San Diego State University

ADAPTdb/ADAPT - A Framework for the Analysis of ARISA Data Sets

Robert Schmieder, Matthew Haynes, Elizabeth Dinsdale, Forest Rohwer, and Robert Edwards

ACEL LAB AT COMPUTATIONAL SCIENCE RESEARCH CENTER

Introduction

The automated ribosomal intergenic spacer analysis (called ARISA) is an analytical fingerprinting technique for characterizing microbial community composition. ARISA analyzes the length of the intergenic spacer region between 16S and 23S rRNA genes (called ITS region) present in almost all Bacteria and Archaea. ARISA has been used to analyze several bacterial communities, including samples from freshwater, marine, and terrestrial environments. It is used for clinical studies to detect and identify microbes species. The main advantages of ARISA are that it is cheaper and quicker than other community profiling techniques.

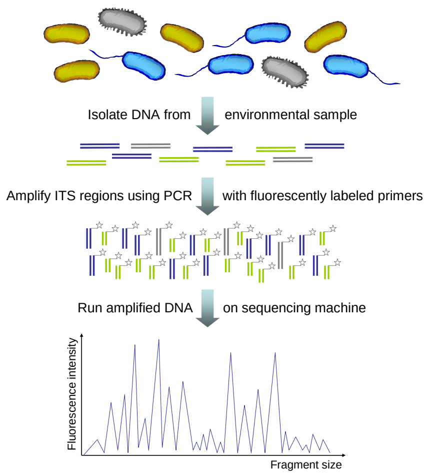


Figure 1. Basic approach of ARISA

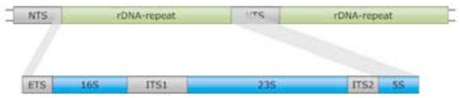


Figure 2. Arrangement of prokaryotic 16S, 23S and 5S rDNA in a tandem repetitive cluster. NTS stands for non-transcribed spacer, ETS for external transcribed spacer and ITS for internal transcribed spacer.

Previously, ARISA data analysis was done manually. Therefore, we have created a database to store and maintain ITS regions along with additional information about their host organisms, which can be used for the analysis of ARISA data sets. This program uses the database to map known ITS regions to new ARISA data sets and performs pathogenic and trophic comparisons based on the taxonomic profiles of the ARISA data sets.

ADAPTdb

ADAPTdb is a collection of annotated ITS regions using a MySQL database as a data storage backend.

Used data resources:

- RISSC (collection of spacer regions)
- Entrez Sequences and Entrez Genomes database
- Pathogenic information manually extracted from literature and from the Lproks file (for completely sequenced bacterial genomes)

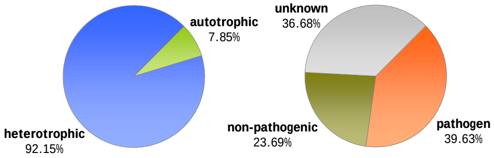


Figure 3. Trophic and pathogenic composition of organism entries in ADAPTdb.

Phylum	# Organisms	# ITS regions	ITS / organism
Crenarchaeota	21 (0.92%)	69 (1.46%)	3.29
Euryarchaeota	43 (1.89%)	46 (0.97%)	1.07
Actinobacteria	452 (19.83%)	628 (13.30%)	1.39
Bacteriodetes	33 (1.45%)	75 (1.59%)	2.27
Chlamydiae	55 (2.41%)	71 (1.50%)	1.29
Cyanobacteria	173 (7.59%)	223 (4.72%)	1.29
Firmicutes	569 (24.97%)	1467 (31.07%)	2.58
Fusobacteria	38 (1.67%)	46 (0.97%)	1.21
Proteobacteria	853 (37.43%)	2020 (42.79%)	2.37
Rare Bacterial spp.	42 (1.84%)	76 (1.63%)	1.81
Total	**2279**	**4721**	**2.07**

Table 1. Taxonomic composition of ADAPTdb entries.

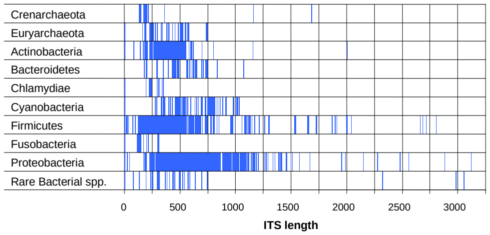


Figure 4. ITS length distribution of each phylum in ADAPTdb. The blue bars mark a certain ITS length if there is at least one database entry for this length, otherwise its white.

ADAPT

<u>A</u>RISA <u>D</u>ata <u>A</u>nalysis program for <u>P</u>athogenic and <u>T</u>rophic comparison

ADAPT is a web-based analyzing program for ARISA data sets using a fingerprint database (ADAPTdb) in order to:

- Predict the taxonomic composition of the underlying microbial communities
- Perform trophic and pathogenic comparisons of the samples from which the ARISA data was obtained

Analyzing steps include:

- ARISA profile filtering (remove noise and uninformative peaks)
- Database search (mapping ITS regions to ADAPTdb to find similar regions using a "binning" strategy)
- Calculate autotrophic/heterotrophic and pathogenic/non-pathogenic values using two different methods
- Performing multiple samples analysis (if more than one sample is provided in the ARISA data set)

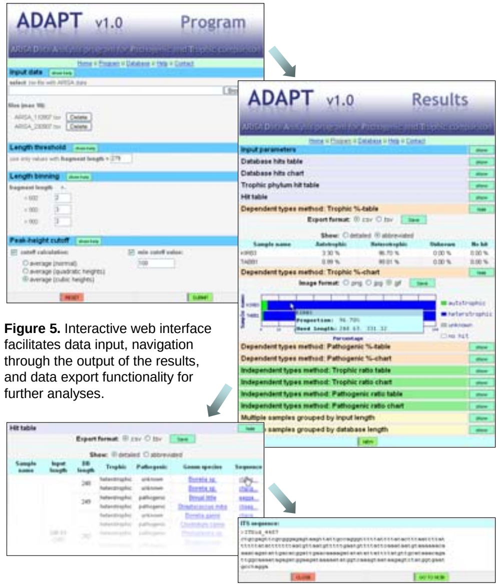


Figure 5. Interactive web interface facilitates data input, navigation through the output of the results, and data export functionality for further analyses.

Try out ADAPT on the web: **edwards.sdsu.edu/adapt**

Results

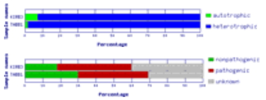
The ADAPTdb/ADAPT framework:

- automatically updates the database of ITS regions and host information
- maps known ITS regions to new ARISA data sets
- allows onsite data analysis of ARISA data sets and computation of the output
- is publicly available through a user-friendly web interface

Application example

Water samples were collected in 2005 at the coral reefs near the two Line Islands Kiritimati and Tabuaeran. The islands lie within the same biogeographic region, but show a gradient of human disturbance. The population gradient was used to investigate the influence of human activities on the microbes living in the coral reefs.
Bacteria were isolated by filtering about 1l of seawater through a filter. The DNA was then isolated from the bacterial cells, processed using ARISA, and analyzed using the ADAPT program.

Figure 6. Charts of the pathogenic and trophic proportion for each sample of the ARISA data set. KIRB3 represents samples taken near Kiritimati and TABB1 the samples taken near Tabuaeran.



The results identified that a range of pathogenic bacterial species were present on the two islands:

Kiritimati:

- *Bacteroides vulgatus*
- *Clostridium carnis*
- *Enterococcus spp.*
- *Haemophilus somnus*
- *Legionella pneumophila*
- *Mycobacterium spp.*
- *Streptococcus spp.*

Tabuaeran:

- *Bacillus spp.*
- *Clostridium spp.*
- *Enterococcus mundtii*
- *Haemophilus somnus*
- *Streptococcus spp.*
- *Staphylococcus simulans*

Acknowledgment

Thanks to Lutz Krause for database advise, the ACEL and Rohwer labs for funding.

Further Information

Please contact rschmieder@gmail.com. This poster is available online at edwards.sdsu.edu/posters/adapt.pdf