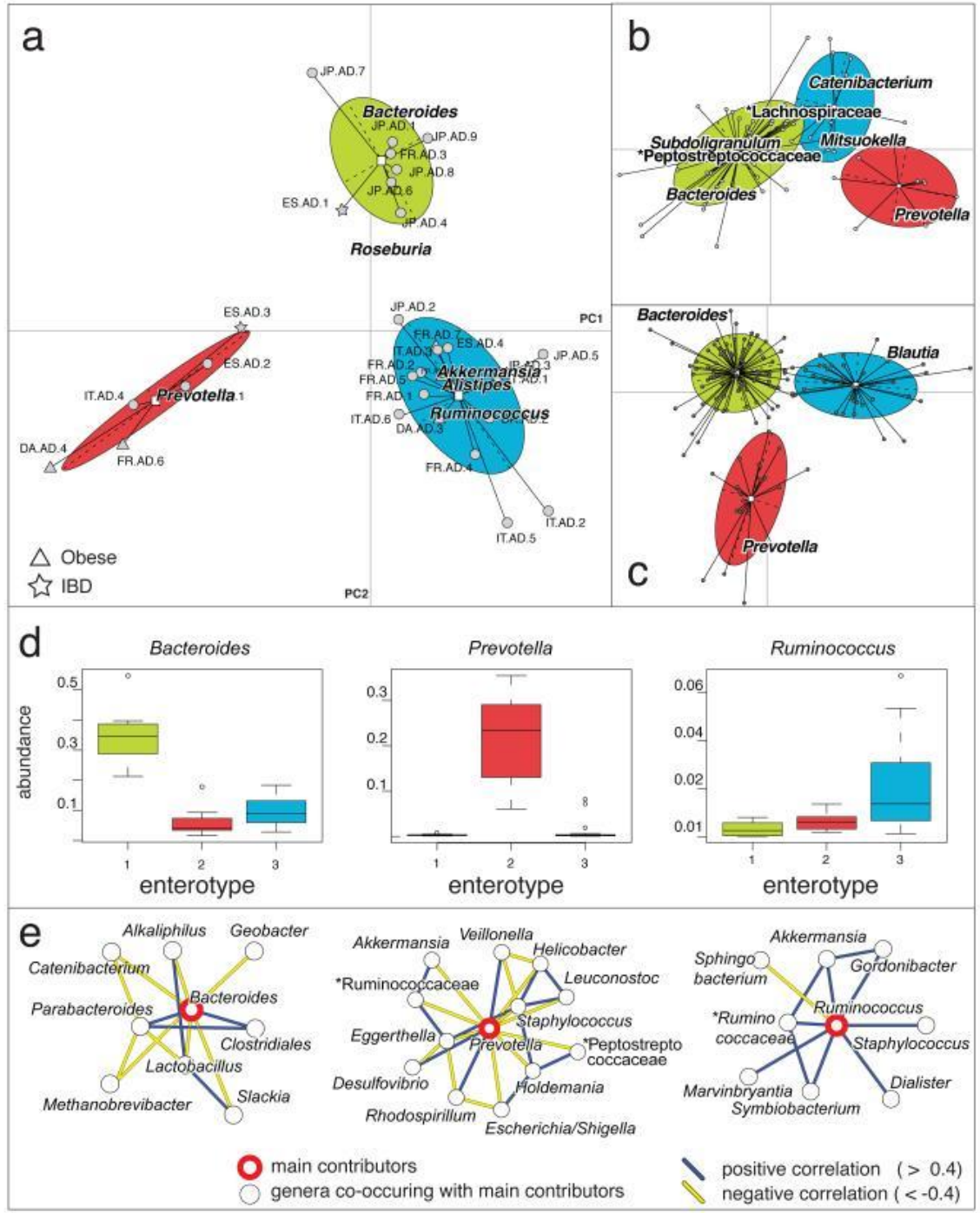
Contact: zeller@embl.de
Version: 1.0

Give feedback

Publications:

Arumugam *et al.* Enterotypes of the human gut microbiome. *Nature* 7346. 2011. PMID: 21508958

Metagenomics


a
JP.AD.7
Bacteroides
JP.AD.1
JP.AD.9
FR.AD.3
JP.AD.8
JP.AD.6
ES.AD.1
JP.AD.4
Roseburia
ES.AD.3
ES.AD.2
IT.AD.4
Prevotella.1
DA.AD.4
FR.AD.6
JP.AD.2
PC1
FR.AD.7
ES.AD.4
IT.AD.3
JP.AD.5
FR.AD.2
Akkermansia
Alistipes
FR.AD.5
FR.AD.1
IT.AD.6
Ruminococcus
DA.AD.3
FR.AD.4
IT.AD.2
IT.AD.5
Obese
IBD
PC2
b
Catenibacterium
*Lachnospiraceae
Subdoligranulum
Mitsuokella
*Peptostreptococcaceae
Bacteroides
Prevotella
c
Bacteroides*
Blautia
Prevotella
d
Bacteroides
Prevotella
Ruminococcus
abundance
enterotype
e
Alkaliphilus
Geobacter
Catenibacterium
Parabacteroides
Bacteroides
Clostridiales
Lactobacillus
Methanobrevibacter
Slackia
Akkermansia
Veillonella
Helicobacter
*Ruminococcaceae
Leuconostoc
Staphylococcus
Eggerthella
Prevotella
*Peptostrepto coccaceae
Desulfovibrio
Holdemania
Rhodospirillum
Escherichia/Shigella
Akkermansia
Sphingo bacterium
Gordonibacter
Ruminococcus
*Rumino coccaceae
Staphylococcus
Dialister
Marvinbryantia
Symbiobacterium
main contributors
genera co-occuring with main contributors
positive correlation (> 0.4)
negative correlation (< -0.4)

SIAMCAT

Statistical analysis of microbiome data

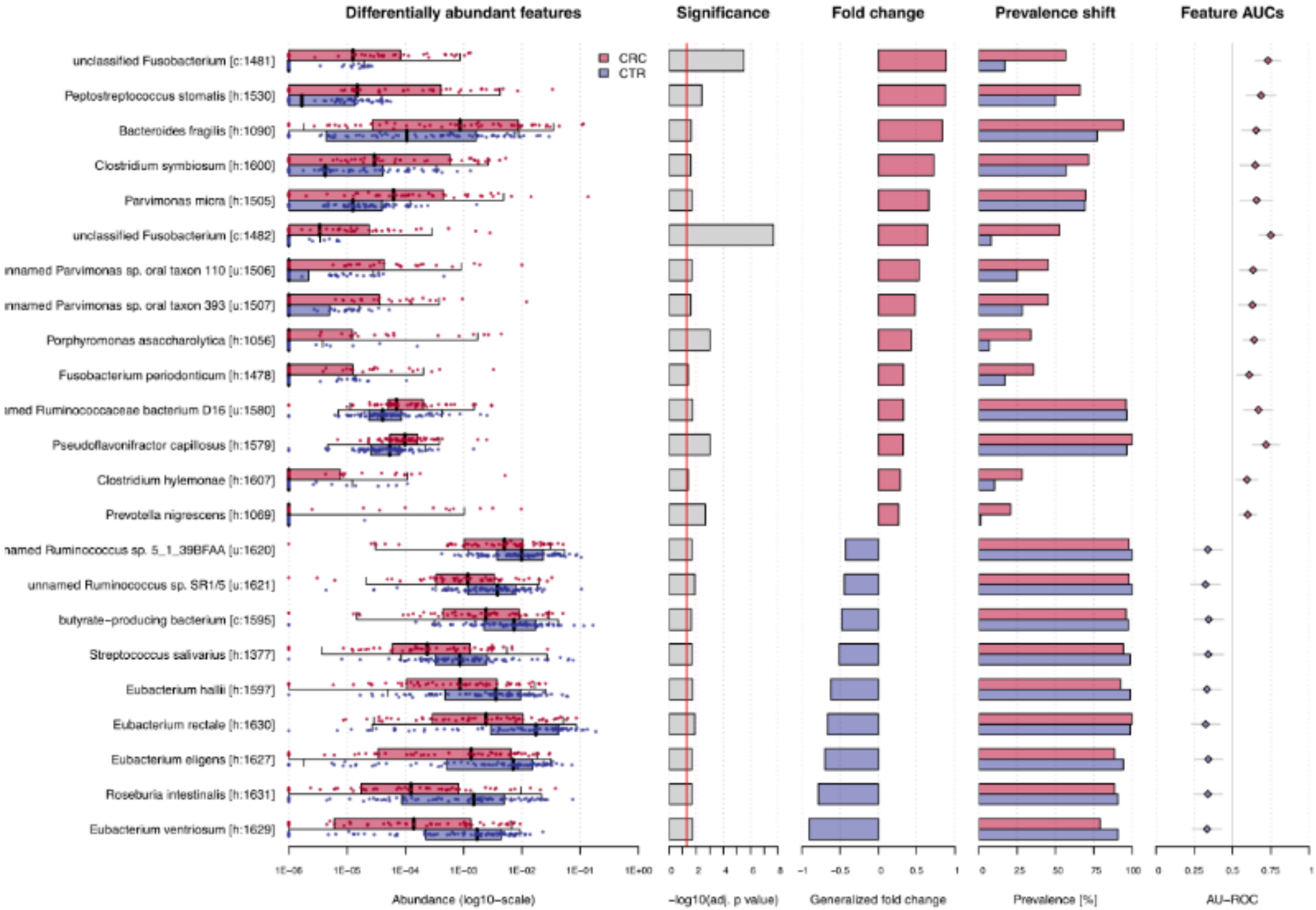
SIAMCAT is a modular framework for the statistical inference of associations between microbial communities and host phenotypes, such as disease states in clinical case-control studies. SIAMCAT is based on LASSO models, which offer distinctive advantages for model interpretation and microbial biomarker selection and avoid overfitting issues that can arise in naive combinations of feature selection and cross-validation.

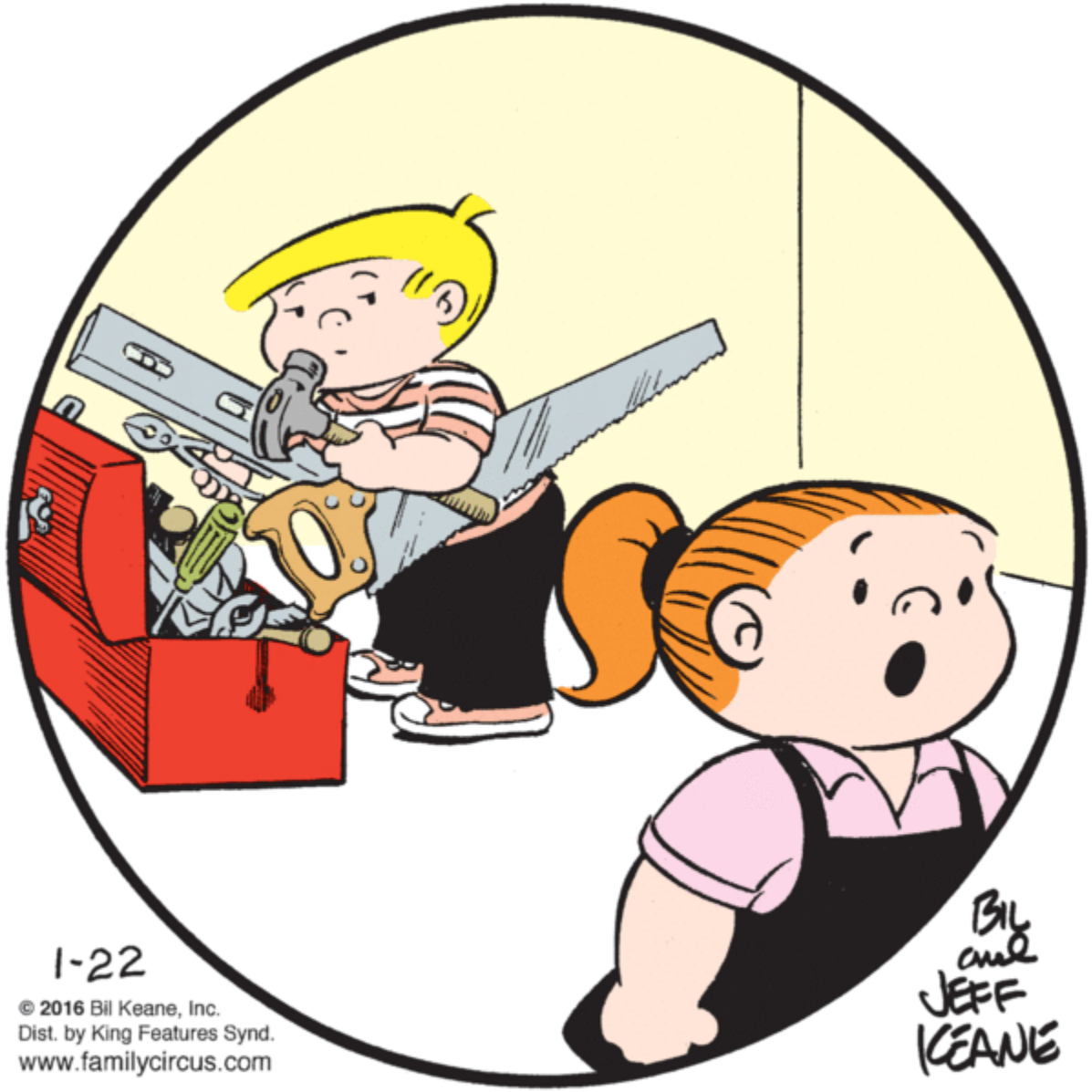
Homepage

Contact: zeller@embl.de
Version: 0.1

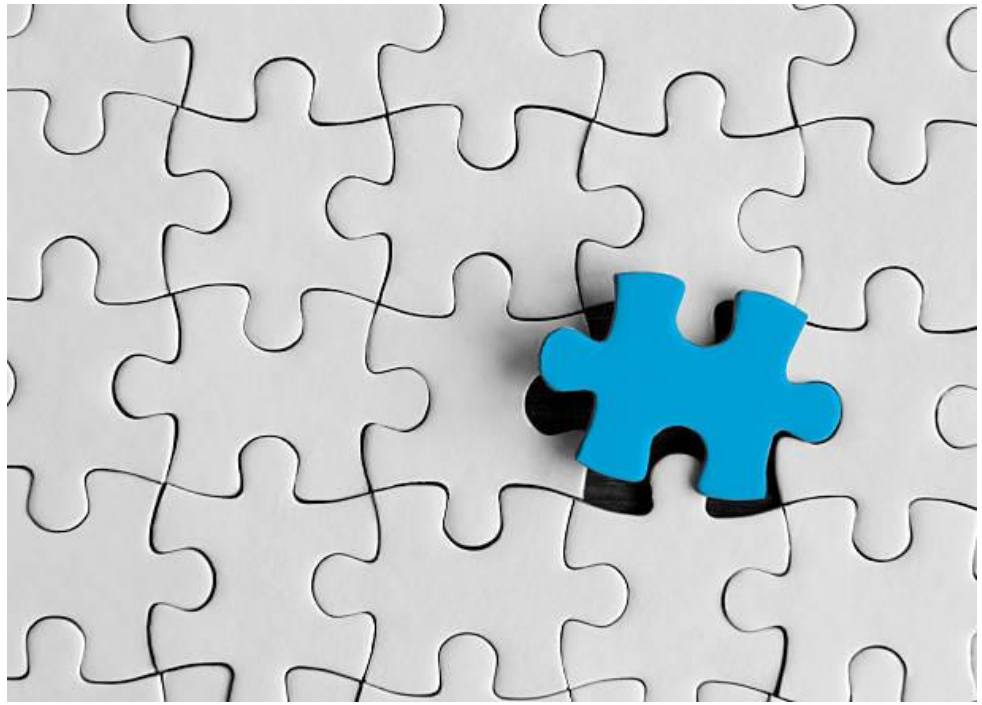
Give feedback

Metagenomics Genomics Statistics and probability


Differentially abundant features
Significance
Fold change
Prevalence shift
Feature AUCs
CRC
CTR
unclassified Fusobacterium [c:1481]
Peptostreptococcus stomatis [h:1530]
Bacteroides fragilis [h:1090]
Clostridium symbiosum [h:1600]
Parvimonas micra [h:1505]
unclassified Fusobacterium [c:1482]
nnamed Parvimonas sp. oral taxon 110 [u:1506]
nnamed Parvimonas sp. oral taxon 393 [u:1507]
Porphyromonas asaccharolytica [h:1056]
Fusobacterium periodonticum [h:1478]
med Ruminococcaceae bacterium D16 [u:1580]
Pseudoflavonifractor capillosus [h:1579]
Clostridium hylemonae [h:1607]
Prevotella nigrescens [h:1089]
named Ruminococcus sp. 5_1_39BFAA [u:1620]
unnamed Ruminococcus sp. SR1/5 [u:1621]
butyrate-producing bacterium [c:1595]
Streptococcus salivarius [h:1377]
Eubacterium hallii [h:1597]
Eubacterium rectale [h:1630]
Eubacterium eligens [h:1627]
Roseburia intestinalis [h:1631]
Eubacterium ventriosum [h:1629]
1E−06
1E−05
1E−04
1E−03
1E−02
1E−01
1E+00
0
2
4
6
8
−1
−0.5
0
0.5
1
0
25
50
75
100
0
0.25
0.5
0.75
1
Abundance (log10−scale)
−log10(adj. p value)
Generalized fold change
Prevalence [%]
AU−ROC



"Mommy, I think you need to keep YOUR eyes on Billy now.



Example - Process, Assemble, Revise Assembly, Predict Genes, cluster genes into gene catalog, annotate gene catalog, profile against gene catalog

```
MOCAT.pl -sf my.samples -rtf
MOCAT.pl -sf my.samples -a
MOCAT.pl -sf my.samples -gp assembly
MOCAT.pl -sf my.samples -make_gene_catalog -assembly_type assembly
MOCAT.pl -sf my.samples -annotate_gene_catalog
MOCAT.pl -sf my.samples -s my.samples.padded -identity 95
MOCAT.pl -sf my.samples -f my.samples.padded -identity 95
MOCAT.pl -sf my.samples -p my.samples.padded -identity 95 -mode functional
```

Example - Process, Assemble, Predict Genes and fetc

```
MOCAT.pl -sf my.samples -rtf
MOCAT.pl -sf my.samples -a
MOCAT.pl -sf my.samples -gp assembly
MOCAT.pl -sf my.samples -fmg assembly
MOCAT.pl -sf my.samples -ss
```

Example - Process, Screen against DB, Assemble, and

```
MOCAT.pl -sf my.samples -rtf
MOCAT.pl -sf my.samples -s hg19 -screened_files
MOCAT.pl -sf my.samples -a -r hg19
MOCAT.pl -sf my.samples -gp assembly -r hg19
MOCAT.pl -sf my.samples -ss
```

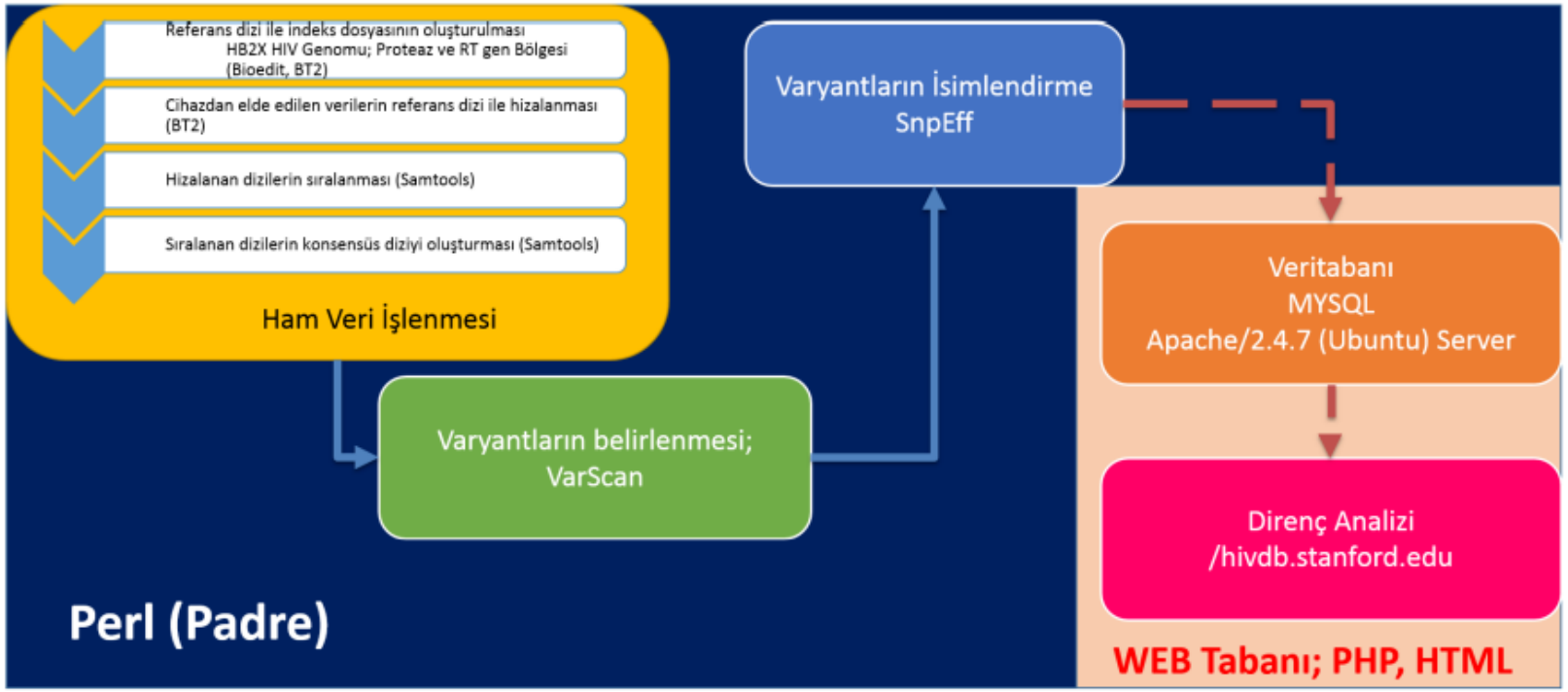
```
> runMOCAT.sh

########################################################################
# WELCOME TO THE MOCAT EXECUTER v1.3 #
########################################################################

This shell script is used to execute a number of MOCAT commands in a row.
Typically this is used to process raw reads up to final taxonomic or mOTU
profiles. Of course you can process each step individually using MOCAT.pl
but we have created this software for your ease to execute these commands
with ease without prior knowledge of how to run MOCAT. Enjoy!

Usage: runMOCAT.sh [-sf SAMPLE_FILE -cfg CONFIG_FILE]

SAMPLE_FILE not specified with option -sf SAMPLE_FILE
Looking for valid sample files in the current folder:
Getting files...
Getting folders...
Processing files..............

SELECT A SAMPLE FILE:
- sample

ENTER SAMPLE FILE:

--- Type 'sample' and press enter ---
```

4. ULUSAL KLİNİK MİKROBİYOLOJİ KONGRESİ
08 -12 KASIM 2017
RIXOS SUNGATE KONGRE MERKEZİ
ANTALYA

SS08

Antiretroviral Direncin Saptanmasına Yönelik Yeni Nesil Dizi Analizi Verilerinin Değerlendirilmesi İçin İlk Türkçe, Açık Kaynak Kodlu Veri Analiz Yazılımının *(DeepM)* Geliştirilmesi

Kuşkucu
Cerrahpa
Tıbbi Mikrobi

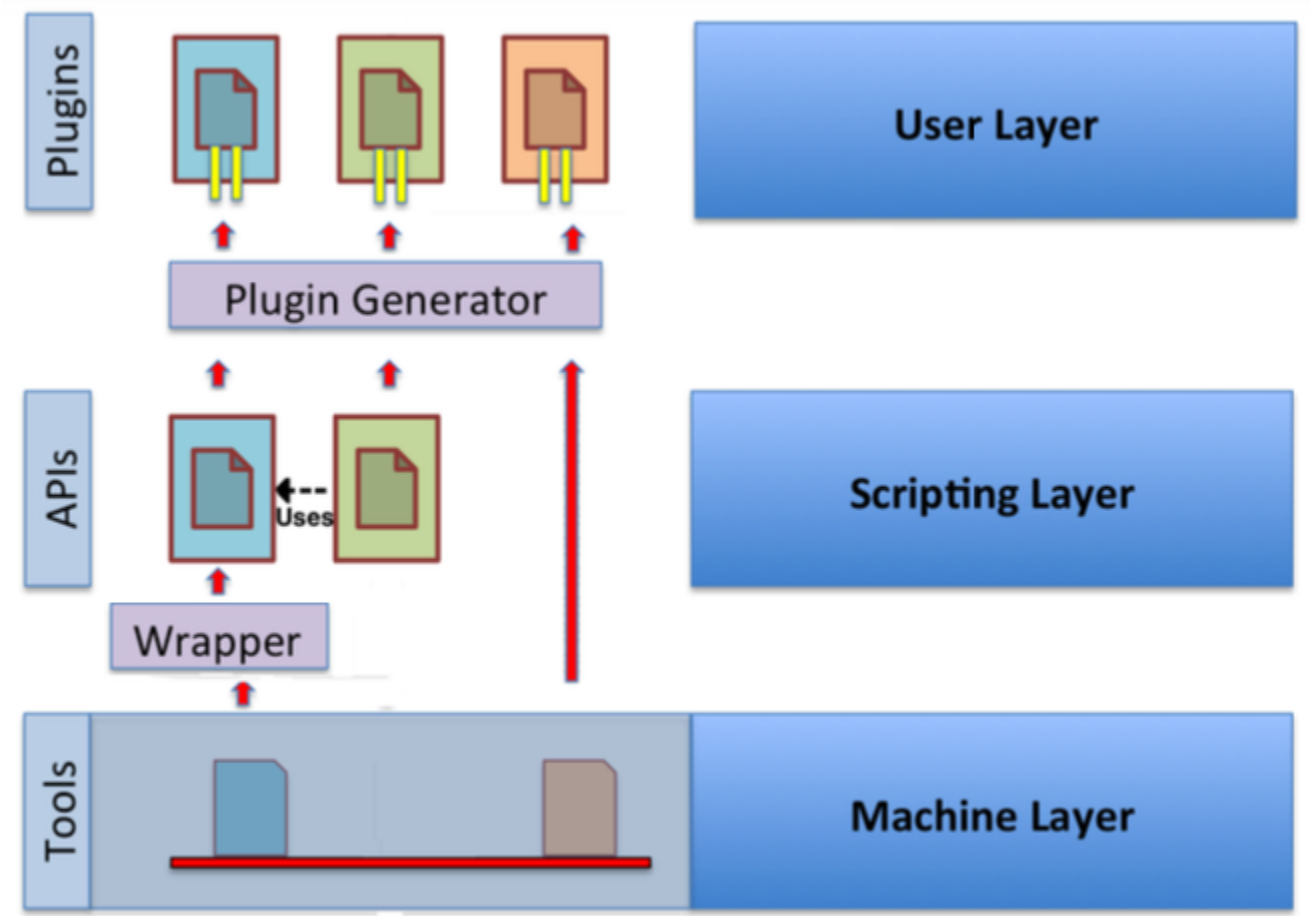


Test your microbiome analysis pipeline with PLuMA

Helene Perrin Follow
May 23, 2018 · 3 min read

Plugin-Based Microbiome Analysis

If you are an algorithm developer who wants to prototype, test and debug a new pipeline stage in your programming language of choice and run alongside existing stages with no overhead for interfacing, the Bioinformatics Research Group (BioRG) at Florida International University would like to introduce you to Plugin-Based Microbiome Analysis (PluMA).



Conceptual design of PLuMA.

PluMA is available at: http://biorg.cis.fiu.edu/pluma. For any questions, please contact lead developer Trevor Cickovski at tcickovs@fiu.edu.

COMAN web-server for COmprehensive Metatranscriptome ANalysis

COMAN | Analysis | Instructions | Example Data | Contact | Citation

COMAN: the COmprehensive Metatranscriptome ANalysis web-server

Your job ID is: **COMAN20151218XMP0TW**

The number of files you have uploaded is : 0 (as listed below) (may also include files generated during analysis)

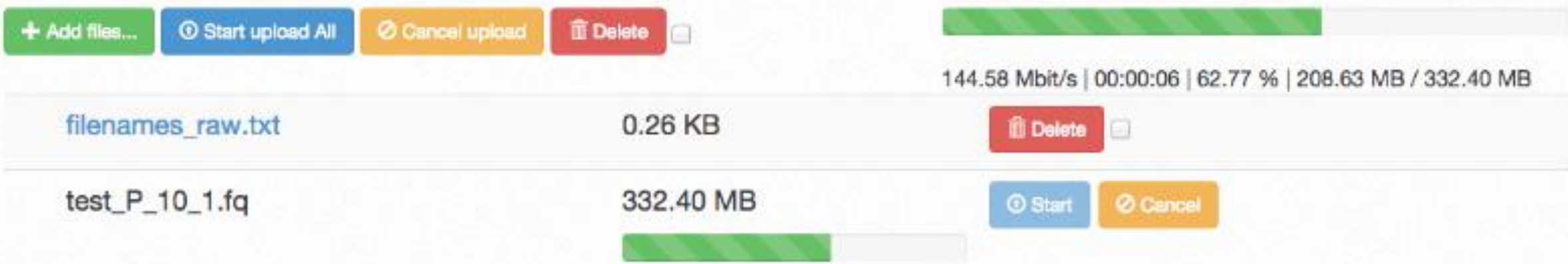
Notice! You have not uploaded the metadata file "filenames_raw.txt".

Make sure you have uploaded all your data to perform the analysis. Please double check the number of sequence files and their file names so that they match the ones provided in metadata file.

If you have not finished uploading all necessary data, please continue: Upload more data

After finishing uploading files, you may go back to **START YOUR ANALYSIS.**

File Upload (based on jQuery) with multiple file selection, drag&drop support and progress bars.



The metadata you uploaded contains the information below (If error messages appear or you are not able to see a table, it is most likely due to the problem with the metadata file "filenames_raw.txt". Please build it according to the example data):

Filename	Condition
baseline_3_100K_1.fq	Control
baseline_3_100K_2.fq	Control
baseline_4_100K_1.fq	Control
baseline_4_100K_2.fq	Control
treatment_3_100K_1.fq	Treatment
treatment_3_100K_2.fq	Treatment
treatment_4_100K_1.fq	Treatment
treatment_4_100K_2.fq	Treatment

Make sure you have uploaded all your data to perform the analysis. Please double check the number of sequence files and their file names so that they match the ones provided in metadata file.

If you have not finished uploading all necessary data, please continue: Upload more data

If all files have been uploaded and double-check has been performed (you may specify the FDR-corrected p-value used for DE analysis), click the "Analyse" button to initiate the analysis.

We will notify you by email (if provided) once the data analysis is completed.

FDR-Corrected P-Value Cutoff explain?

0.10

Analyse

The metadata you uploaded contains the information below (If error messages appear or you are not able to see a table, it is most likely due to the problem with the metadata file "filenames_raw.txt". Please build it according to the example data):

Filename	Condition
baseline_3_100K_1.fq	Control
baseline_3_100K_2.fq	Control
baseline_4_100K_1.fq	Control
baseline_4_100K_2.fq	Control
treatment_3_100K_1.fq	Treatment
treatment_3_100K_2.fq	Treatment
treatment_4_100K_1.fq	Treatment
treatment_4_100K_2.fq	Treatment

Make sure you have uploaded all your data to perform the analysis. Please double check the number of sequence files and their file names so that they match the ones provided in metadata file.

If you have not finished uploading all necessary data, please continue: Upload more data

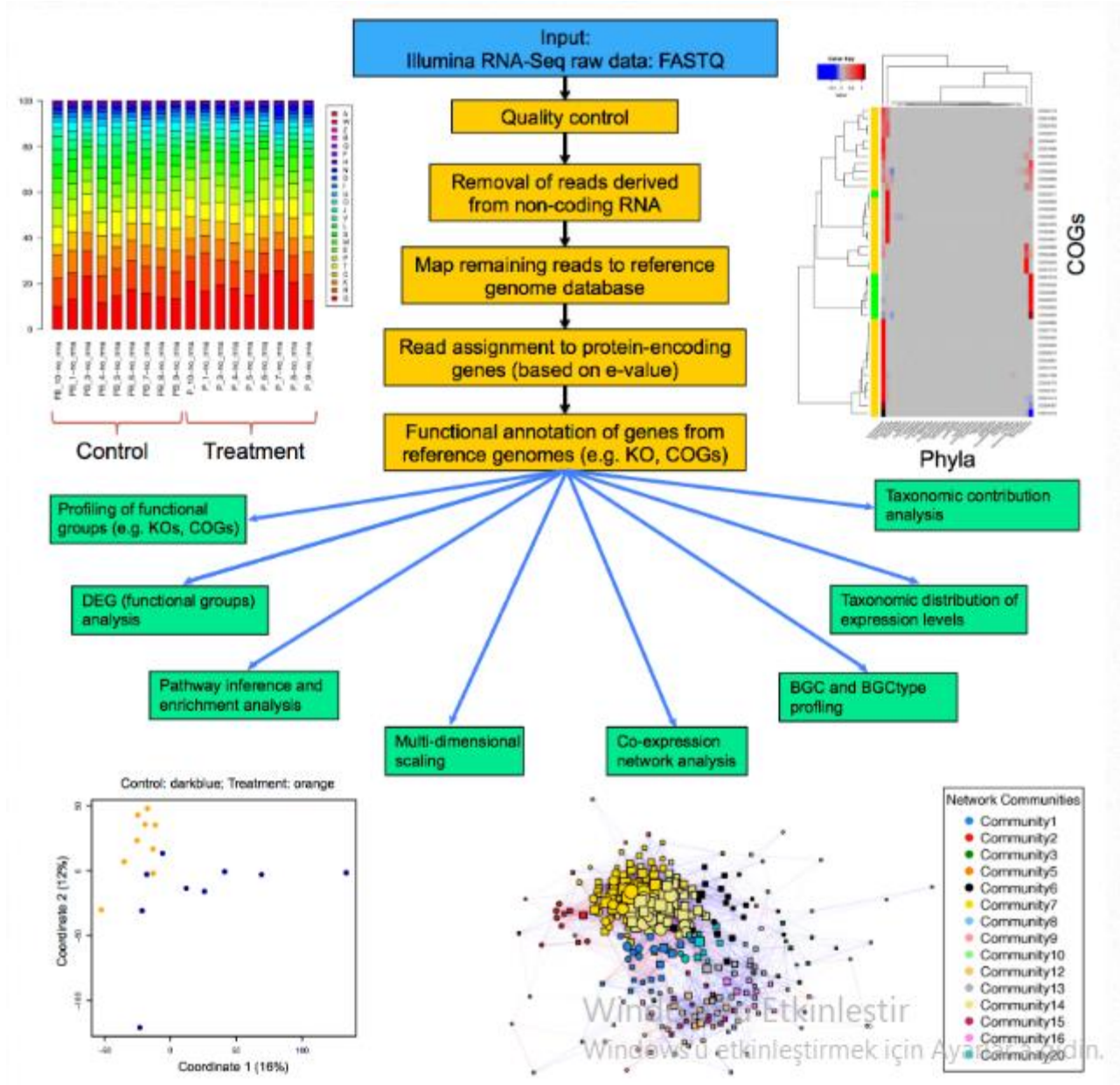
If all files have been uploaded and double-check has been performed (you may specify the FDR-corrected p-value used for DE analysis), click the "Analyse" button to initiate the analysis.

We will notify you by email (if provided) once the data analysis is completed.

FDR-Corrected P-Value Cutoff explain?

0.10

Analyse



https://cge.cbs.dtu.dk/services/

Center for Genomic Epidemiology

Username
Password
New Reset Login

Home | Services | Datasets

Overview of Services

Workflows

Bacterial Analysis Pipeline (Batch Upload)

Phenotyping

ResFinder
LRE-Finder
KmerResistance
PathogenFinder
VirulenceFinder
Restriction-ModificationFinder

Typing

KmerFinder
SpeciesFinder
MLST
PlasmidFinder
pMLST
cgMLSTFinder
SerotypeFinder
FimTyper
CHTyper
spaTyper
PAst
SCCmecFinder

Phylogeny

CSI Phylogeny
NDtree
Evergreen
snpTree (Out of order, use CSI Phylogeny or NDtree)
TreeViewer

Genome Assembly

Assembler
SPAdes

PCR-tools

RUCS

Other

PanFunPro
MyDbFinder
SPIFinder
HostPhinder
DeHumanizer
MetaPhinder

Karbapenem Dirençli İki K.pneumoniae Kökeninin Resistom Analizi

Mert Ahmet Kuşkucu[1], İlker İnanç Balkan[2], Bilgül Mete[2], Hatice Yaşar Arsu[2], Ebru Yücebağ[1], Gökhan Aygün[1], Kenan Midilli[1]

1 İstanbul Üniversitesi-Cerahpaşa, Cerrahpaşa Tıp Fakültesi, Tıbbi Mikrobiyoloji Anabilim Dalı
2 İstanbul Üniversitesi-Cerahpaşa, Cerrahpaşa Tıp Fakültesi, Enfeksiyon Hastalıkları ve Klinik Mikrobiyoloji Anabilim Dalı

Journal of
Proteomics & Bioinformatics

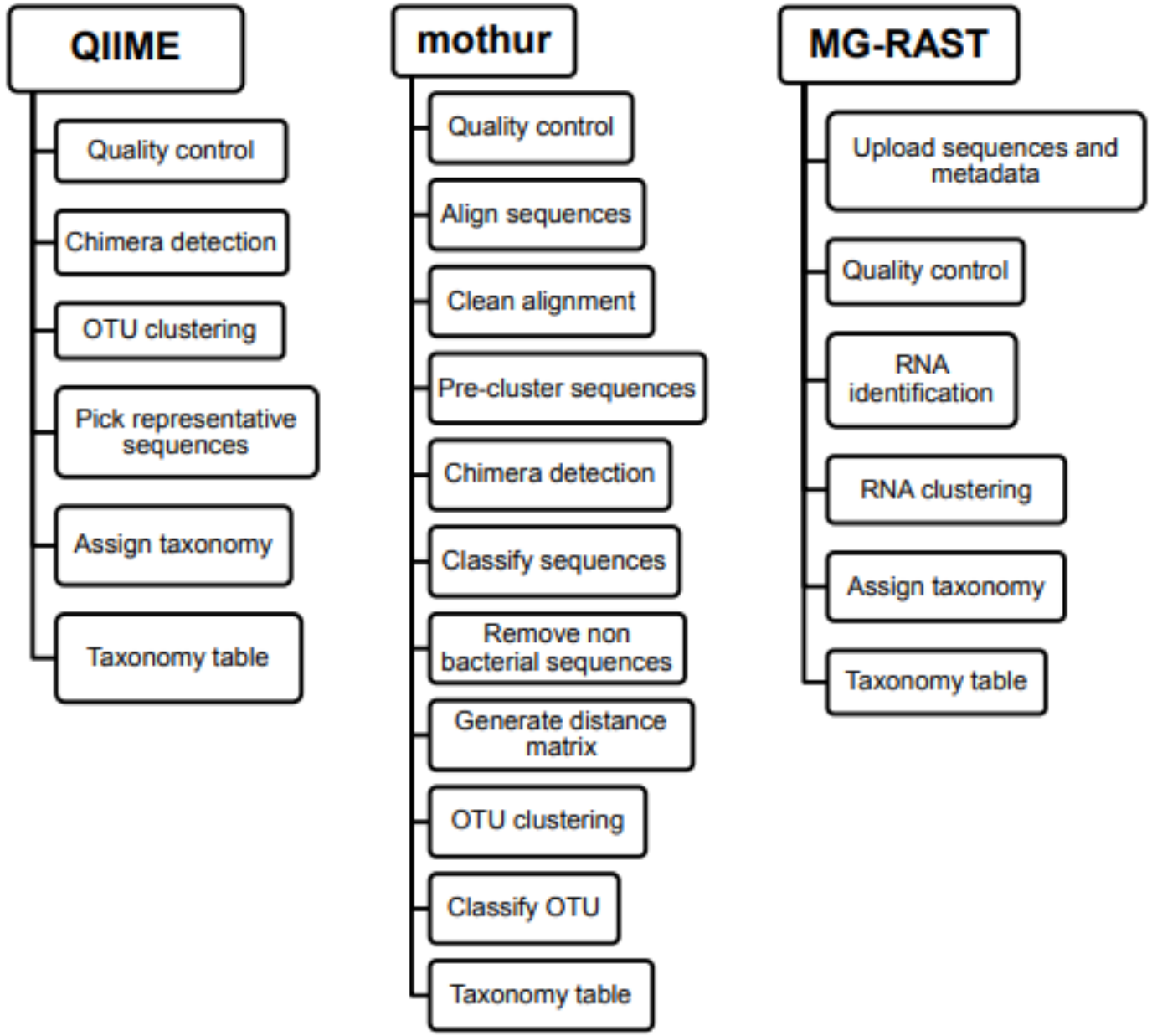
Plummer et al., J Proteomics Bioinform 2015, 8:12
http://dx.doi.org/10.4172/jpb.1000381

Research Article Open Access

A Comparison of Three Bioinformatics Pipelines for the Analysis of Preterm Gut Microbiota using 16S rRNA Gene Sequencing Data

Erica Plummer[1*], Jimmy Twin[1], Dieter M. Bulach[2,3,4], Suzanne M. Garland[1,3,5,6] and Sepehr N Tabrizi[1,3,5,6]

1Murdoch Childrens Research Institute, The Royal Children's Hospital, Flemington Rd, Parkville, Victoria 3052 Australia
[2] Monash University, Victoria 3800, Australia
[3]The University of Melbourne, Victoria 3050, Australia
[4]Victorian Life Sciences Computation Initiative, The University of Melbourne, Parkville Campus, LAB-14, 700 Swanston St, Carlton, Victoria 3053, Australia
[5]The Royal Women's Hospital, 20 Flemington Rd, Parkville, Victoria 3052, Australia
[6]The Royal Children's Hospital, 50 Flemington Rd, Parkville, Victoria 3052, Australia


QIIME
Quality control
Chimera detection
OTU clustering
Pick representative sequences
Assign taxonomy
Taxonomy table
mothur
Quality control
Align sequences
Clean alignment
Pre-cluster sequences
Chimera detection
Classify sequences
Remove non bacterial sequences
Generate distance matrix
OTU clustering
Classify OTU
Taxonomy table
MG-RAST
Upload sequences and metadata
Quality control
RNA identification
RNA clustering
Assign taxonomy
Taxonomy table

	QIIME	mothur	MG-RAST	p-value
Approximate analysis time (for the study dataset)	1 hour	10 hours	2 days	-
Number of reads uploaded	159,691	159,691	159,691	-
Number of reads post demultiplexing and QC	131,661	132,314	131,368	0.695
Number of reads assigned identity	123,909	128,064	123,022	0.0547
Number of unclassified reads at phylum level (%)	104 (0.08)	155 (0.12)	14,199 (11.54)	<0.0001*
Number of unclassified reads at genus level (%)	12,724 (10.27)	37,039 (28.92)	20,253 (16.46)	0.0814
Number of genera identified	60	50	57	-
Genus Richness (median, IQR)	10 (9-15)	8 (5-12)	9 (7-14)	<0.0001*
Effective number of genera (median, IQR)	3 (2-4)	2 (2-3)	3 (3-4)	<0.0001*

QIIME	mothur	MG-RAST
Bifidobacterium (35.79%)	*Bifidobacterium* (35.41%)	*Bifidobacterium* (36.96%)
Enterobacter (28.46%)	*Escherichia-Shigella* (12.38%)	*Klebsiella* (12.20%)
Enterococcus (4.09%)	*Enterococcus* (4.27%)	*Escherichia-Shigella* (7.20%)
Clostridium (3.44%)	*Staphylococcus* (3.41%)	*Enterococcus* (4.17%)
Staphylococcus (3.41%)	*Clostridium* (3.26%)	*Veillonella* (3.57%)

NB: abundance of each genus is expressed in parentheses following the genus name *Escherichia* and *Shigella* are difficult to resolve using 16S rRNA gene analysis [31] and as such have been grouped together as one genus.

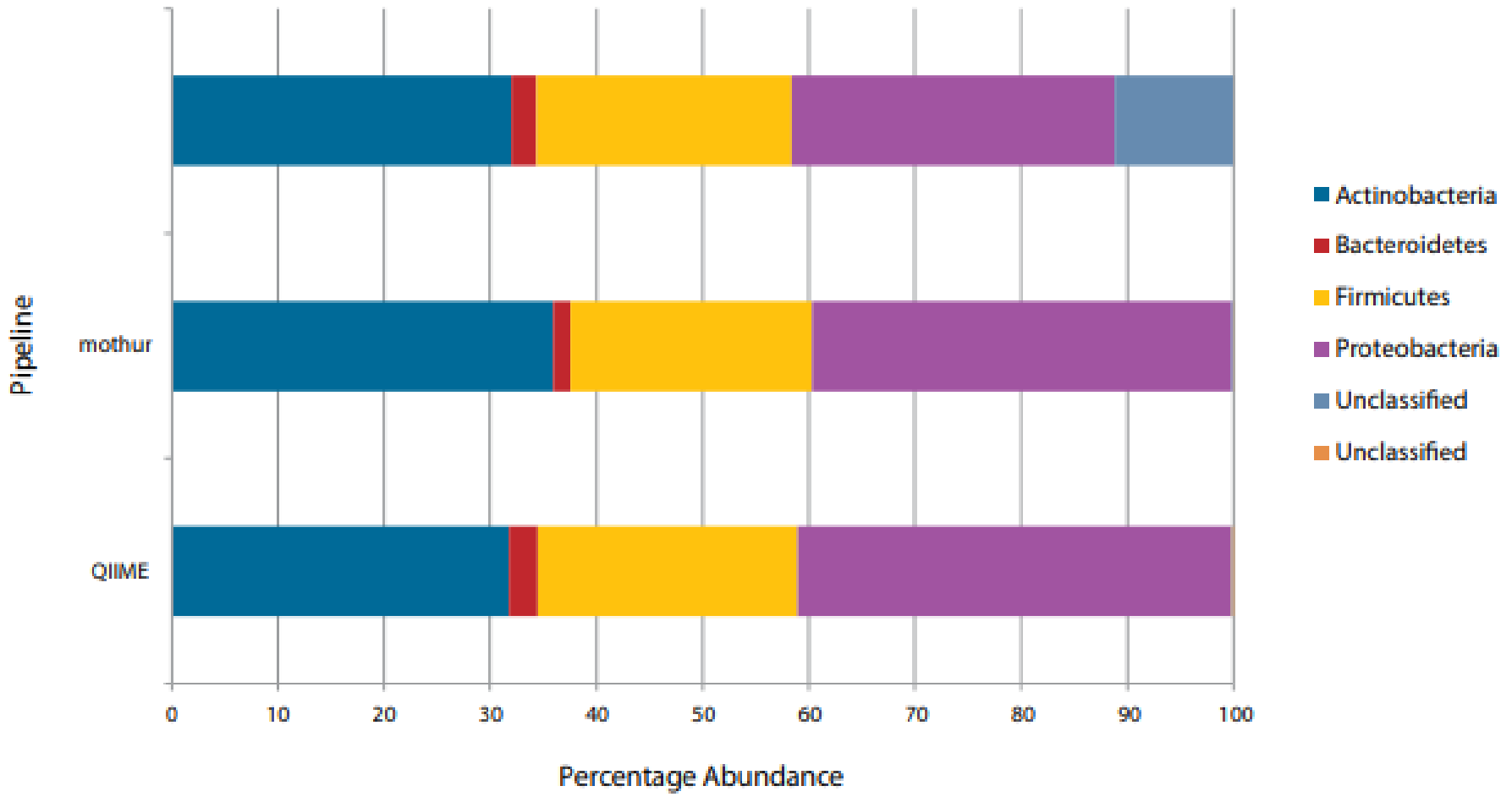


Figure 2: Comparison of taxonomic classification by QIIME, mothur and MG-RAST at the phylum level. This presents the abundance of different phyla detected by QIIME, mothur and MG-RAST. The phyla present in the less than 1% category are: Verrucomicrobia and Fusobacteria.

" I'M A DO-IT-YOURSELFER, BUT I'VE NEVER BEEN A DONE-IT-YOURSELFER . . . "

Teşekkürler. . .