Csaba Kerepesi, Dániel Bánky, Vince Grolmusz:

AmphoraNet: Taxonomic Composition Analysis of Metagenomic Shotgun Sequencing Data

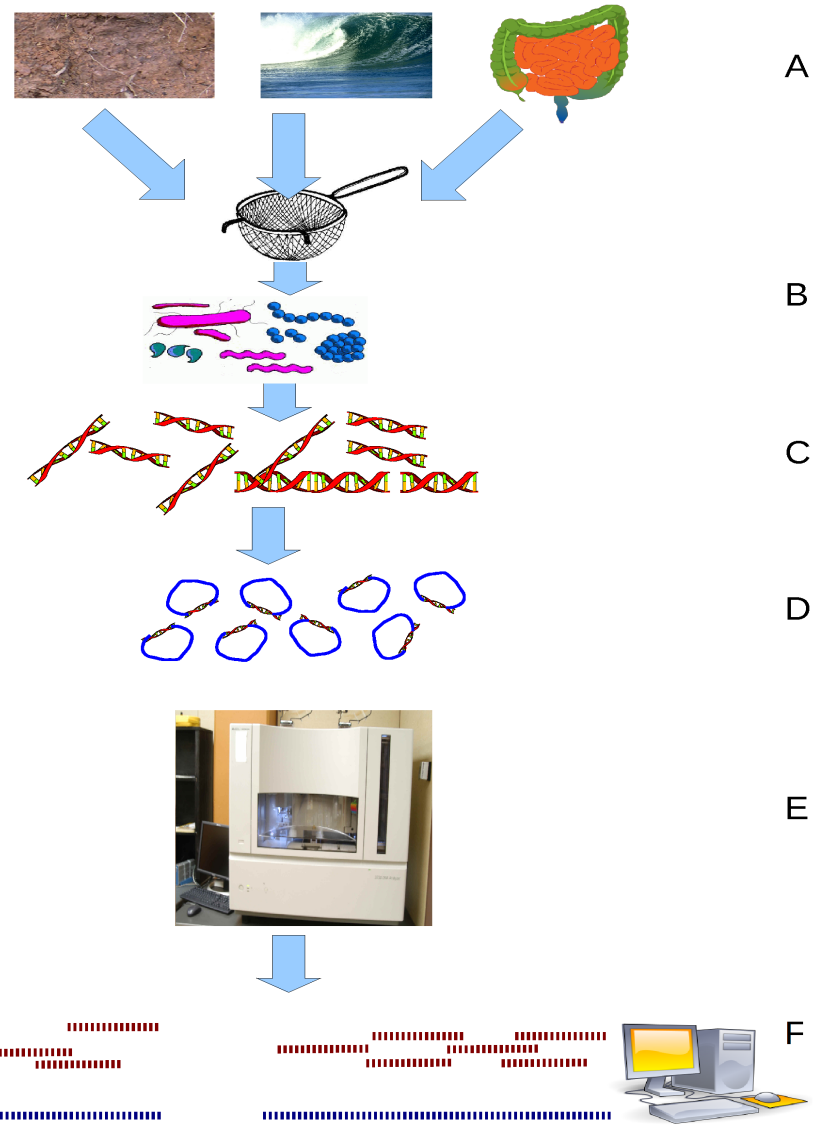
http://pitgroup.org/amphoranet/

PIT Bioinformatics Group,
Department of Computer Science
Eötvös Loránd University, Budapest, Hungary

Metagenomics

- Past 4-5 years: only the 10 percent of the bacterial species are culturable
- Solution: bioinformatics, metagenomics analysis that is the direct analysis of genetic material of the microorganisms in environmental or clinical samples
- Metagenomics samples: from seas, lakes, mines, animals, humans→ associations with diseases
- Cost of sequencing: in the year 2000 one billion base pairs cost about $400000, recently 1 billion base pairs cost about $50 → amount of the public data has grown rapidly
- Metagenomics analyzing tools: AMPHORA2 and its webserver implementation AmphoraNet
- The analysis is difficult because it is not possible to sequence DNA molecules separately, only short fragments mixed randomly (shotgun sequencing)

Shotgun sequencing ([6])



A (A) Sampling

B (B) Filtering by size

C (C) DNA extraction and lysis

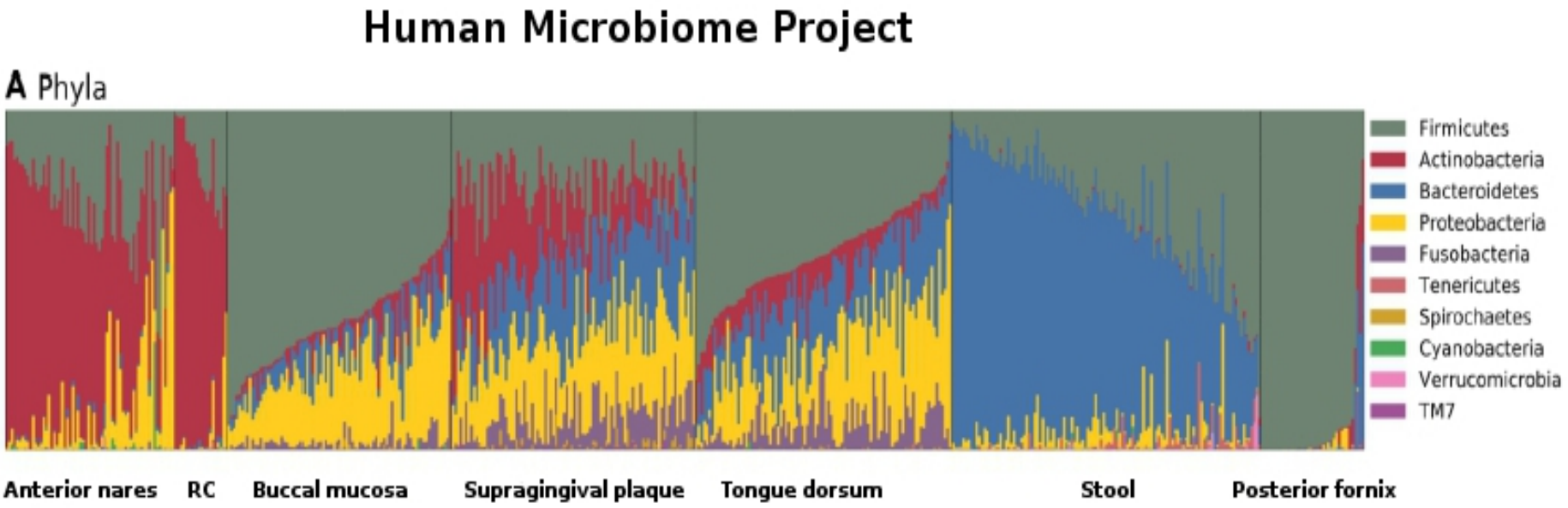
D (D) Cloning and library

E (E) Sequencing

F (F) Sequence assembly

Healthy microbiome

- Human microbiome: 100 times more bacteria than self cells. They are important in our subsistence
- Human Microbiome Project (USA, 2008-2012): samples from 15, 18 habitats of 242 healthy individuals at two different time points [1]
- Observation: there are differences at the level of individuals, but we can observe a typical average composition pattern in each body habitat



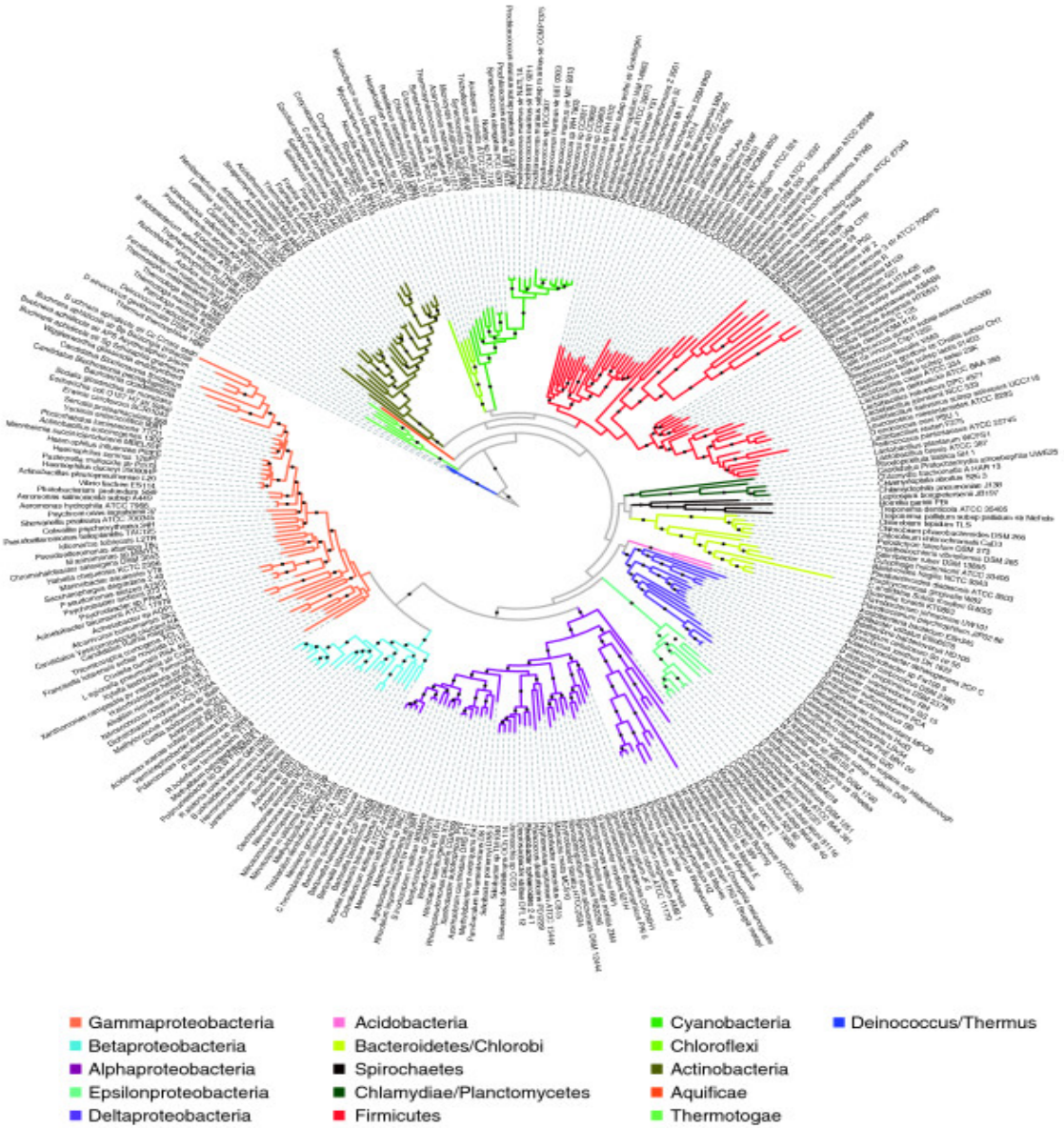
Microbiome of individuals suffered from diseases

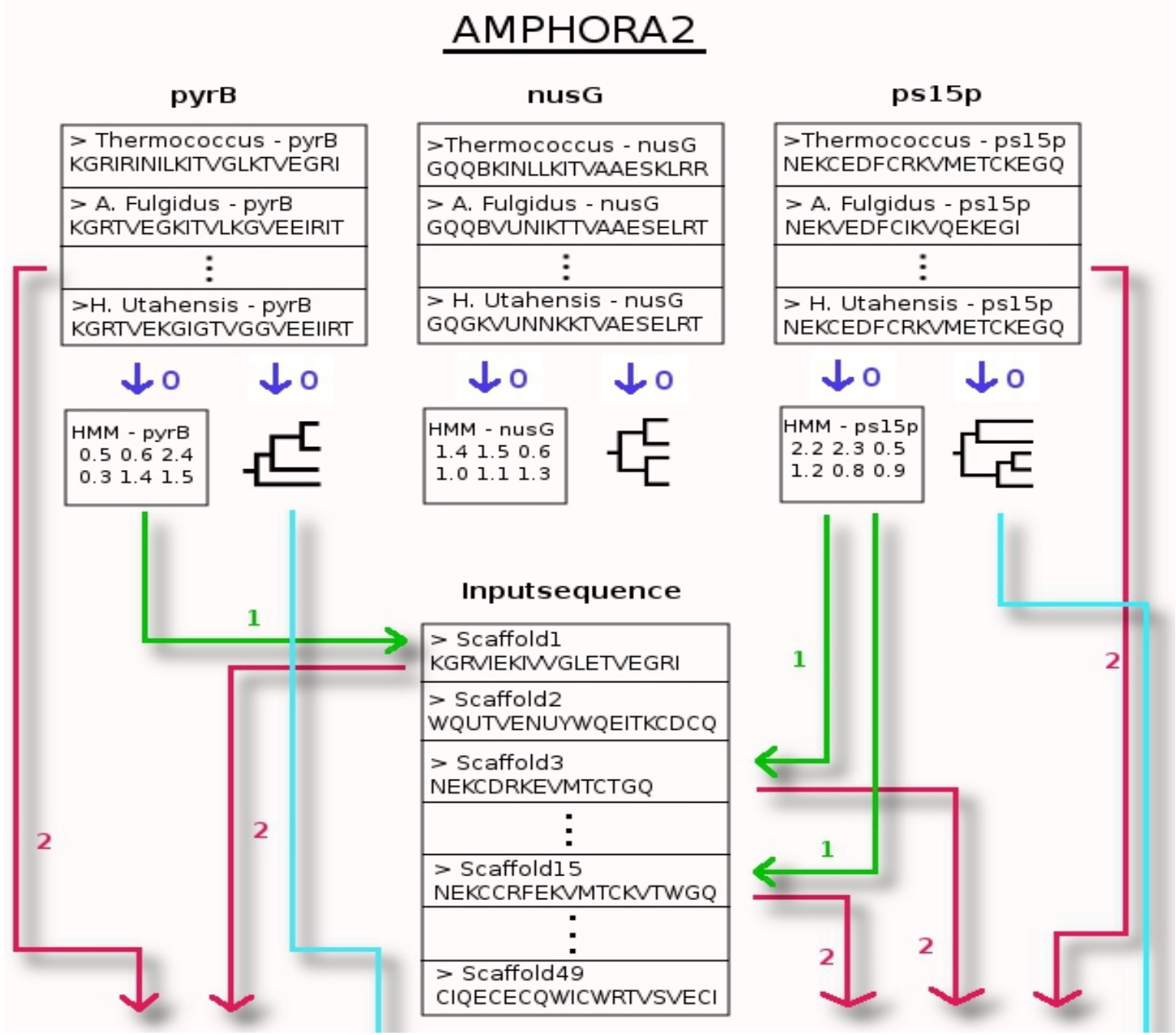
- Obesity: an increase in the phylum Firmicutes and the phylum Actinobacteria and a decrease in the phylum Bacteroidetes in the gut (2009, 145 sample from twins and parents, [2])
- Type 2 diabetes: reduced proportion of the class Clostridiales in the phylum Firmicutes (2010, 36 gut samples from Danish adults, [3]), a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens (2012, 345 gut samples from Chinese individuals, [4])
- Autism: lower abundances of the genera Prevotella, Coprococcus, and unclassified Veillonellaceae in the gut microbiome of autistic children (2013, 40 samples, [5])
- Application of the results: forecast, prevention or cure of diseases. But a lot of new samples and testing would be needed to confirm the current results and explore many other associations.

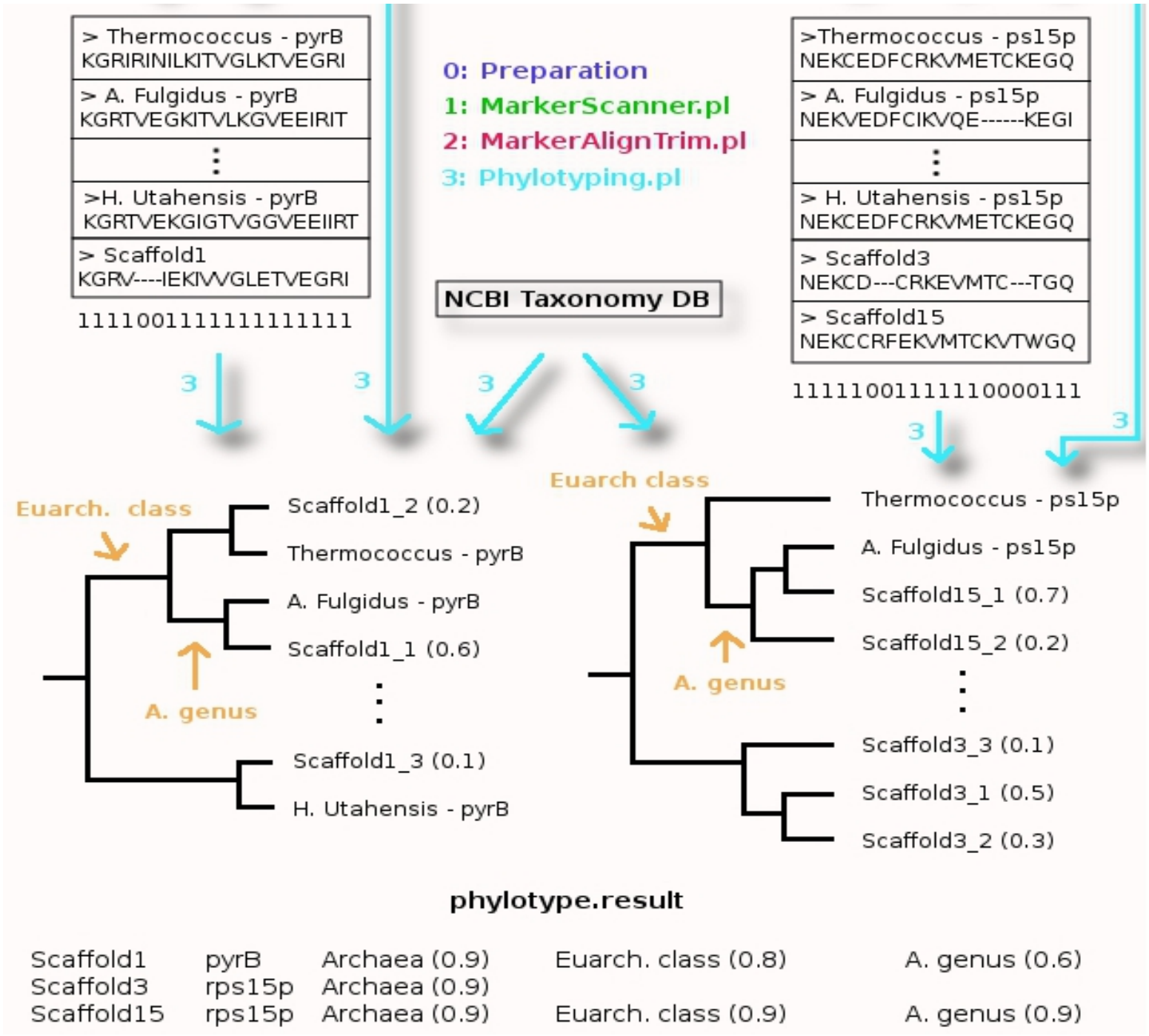
Metagenomics analyzing tools

- Most of the above-mentioned studies used 16S rRNA gene sequencing with the 454 platform. But there are tools for analyze shotgun sequencing data as well.
- Comparison based phylotyping (comparing DNA reads against known complete genomes represented species): MEGAN [7]
- Phylogeny based phylotyping (finding phylogenetic marker genes in the shotgun sequencing data and using evolutionary trees): AMPHORA2 [8] and its web server implementation AmphoraNet
- AMPHORA2 uses 31 bacterial and 104 archaeal protein-coding phylogenetic marker genes for metagenomic phylotyping
- Most of these marker genes are single copy genes, therefore AMPHORA2 is suitable for inferring the taxonomic composition of shotgun sequencing data

The Bacterial Tree of Life [9]




AMPHORA2
pyrB
> Thermococcus - pyrB
KGRIRINILKITVGLKTVEGRI
> A. Fulgidus - pyrB
KGRTVEGKITVLKGVEEIRIT
>H. Utahensis - pyrB
KGRTVEKGIGTVGGVEEIIRT
HMM - pyrB
0.5 0.6 2.4
0.3 1.4 1.5
nusG
>Thermococcus - nusG
GQQBKINLLKITVAAESKLRR
> A. Fulgidus - nusG
GQQBVUNIKTTVAAESELRT
> H. Utahensis - nusG
GQGKVUNNKKTVAESELRT
HMM - nusG
1.4 1.5 0.6
1.0 1.1 1.3
ps15p
>Thermococcus - ps15p
NEKCEDFCRKVMETCKEGQ
> A. Fulgidus - ps15p
NEKVEDFCIKVQEKEGI
> H. Utahensis - ps15p
NEKCEDFCRKVMETCKEGQ
HMM - ps15p
2.2 2.3 0.5
1.2 0.8 0.9
Inputsequence
> Scaffold1
KGRVIEKIVVGLETVEGRI
> Scaffold2
WQUTVENUYWQEITKCDCQ
> Scaffold3
NEKCDRKEVMTCTGQ
> Scaffold15
NEKCCRFEKVMTCKVTWGQ
> Scaffold49
CIQECECQWICWRTVSVECI
1
2



phylotype.result

Scaffold1	pyrB	Archaea (0.9)	Euarch. class (0.8)	A. genus (0.6)
Scaffold3	rps15p	Archaea (0.9)		
Scaffold15	rps15p	Archaea (0.9)	Euarch. class (0.9)	A. genus (0.9)

AmphoraNet webserver: amphoranet.pitgroup.org

About us:
- Overview
- Publicity
- Pictures gallery
- Downloads

Web servers:
- WikiPDB→
- ProtDict
- QuickKEGG
- Decomp
- Triad database
- Nascent
- SwissAlign
- Brownian
- AmphoraNet
- PSA

AmphoraNet

The webserver implementation of the AMPHORA2 workflow for phylogenetic analysis of metagenomic shotgun sequencing data

AmphoraNet is the webserver implementation of the AMPHORA2 workflow. It uses 31 bacterial and 104 archaeal protein-coding phylogenetic marker genes for metagenomic phylotyping. Most of these marker genes are single copy genes, therefore AmphoraNet is suitable for inferring the taxonomic composition of bacterial and archaeal communities from metagenomic shotgun sequencing data.

Input values needed

FASTA file (max size: 10.0 MB) Tallózás... Nincs kijelölve fájl.

Sequence: (•) DNA sequence ( ) Amino acid sequence

Marker genes: (•) Bacteria ( ) Archaea ( ) Archaea and bacteria

Check values

Using AmphoraNet step by step

1. Upload a metagenomic sequence file in FASTA or FASTAQ format and at most 10.0 MB size. Sample input file (HMP - buccal mucosa metagenome). Other input files.
2. Choose the type of your input data: DNA sequence or amino acid sequence.
3. Choose the type of the marker genes you would like to search: bacteria, archaea or both.
4. Click the "Check values" button. If everything is OK then click the "Schedule job" button.

AmphoraNet sample input files

AmphoraNet ›

Sample input files to try AmphoraNet

1 bejegyzés 1 szerzőtől

Csaba Kerepesi

Az üzenet fordítása magyar nyelvre

Here we collected input files and databases to try AmphoraNet.

Bacterial genomes:

- Treponema pallidum subsp. pallidum DAL-1, complete genome (1.1 MB, 1.1 million bp, running time: about 30 min.)
- More Bacteria Genomes from EBI

Archaeal genomes:

- Archaeoglobus profundus DSM 5631, complete genome (1.5 MB, 1.5 million bp, running time: about 15 min.)
- More archaeal Genomes from EBI

Metagenomic sequences:

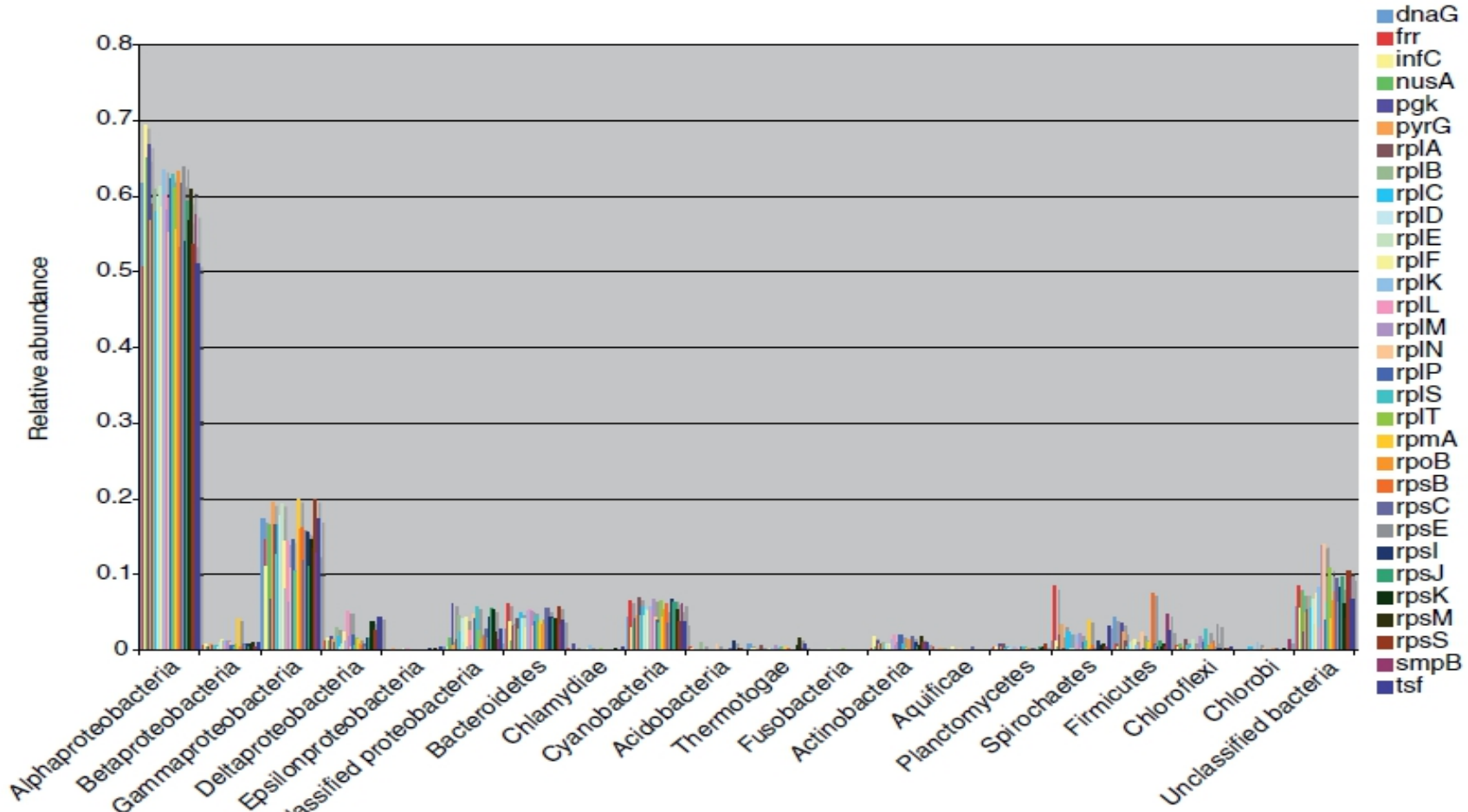
- Human Genome Project - Buccal Mucosa sample (SRS050007) (0.7 MB, 0.7 million bp, running time: about 20 min.)
- More Human Microbiome Project: Illumina WGS Reads and Assemblies
- CAMERA2 - Community Cyberinfrastructure for Advanced Microbial Ecology Research & Analysis
- Type 2 Diabetes gut metagenome (microbiome) data from 368 Chinese samples and updated metagenome gene catalog

AmphoraNet sample output file

Query	Marker	Superkingdom	Phylum	Class	Order	Family	Genus	Species
C341382_2	dnaG	Bacteria(1.00)	Firmicutes(1.00)	Bacilli(1.00)	Lactobacillales(1.00)	Streptococcaceae(1.00)	Streptococcus(1.00)	Streptococcus pneumoniae(1.00)
C342110_2	dnaG	Bacteria(0.99)	Firmicutes(0.99)	Bacilli(0.99)	Lactobacillales(0.99)	Streptococcaceae(0.99)	Streptococcus(0.99)	Streptococcus pneumoniae(0.44)
C343864_5	infC	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.76)
C341888_4	pgk	Bacteria(1.00)	Proteobacteria(1.00)	Gammaproteobacteria(1.00)	Pasteurellales(1.00)	Pasteurellaceae(1.00)	Aggregatibacter(0.71)	Aggregatibacter actinomycetemcomitans(0.24)
C343484_9	pgk	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus oralis(0.88)
C341726_4	pyrG	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus mitis(0.94)
SRS050007_WUGC_scaffold_2240_7	pyrG	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus pseudopneumoniae(0.11)
C343398_7	rplA	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus pneumoniae(0.62)
SRS050007_WUGC_scaffold_2153_2	rplA	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.77)
C343640_5	rplB	Bacteria(1.00)	Proteobacteria(1.00)	Gammaproteobacteria(1.00)	Pasteurellales(1.00)	Pasteurellaceae(1.00)	Haemophilus(0.87)	Haemophilus influenzae(0.73)
C342898_5	rplB	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.66)
C344044_3	rplB	Bacteria(0.99)	Proteobacteria(0.99)	Gammaproteobacteria(0.99)	Pasteurellales(0.99)	Pasteurellaceae(0.99)	Haemophilus(0.99)	Haemophilus parainfluenzae(0.99)
C343982_3	rplB	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.82)
C341820_3	rplC	Bacteria(0.95)	Actinobacteria(0.95)	Actinobacteria(0.95)	Actinomycetales(0.95)	Micrococcaceae(0.95)	Rothia(0.95)	Rothia mucilaginosa(0.91)
C343034_1	rplC	Bacteria(0.99)	Proteobacteria(0.99)	Gammaproteobacteria(0.99)	Pasteurellales(0.99)	Pasteurellaceae(0.99)	Haemophilus(0.99)	Haemophilus influenzae(0.81)
C343580_5	rplC	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.79)
C344220_6	rplE	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.82)
C344106_5	rplE	Bacteria(1.00)	Proteobacteria(1.00)	Gammaproteobacteria(1.00)	Pasteurellales(1.00)	Pasteurellaceae(1.00)	Haemophilus(0.81)	Haemophilus influenzae(0.68)
C342820_2	rplF	Bacteria(1.00)	Actinobacteria(1.00)	Actinobacteria(1.00)	Actinomycetales(1.00)	Micrococcaceae(1.00)	Rothia(1.00)	Rothia mucilaginosa(1.00)
C343946_4	rplF	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pseudopneumoniae(0.11)
C342602_1	rplL	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.82)
C342866_6	rplL	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.50)
C342108_5	rplM	Bacteria(1.00)	Proteobacteria(1.00)	Gammaproteobacteria(1.00)	Pasteurellales(1.00)	Pasteurellaceae(1.00)	Haemophilus(1.00)	Haemophilus influenzae(0.84)
C344016_1	rplM	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.77)
C342158_2	rplN	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.92)	Streptococcus mutans(0.18)
C342200_4	rplN	Bacteria(0.96)	Proteobacteria(0.96)	Gammaproteobacteria(0.96)	Pasteurellales(0.96)	Pasteurellaceae(0.96)	Actinobacillus(0.24)	Actinobacillus pleuropneumoniae(0.20)
C342196_2	rplS	Bacteria(0.95)	Proteobacteria(0.95)	Gammaproteobacteria(0.95)	Pasteurellales(0.95)	Pasteurellaceae(0.95)	Haemophilus(0.95)	Haemophilus influenzae(0.95)
C342138_1	rplT	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus suis(0.27)
C343464_7	rpmA	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.63)
C343390_4	rpoB	Bacteria(1.00)	Actinobacteria(1.00)	Actinobacteria(1.00)	Actinomycetales(1.00)	Micrococcaceae(1.00)	Rothia(1.00)	Rothia mucilaginosa(1.00)
C343814_6	rpoB	Bacteria(0.95)	Firmicutes(0.95)	Bacilli(0.95)	Lactobacillales(0.95)	Streptococcaceae(0.95)	Streptococcus(0.95)	Streptococcus mitis(0.73)
C341284_7	rpoB	Bacteria(1.00)	Actinobacteria(1.00)	Actinobacteria(1.00)	Actinomycetales(1.00)	Micrococcaceae(1.00)	Rothia(1.00)	Rothia mucilaginosa(0.85)
C342178_6	rpoB	Bacteria(1.00)	Actinobacteria(1.00)	Actinobacteria(1.00)	Actinomycetales(1.00)	Micrococcaceae(1.00)	Rothia(1.00)	Rothia mucilaginosa(0.88)
C342392_1	rpoB	Bacteria(1.00)	Fusobacteria(1.00)	Fusobacteriia(1.00)	Fusobacteriales(1.00)	Fusobacteriaceae(1.00)	Fusobacterium(1.00)	Fusobacterium nucleatum(1.00)
C341308_2	rpsC	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.72)
C344016_3	rpsI	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.55)
C344232_2	rpsK	Bacteria(0.97)	Firmicutes(0.97)	Bacilli(0.97)	Lactobacillales(0.97)	Streptococcaceae(0.97)	Streptococcus(0.97)	Streptococcus pneumoniae(0.69)
C344056_3	rpsM	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pneumoniae(0.52)
C342284_5	rpsS	Bacteria(0.96)	Firmicutes(0.96)	Bacilli(0.96)	Lactobacillales(0.96)	Streptococcaceae(0.96)	Streptococcus(0.96)	Streptococcus pyogenes(0.58)
C343558_3	smpB	Bacteria(0.99)	Firmicutes(0.99)	Bacilli(0.99)	Lactobacillales(0.99)	Streptococcaceae(0.99)	Streptococcus(0.99)	Streptococcus oralis(0.99)

Visualization of the output files

We are working on the automated visualization of the output files generated AmphoraNet as Eisen et al's figure below [9]:



Thank you for your attention!

References

- [1] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214 (2012).
- [2] Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. Nature 457,480–484 (2009).
- [3] Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, et al. (2010) Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-Diabetic Adults. PLoS ONE 5(2): e9085.
- [4] Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60 (2012).
- [5] Kang D-W, Park JG, Ilhan ZE, Wallstrom G, LaBaer J, et al. (2013) Reduced Incidence of Prevotella and Other Fermenters in Intestinal Microflora of Autistic Children. PLoS ONE 8(7): e68322.
- [6] Wooley JC, Godzik A, Friedberg I (2010) A Primer on Metagenomics. PLoS Comput Biol 6(2): e1000667
- [7] D.H. Huson, A.F. Auch, Ji Qi and S.C. Schuster, MEGAN Analysis of Metagenomic Data, Genome Research. 17:377-386, 2007.
- [8] Wu M, Scott AJ.: Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics. 2012 Apr 1;28(7):1033-4. doi: 10.1093/bioinformatics/bts079. Epub 2012 Feb 12.
- [9] Wu M, Eisen JA.: A simple, fast, and accurate method of phylogenomic inference. Genome Biol. 2008 Oct 13;9(10):R151.