

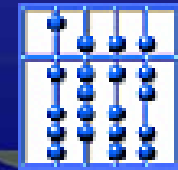
UGC UCU GAC CSU GUU GAG AAU AUC GGG UUU CGG GCU CAU
CUA CGA UGC GUA ACG GAC CCA UAC GCC CAU GCG UGA UGU
AAU UGU AAA AUA GGG UUG GGU CCG CAA ACA GCC CAU ACU
CUC UUA CCG CGU UAA UAC UUG AAU ACA UCC GUC UUA AAC
GAA GUA UUU CUA AUC AUU GCG UCC UCG AUU ACA AUU CGG
AAC CCA AUG UGU ACA GCU GGC GGA CUU AAU GAU UGA AGU GCU
CGG CAA AGG CAA GCA CGA AAA UUG AUC UCC CGA UUC UAU CCA
GUG GAA GAA CAU CAU GGA AAU CAG GUC CCA CUU UCA UUA UCA
UGA GAA GCA CCA AAA CCC GUA UUC GCA GCA UCA UCA UCC AGU
GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA

Computational Analysis of Data from Microbial Phylochip Microarrays

ChipAnalyser

Harald Meier

Bioinformatik, Lehrstuhl für Rechnertechnik und
Rechnerorganisation, Institut für Informatik,
Technische Universität München





Outline

- Introduction
- General Problems
- Analysis Methods and Algorithms
- Features and Solutions - The ChipAnalyser
- Biological Application
- Summary
- Future prospects

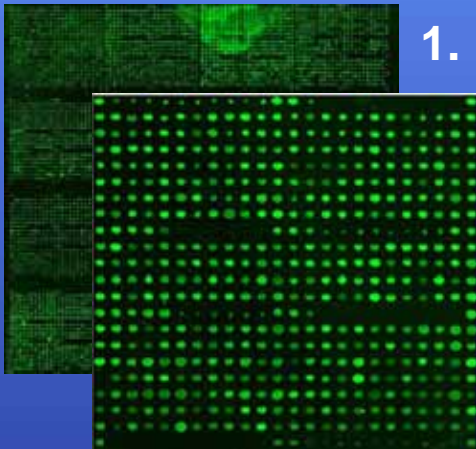
Introduction - Application

- Analysis of microbial communities in complex samples
- Application of diagnostic chips (conserved genes, i.e. ribosomal ribonucleic acid (rRNA))
- Interpretation of Complex Hybridization patterns

GOAL:

Qualitative biological meaning (i.e. presence/ absence of gene/organism in sample examined)

Hybridization Patterns → Biological Meaning



1. Image

A. Image Analysis



2. (Raw) Chip Data

Spot	Probe-ID	signal	noise
A 1 1	EUB-1	5802.93	402.73
A 1 2	SRB-2	492.625	362.11
A 1 3	LGC-3	544.218	317.35
A 1 4	SRP-152	2441.46	384.28
B 1 1
		.	.
		.	.
		.	.
		.	.
		.	.

C. Representation

Organism	Significance
D. geothermicum	75.67
D. gibsoniae	72.70
D. guttoideum	65.00
D. halophilum	64.32
.	.
.	.
E. coli	12.63
.	.
.	.

B. Data Analysis



3. Probe Data

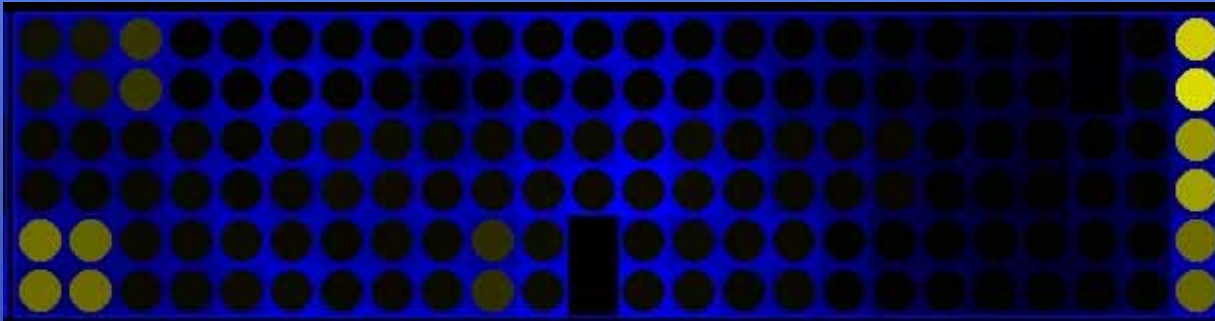
Probe-ID	Sequence	Specificity
EUB-1	acggtgcaaggtt	Bacteria
SRB-2	cggaataaattc	Group A
LGC-3	ggcctaaatggc	Sulfate Red. Bact.
SAB-4	cgtattccgtacc	D. guttoideum
.	.	.
SRP-152	acgtacgaacta	D. gibsoniae
.	.	.
.	.	.

4. Result Data

Image Analysis Software

- GenePixPro, Molecular Devices Corporation, Sunnyval, CA
- ImaGene (and GeneSight), BioDiscovery Corporation, Los Angeles, CA.
- GeneChip Analysis Suite, Affymetrix, Santa Clara, CA
- ScanArray Express, Perkin Elmer, Wellesley, MA
- QuantArray, Packard Biochip Technologies, Billerica, MA
- ScanAlyse <http://rana.lbl.gov/EisenSoftware.htm>
(Michael Eisen, Lawrence Berkeley National Lab, Berkeley, CA)

Data from Image Analysis – Signal and background



- Signal Pixel intensity
- Local Background Pixel Intensity (Unspecific)
 - Median, Mean, Standard Deviation of mean local background
- Global Background Pixel Intensity (Specific)

Chipanalysis of Diagnostic Arrays

- Bioconductor (Statistics)
- Cluster analysis and Treeview
 - Murray et al., PNAS, 2001
- GeneChip Analysis Suite, Affymetrix
 - Wilson et al., Mol Cell Probes, 2002; AEM, 2002
- Excel-Sheets
 - Loy et al., AEM, 2002, 2005
 - Bodrossy et al., Env Microbiol, 2003
 - Stralis-Pavese et al., Env Microbiol, 2004
 - Francois et al., Mol Cell Probes, 2006
 - Taroncher-Oldenburg et al., AEM, 2003

Affymetrix

Spotted Arrays

Outline

- Introduction
- **General and Particular Problems**
- Analysis Methods and Algorithms
- Features and Solutions - The ChipAnalyser
- Biological Application
- Summary
- Future prospects

General Data Analysis Problems

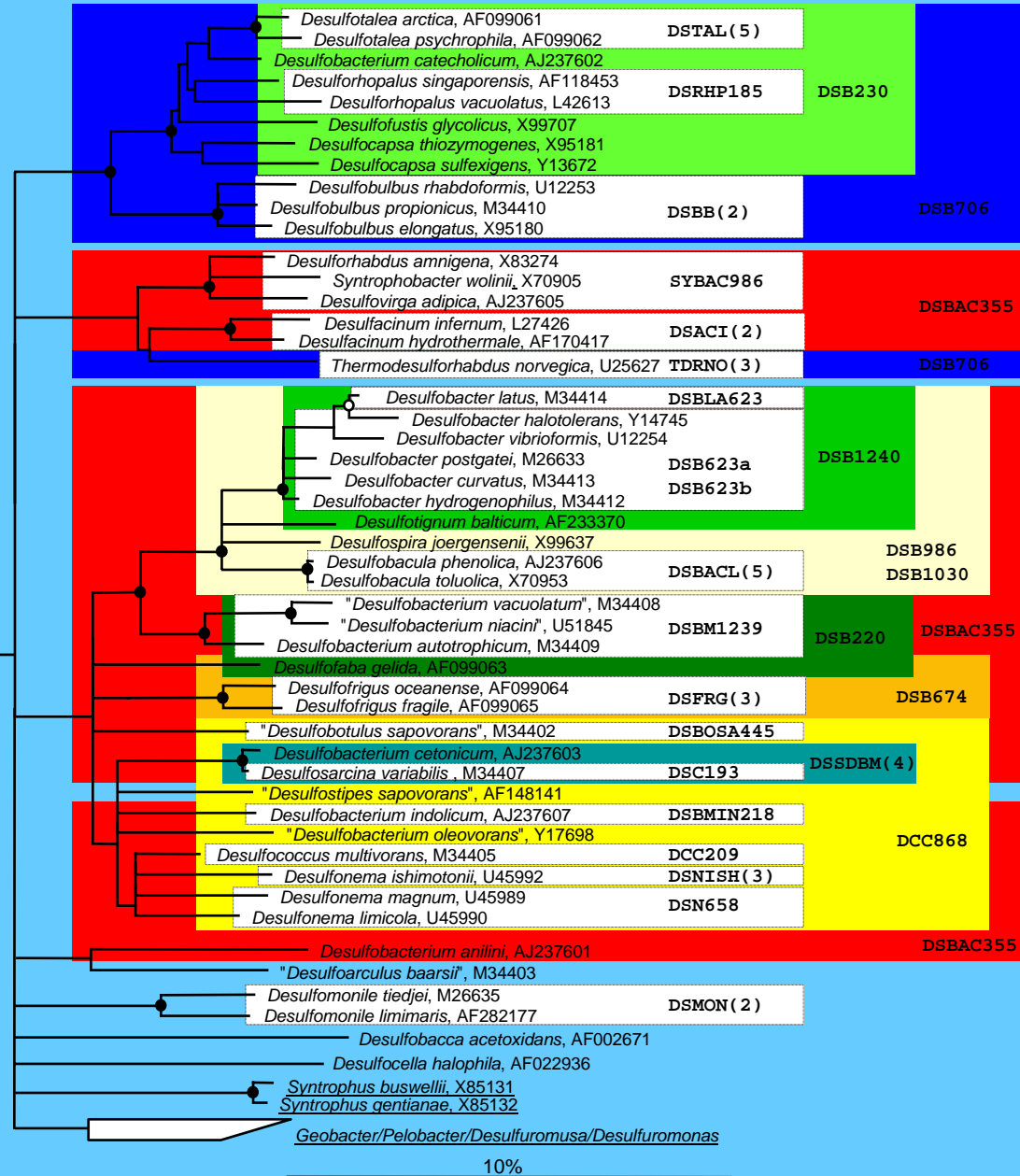
- Different data formats
- Handling large data lists
- Image Data Analysis
 - General Controls
 - Which spots to take into account ?
 - Threshold setting
 - Which analysis method (which data) to use ?
- Proper Visualization of Result Representation

Complexity Factors (using Phylochips and microbial community profiling)

- Probe dependant complexity factors
- Sequence databases increase rapidly → Evaluation of probe sets and specificities list
- “Biological” / methodological problems
 - Same probe, different target organisms → different signals intensities
 - Methodological Biases
 - different probes → different hybridisation behaviour
 - Any sequence could be in the sample

DELTA495a
 DELTA495b
 DELTA495c

16S rRNA-Based Phylogenetic Tree



**Multiple
 Probe
 Concept !**



Outline

- Introduction
- General and Particular Problems
- **Analysis Methods and Algorithms**
- Features and Solutions - The ChipAnalyser
- Biological Application
- Summary
- Future prospects

Data Evaluation and Processing

- Hybridisation successful ?
 - signal/noise of "positive control" $>$ Threshold_{PC}
(Loy et al., AEM, 2002, 2005)
- Hybridisation specific ?
 - signal/noise of "negative control" $<$ Threshold_{NC}
(Peplies, AEM, 2006; Loy et al., AEM, 2002, 2005)
 - individual mismatch probe ? (Wilson et al, Mol Cell Probes, 2002; AEM 2004; DeSantis et al, FEMS Microb Lett, 2004)

Data Evaluation and Processing II

- Uniform hybridisation ? Replicates and Outliers
 - coefficient of variation (CV)
 - median filtering (Cherkasova et al., PNAS 2003)
 - Deviation threshold from mean signal
- Normalisation
 - high signal sample (Liu et al., Env. Microbiol, 2001)
 - total signal (Bodrossy et al., Env. Microbiol. 2003)

Corrected Signal / Signal to Noise

- Background Correction (Francois et al, Cell Mol Probes 2006)

$$CS_{PB} = F635_{PB} - B635_{PB}$$

- Signal to Noise (Francois et al, J Microb Meth 2001; Oh et al., J Clin Microbiol 2004)

$$SNR_{PB} = F635_{PB} / B635_{PB}$$

- Corrected Signal to Noise (Loy et al, AEM 2002)

$$CSNR_{PB} = (F635_{PB} - (F635_{NS} - B635_{NS})) / B635_{PB}$$

- Signal to Noise - GenePix (Murray et al, PNAS 2001)

$$SNR-GP = (F635_{PB} - B635_{PB}) / B635 \text{ StandDev}_{PB}$$

$F635_{PB}$ = 635nm Fluorescence Pixel Intensity of Probe PB
 $B635_{PB}$ = 635nm Background Pixel Intensity of Probe PB
 $B635\text{StandDev}_{PB}$ = Standard Deviation of Mean Background Pixel Intensity
 SNR_{PB} = Signal to Ratio of Probe PB
 $CSNR_{PB}$ = Corrected Signal to Noise Ratio of Probe PB

When is a signal positive? – The Threshold

- SNR: 1.5 (Francois et al, J Microb Meth 2001; Oh et al., J Clin Microbiol 2004)
- CSNR: 2.0 (Loy et al, AEM 2002)
- SNR-GP: 2 (Murray et al, PNAS 2001)
- $SI(PR) > 3x SI(NC)$ (Kakinuma et al., BIOTECHN AND BIOENG, 2001)

Above threshold → include analysis

→ **Below or equal to threshold → exclude from analysis**

And then ?

Outline

- Introduction
- General and Particular Problems
- Analysis Methods and Algorithms
- **The ChipAnalyser - Features and Solutions**
- Biological Application
- Summary
- Future prospects

Basic Design ChipAnalyser

- Data environment for phylochip data
- Graphical user interface
- Various functions and analysis methods
- Interfaces to foreign software applications
- Proper data representation
- Error logging

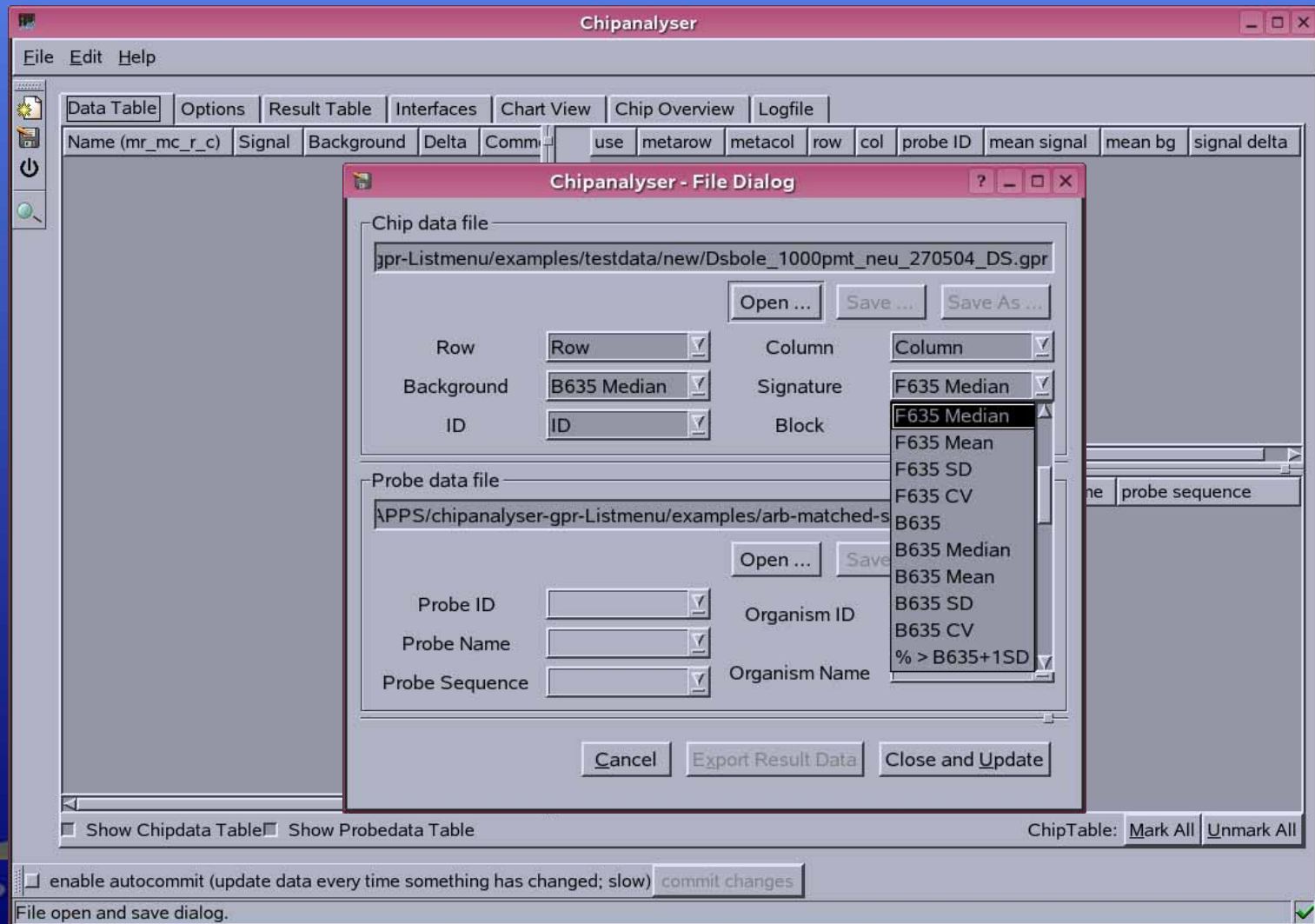
Technical Features

- Platform independency (Windows + Linux)
- Runtime optimisation
 - Structured internal data management (tripartite graph)
 - Adapted search algorithms
 - Processing of large data files

Algorithm Chipanalyser

- Open (Raw) Data from Image Analysis
- Select the data for analysis
- Select method + Set parameters for calculation
- Prove positive/negative control
- Calculate values
- Translate into biological meaning
- Display results

Open Raw Data File



DataTable and Visualisation

The screenshot displays the ChipAnalysier software interface, which is used for analyzing microarray data. The main window is titled "ChipAnalysier" and contains several panes:

- Data Table:** A table with columns for Name (nr, mc, f, c), use, metarow, metacol, row, col, probe ID, mean signal, mean bg, signal delta, and comment. The table lists various probes and their associated signal values.
- Chart View:** A line graph showing signal intensity across different probes. The x-axis represents probe positions, and the y-axis represents signal intensity. The chart shows several peaks, indicating high signal for specific probes.
- Probe Visualization:** A grid of colored dots representing individual probes. The dots are colored based on their signal intensity, with yellow and red indicating higher signal and blue indicating lower signal.

The interface also includes a menu bar (File, Edit, Help) and a toolbar with various icons for navigation and analysis. The status bar at the bottom provides information about the current probe and signal values.

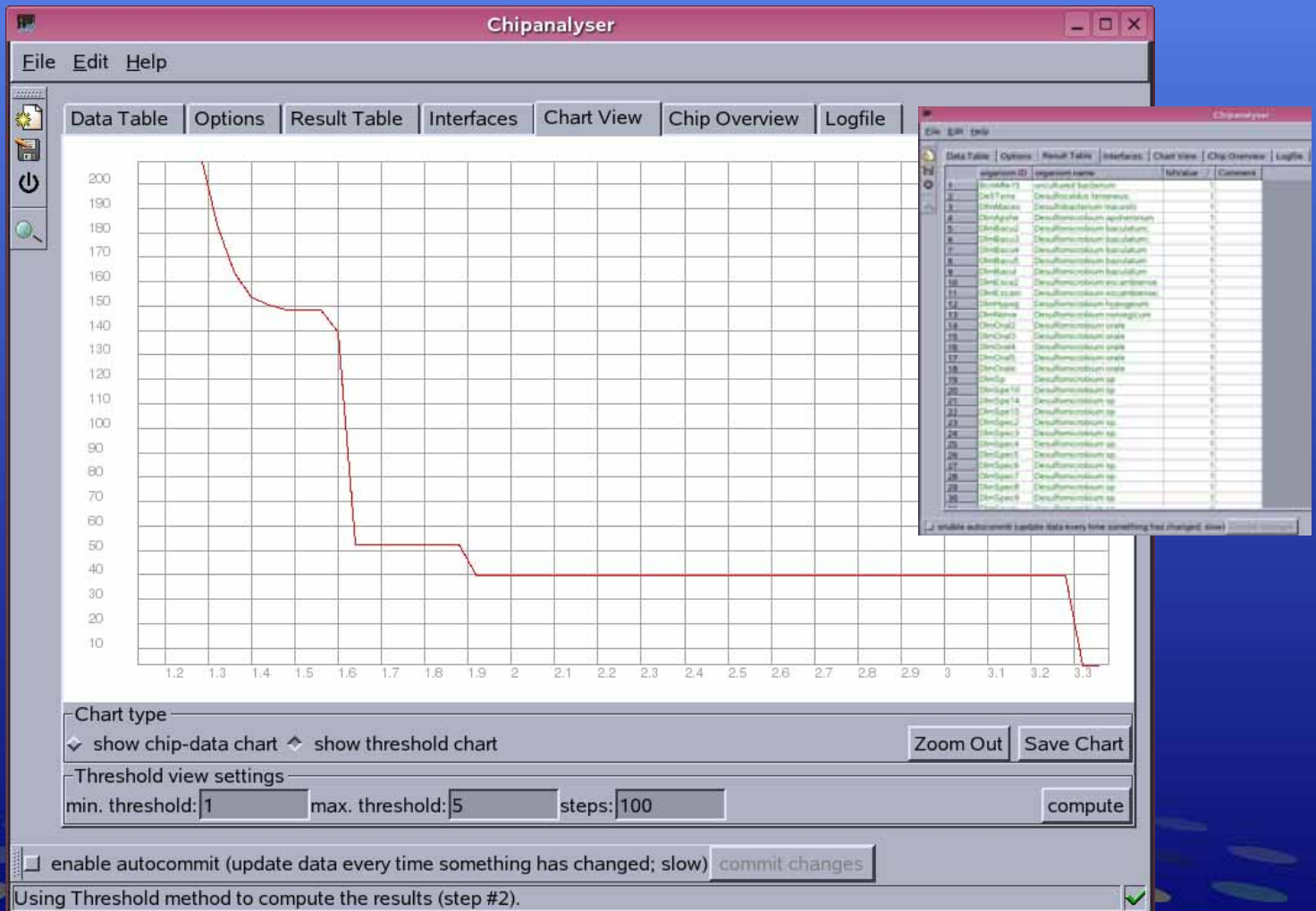
"preThreshold" based computation method

- Signal/Noise (or CSNR or SNR-GP)
- Threshold determines status of each spot (individual / mean / median filtering)
- Probes are weighted equally
- Result: Clear statement (positive / negative)

"preThreshold" based computation method

- Problems:
- Proper Threshold setting ("Weak" Signal/Noise!)
- One negative probe → gene/organism negative
- Group specific probes → induce "False positives"

GUI: Threshold graph



Hit Value Significance Method (postThreshold)

- Signal intensities
- Normalised by number of matched organisms
- Species specific probes > group specific probes
- Normalization → Number of relevant probes
- Result:
 - comparable hit value for each organism
 - Fault tolerance (→ group specific probes)

Properties of "postThreshold" method

- Hit value for each target gene/organism
- All signal values and probe specificities acknowledged
- Final evaluation of results → user
(Sets threshold for hit value)
- Higher fault tolerance (?)
- Full probe list required
- Error for saturated signals

Result Representation

Chipanalyser

File Edit Help

Data Table Options Result Table Interfaces Chart View Chip Overview Logfile

	organism ID	organism name	hitValue ▲	# relevant probes
1	DfmCeton	Desulfobacterium cetonicum	114.185	5
2	DetProt6	delta proteobacterium	114.185	5
3	DlmEzca2	Desulfomicrobium escambiense	97.8549	4
4	DfsVaria	Desulfosarcina variabilis	95.3521	6
5	DfsVari2	Desulfosarcina variabilis	95.3521	6
6	DlmSpec6	Desulfomicrobium sp.	94.916	4
7	DlmSpeci	Desulfomicrobium sp.	78.7834	5
8	DlmSpec5	Desulfomicrobium sp.	78.7834	5
9	DlmSpec4	Desulfomicrobium sp.	78.7834	5
10	DlmSpec3	Desulfomicrobium sp.	78.7834	5
11	DlmSpec2	Desulfomicrobium sp.	78.7834	5
12	DlmBacul	Desulfomicrobium baculatum		
13	DlmSpe14	Desulfomicrobium sp		
14	DlmSpec7	Desulfomicrobium sp.		
15	UncDe264	uncultured delta		
16	UncDe260	uncultured delta		
17	UncDe254	uncultured delta		
18	UncDe248	uncultured delta		
19	UncDe245	uncultured delta		
20	UncDe244	uncultured delta		
21	DlmSpec8	Desulfomicrobium sp		
22	DlmOrale	Desulfomicrobium orale		
23	DlmOral5	Desulfomicrobium orale		
24	DlmOral4	Desulfomicrobium orale		
25	DlmOral3	Desulfomicrobium orale		
26	DlmOral2	Desulfomicrobium orale		
27	DlmSpec9	Desulfomicrobium sp		
28	DlmSpe15	Desulfomicrobium sp		
29	DlmEscam	Desulfomicrobium escambiense;		
30	DsfDesul	Desulfovibrio desulfuricans		
31	DlmBacu5	Desulfomicrobium baculatum		

enable autocommit (update data every time something has changed; slow)

Chipanalyser

File Edit Help

Data Table Options Result Table Interfaces Chart View Chip Overview Logfile

```

[09:55:02] Probe C from the chiptable at position 2_2_1_24 is unreferenced.
[09:55:05] Results have been computed with 'Approximation' Method

[09:59:08] Results have been marked in the ARB
[10:15:02] Global CSNR Value modified.
[10:15:10] Threshold for negative probes modified.
[10:15:16] Threshold for negative probes modified.
[10:15:39] Threshold for negative probes modified.
[10:16:01] Threshold for negative probes modified.
[10:16:03] Threshold for negative probes modified.
[10:16:06] Threshold for negative probes modified.
[10:16:41] Threshold for negative probes modified.
[10:17:40] Approximation border-value modified.
[10:19:02] Approximation border-value modified.
[10:19:28] Probe C from the chiptable at position 2_2_1_24 is unreferenced.
[10:19:30] Results have been computed with 'Approximation' Method

[10:20:56] Results have been marked in the ARB
[10:33:17] Probe C from the chiptable at position 2_2_1_24 is unreferenced.
[10:33:21] Results have been computed with 'Threshold' method

[10:37:27] Probe C from the chiptable at position 2_2_1_24 is unreferenced.
[10:37:29] Results have been computed with 'Threshold' method

[10:38:50] Results have been marked in the ARB
[10:39:08] Results have been marked in the ARB
[10:42:38] Probe C from the chiptable at position 2_2_1_24 is unreferenced.
[10:42:40] Results have been computed with 'Threshold' method
    
```

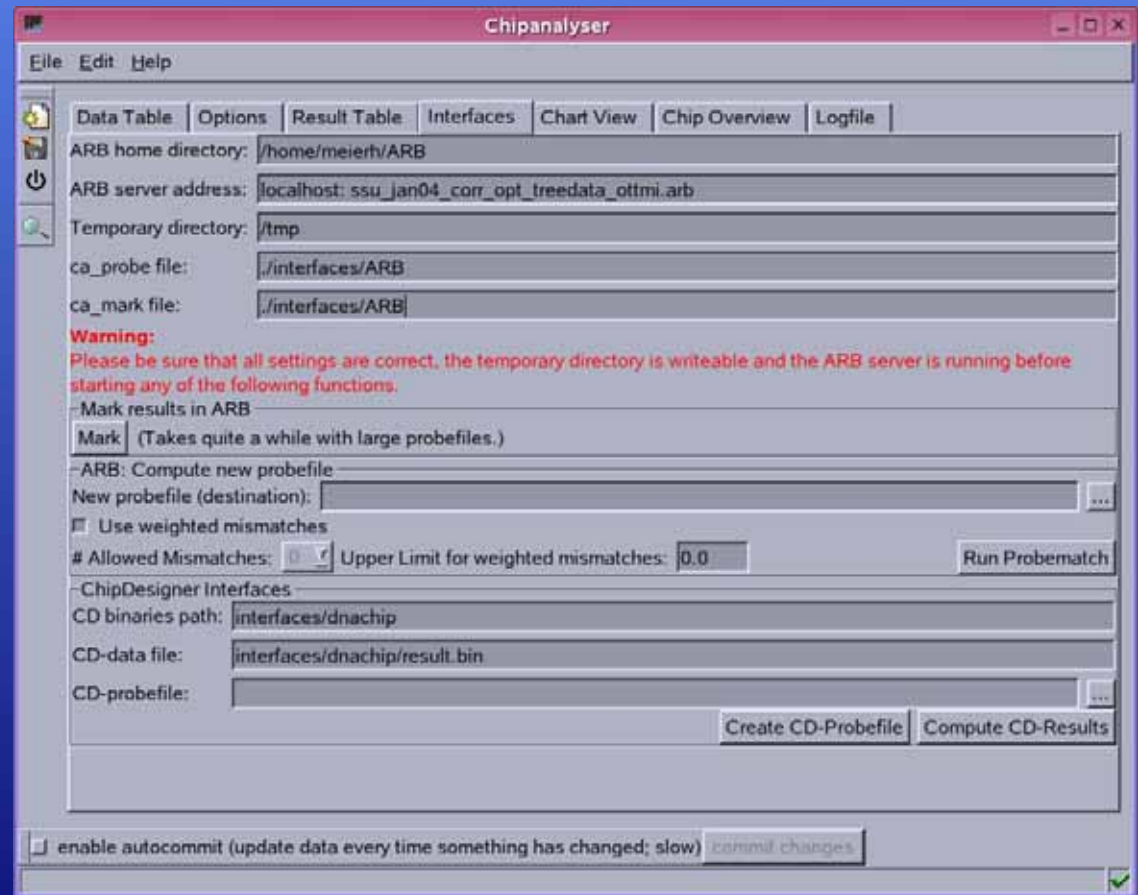
enable autocommit (update data every time something has changed; slow)

save Log clear Log

Using Threshold method to compute the results (step #2).

ARB-Interface

- Visualise results in ARB
- Create/Update probe specificity



Chipanalyser Results visualised in ARB

The image shows a screenshot of the ARB (Arb Research Bioinformatics) software interface. The main window displays a phylogenetic tree with various bacterial species names, including *Desulfonatronum* and *Desulfomicrobium*. A search window titled "SEARCH and QUERY" is open, showing search criteria and a list of 44 hits. A "Species Selections" window is also open, displaying a list of selected species for further analysis.

ARB_NT

File Species Sequence SBI Probes Tree Tools Properties

corr_opt_treedata_ tree_1000_jan05 ali_16s DlnApshe HELP ? LESS

UNDO REDO SEARCH JUMP WWW Protect 0

44 marked

SEARCH and QUERY

More functions: More search HELP

CLOSE DATABASE SEARCH HELP

Search species that match the query

Add species that dont match the q.

Keep species that are marked

QUERY

Search fields Search strings

name = *

name = *

name = *

SEARCH

Hits: 44

HITLIST

MARK LISTED UNMARK REST

DFsVar:12 :DFsVar:12

DfncCetlon :DfncCetlon

DetProt6 :DetProt6

DlnBacu4 :DlnBacu4

DlnBacu1 :DlnBacu1

DfncMaces :DfncMaces

DlnSpe15 :DlnSpe15

DlnSpe3 :DlnSpe3

DlnApshe :DlnApshe

DlnSpe14 :DlnSpe14

UdnB1622 :UdnB1622

DlnNorve :DlnNorve

DlnSpe7 :DlnSpe7

DlnSpe2 :DlnSpe2

DlnSpe6 :DlnSpe6

DlnSpe4 :DlnSpe4

DlnSpeci :DlnSpeci

REFRESH Note: double click to mark/unmark.

Species Selections

CLOSE HELP

Selections

stap15

stap10

stap5

stap5_2

stap_Fin

toali

toali_pre

ganna_nov04

helix

vibr io

bergoy

dsbole_1000pnt_270504_hitval_th1_3

arb_dmtal_hitval_th2_nov_valum35

STORE

EXTRACT

MARK

UNMARK

INVERT

COMBINE

DELETE

RENAME

Create/Update Probe Specificity List

Probe Match

133 ARGLO972	CCC CGG TAA GCT TCC CGG
136 DSRHP185	CCA CCT TTC CTG TTT CCA
138 DSBACL143	TCG GGC AGT TAT CCC GGG
139 DSBACL225	GGT CCG CAA ACT CAT CTC



133 ARGLO972	UncArc27	uncultured archaeon	CCC CGG TAA GCT TCC CGG
133 ARGLO972	UncArc28	uncultured archaeon	CCC CGG TAA GCT TCC CGG
133 ARGLO972	UncArc29	uncultured archaeon	CCC CGG TAA GCT TCC CGG
133 ARGLO972	UncArc30	uncultured archaeon	CCC CGG TAA GCT TCC CGG
133 ARGLO972	UncArc38	uncultured archaeon	CCC CGG TAA GCT TCC CGG
133 ARGLO972	UncArc39	uncultured archaeon	CCC CGG TAA GCT TCC CGG
136 DSRHP185	DfhSinga	Desulforhopalus singaporensis	CCA CCT TTC CTG TTT CCA
136 DSRHP185	DfhSp	Desulforhopalus sp	CCA CCT TTC CTG TTT CCA
136 DSRHP185	DfhVacuo	Desulforhopalus vacuolatus	CCA CCT TTC CTG TTT CCA
138 DSBACL143	DslPheno	Desulfobacula phenolica	TCG GGC AGT TAT CCC GGG
138 DSBACL143	DslTolu2	Desulfobacula toluolica	TCG GGC AGT TAT CCC GGG
138 DSBACL143	DslToluo	Desulfobacula toluolica	TCG GGC AGT TAT CCC GGG
139 DSBACL225	DslPheno	Desulfobacula phenolica	GGT CCG CAA ACT CAT CTC
139 DSBACL225	DslTolu2	Desulfobacula toluolica	GGT CCG CAA ACT CAT CTC
139 DSBACL225	DslToluo	Desulfobacula toluolica	GGT CCG CAA ACT CAT CTC
139 DSBACL225	Unc12262	uncultured bacterium	GGT CCG CAA ACT CAT CTC
139 DSBACL225	UncDes21	uncultured Desulfobacula	GGT CCG CAA ACT CAT CTC



Microarray-based examination of tooth pocket samples of patients with adult periodontitis

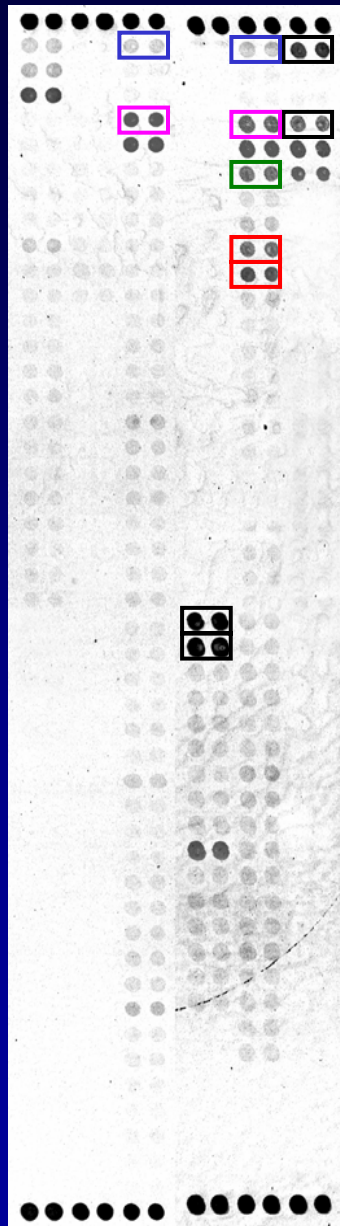
Loy et al., Appl. and Environ. Microbiol., 2002.

Microarray for detection and identification of sulfate reducing prokaryotes (SRP)



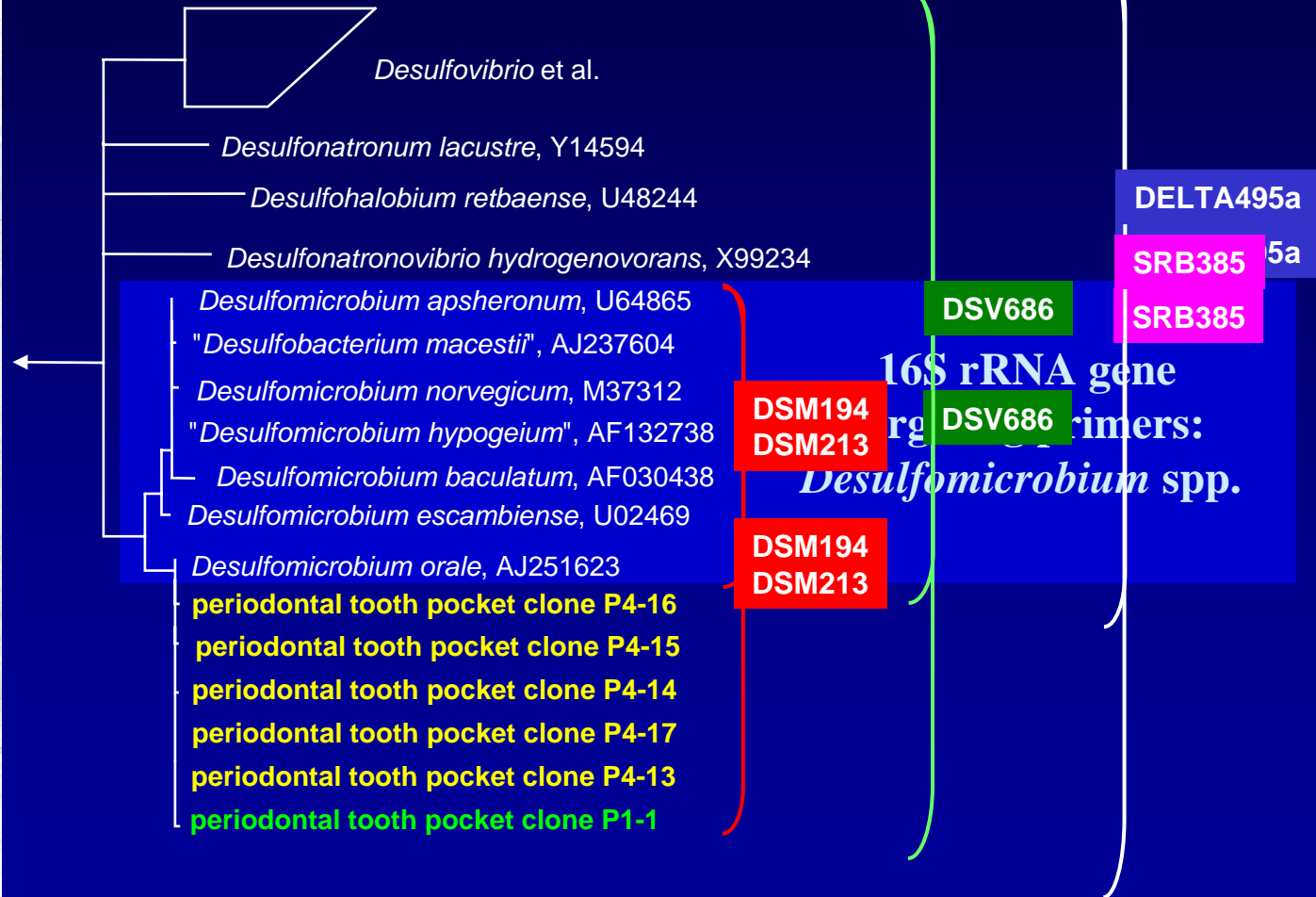
- 132 ssu-rRNA-targeted oligonucleotide probes
- Duplicates
- Positive/negative control probes
- Detection and differentiation capacity > 130 described species of SRB organisms
- Reverse hybridization of CY5-labelled PCR – amplicates of ssu-rRNA gene

Desulfomicrobium orale in Patients with Adult Periodontitis

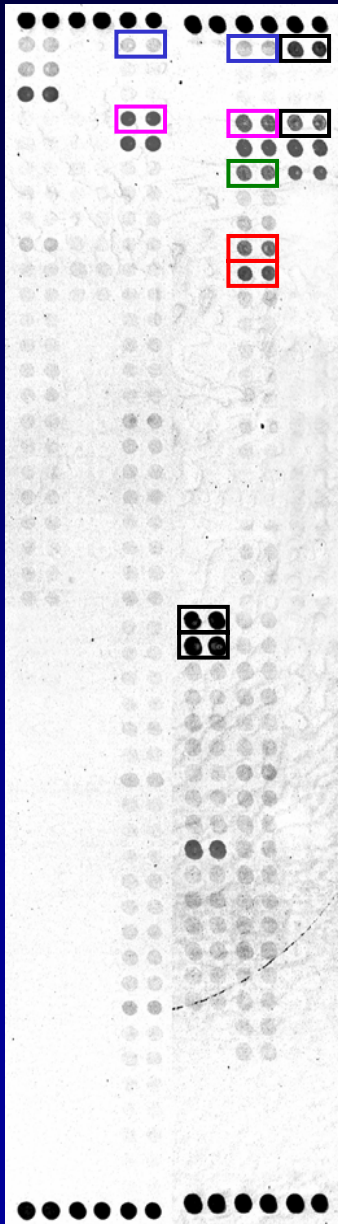


16S rRNA

10 %



Desulfomicrobium orale in Patients with Adult Periodontitis



Chipanalyser

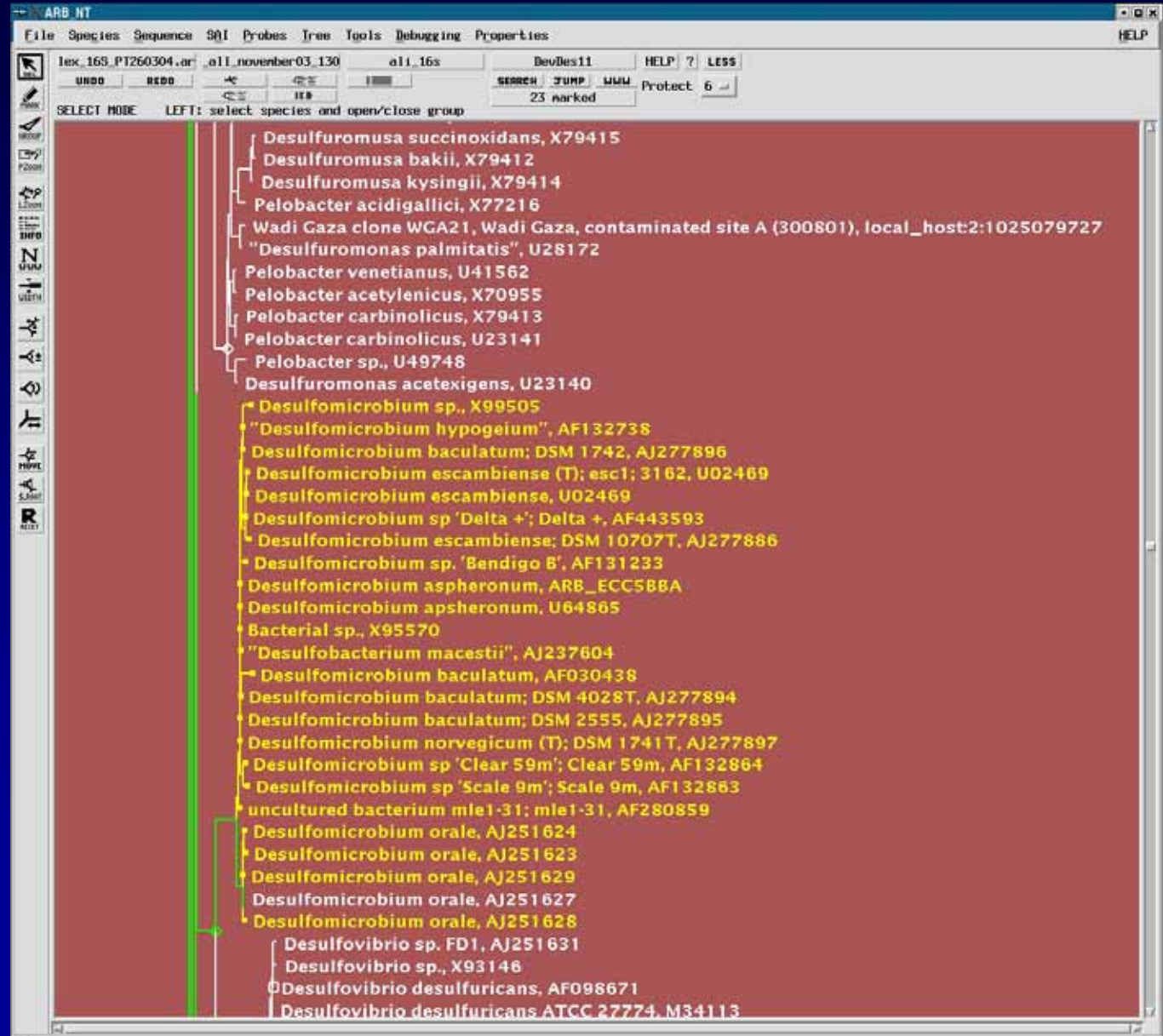
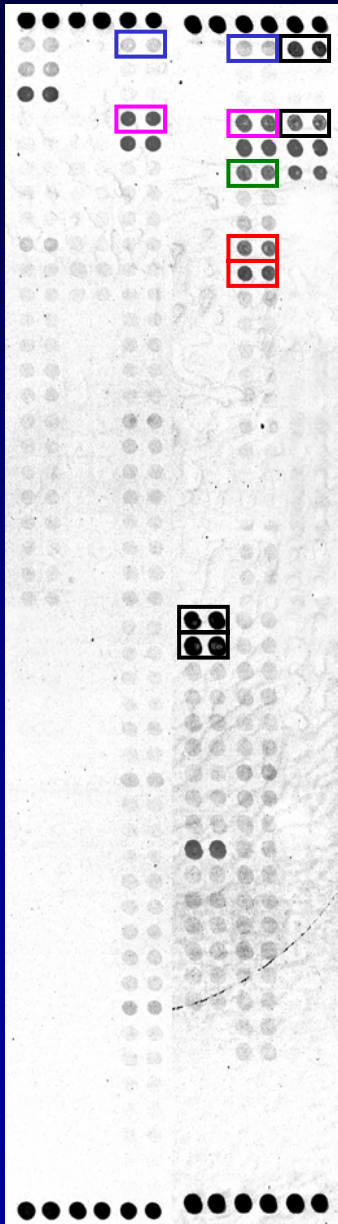
File Edit Help

Data Table Options Result Table Interfaces Chart View Chip Overview Logfile

	organism ID	organism name	hitValue	# relevant probes
1	S000016044	4 Desulfomicrobium escambiense	1	4
2	DeiEscam	Desulfomicrobium escambiense	1	4
3	DeiBacu3	Desulfomicrobium baculatum	0.789416	5
4	DeiSpeci	Desulfomicrobium sp. 'Bendigo'	0.771945	5
5	DeiOr	Desulfomicrobium sp.	0.689406	6
6	DeiOrali	Desulfomicrobium orale	0.689406	6
7	DeiOra7	Desulfomicrobium orale	0.689406	6
8	DeiOra6	Desulfomicrobium orale	0.689406	6
9	DeiOra2	Desulfomicrobium orale	0.689406	6
10	S000017587	7 Desulfomicrobium sp 'Clear 59'	0.672863	6
11	S000017586	6 Desulfomicrobium sp 'Scale 9m'	0.672863	6
12	S000008127	7 Desulfomicrobium escambiense:	0.662207	6
13	S000020725	5 Desulfomicrobium baculatum; D	0.592957	7
14	S000020193	3 Desulfomicrobium baculatum; D	0.592957	7
15	S000016904	4 Desulfomicrobium baculatum; D	0.592957	7
16	S000016859	9 uncultured bacterium mle1-31;	0.592957	7
17	S000005487	7 Desulfomicrobium norvegicum (0.592957	7
18	S000000541	1 Desulfomicrobium sp 'Delta +'	0.592957	7
19	DeoMaces	"Desulfobacterium macesti"	0.592957	7
20	DeiHypog	"Desulfomicrobium hypogeium"	0.592957	7
21	DeiAsphe	Desulfomicrobium asphaerum	0.592957	7
22	DeiAsphe	Desulfomicrobium asphaerum	0.592957	7
23	BacSpe131	Bacterial sp.	0.592957	7
24	S000088251	1 Desulfobacterium cetonicum (T	0.371189	5
25	S000015701	1 delta. proteobacterium oXyS1;	0.371189	5
26	DeoCeton	Desulfobacterium cetonicum	0.371189	5
27	1dProt6	delta. proteobacterium oXyS1	0.371189	5
28	S000014246	6 Desulfosarcina variabilis (T)	0.312592	6
29	S000000231	1 Desulfosarcina variabilis (T)	0.312592	6
30	DeaVaria	Desulfosarcina variabilis	0.312592	6
31	DeaVar2	Desulfosarcina variabilis	0.312592	6
32	DeiOra3	Desulfomicrobium orale	0.289151	6
33	VagFluvi	Vagococcus fluvialis	0.228125	2
34	StoUber2	Streptococcus uberis	0.228125	2
35	StoSang2	Streptococcus sanguinis	0.228125	2
36	StoSaliV	Streptococcus salivarius	0.228125	2
37	StoSali2	Streptococcus salivarius	0.228125	2
38	StoParaS	Streptococcus parasanguinis	0.228125	2
39	StoPara2	Streptococcus parauberis	0.228125	2
40	StoMutax	Streptococcus mutans	0.228125	2

enable autocommit (update data every time something has changed; slow)

Desulfomicrobium orale in Patients with Adult Periodontitis



Application and Evaluation

- “Pre-Threshold” result agrees with original result
- Result by “Post-Threshold” method indicates existence of another SRB group
- Pre-Threshold method works with “reduced” probe lists
- Factors that influence results by “Post-Threshold” method
 - quality of *in silico* probe specificity list
 - degree of saturation of spot signals

Summary

- Software for analysis of phylochip data from community analysis
 - Extensive GUI (data handling/editing, visualization)
 - Various analysis methods
 - Supports threshold setting
 - Result export
- ARB-Interface (chip result tree, probe list update)
- Biological example study showed in principal applicability of both methods



Future prospects

- Extension by data types
- Extension of the results table
- Extension for data evaluation (control function)
- Evaluation/optimization of analysis algorithms

Thank you!

- Kai Bader, Wolfgang Thomas, Technische Universität München, Informatik
- Wolfgang Ludwig, Technische Universität München, Lehrstuhl für Mikrobiologie
- Alex Loy, Michael Wagner, University of Vienna, Department of Microbial Ecology



Harald Meier
meierh@in.tum.de
www.lrr.in.tum.de/~meierh