

Analysis of a Diagnostic
Microarray:
The Burkholderia Phylochip

Dr. Hubert Rehrauer

Outline

Introduction

Challenges

Data analysis

Optimization details

Outlook & summary

Introduction

Burkholderia

The genus *Burkholderia* are widely distributed beta-Proteobacteria ...

- hospitals, industrial sites
- rhizosphere, soils
- animals, water

and are important as

- human and plant pathogens
- nitrogen fixation and other symbiotic plant interactions
- soil remediation bacteria

Introduction

Diagnostic Microarrays

Microarrays

- enable the fast parallel quantification of thousands of RNA or DNA sequences
- are cost-effective
- are reliable (if applied under well controlled conditions)

Introduction

The *Burkholderia* Phylochip

The Burkholderia Phylochip is a taxonomic microarray based on the 16S rRNA

It contains 160 probes (18-mer oligos) matching *Burkholderia* and *Pandorea* at taxonomic levels ranging from higher taxa to individual Burkholderia species

Applications

- detection of pathogenic, beneficial or contaminating Burkholderia species
- assessment of *Burkholderia* diversity

Introduction

Project Steps

- Probe design
- Optimization of hybridization conditions
 - hybridization temperature
 - washing temperature
 - ...
- Validation with 34 *Burkholderia* typestrains + 8 outgroups
 - identifiability of individual strains/species/taxa
 - reliability of individual probes
- Species detection from natural samples
- Differential analysis of natural samples

Challenges

Probe Design

Probe design:

- specificity (no cross-hyb)
- sensitivity (good binding)
- hybridization properties (same thermodynamic characteristics)

Recent publications suggest that

- **in-silico prediction of hybridization properties is unreliable**

Challenges

Limited Signal Range

Applications with different requirements

- Strain detection → maximize sensitivity
- Quantitative biodiversity → signal linearity

Usable signal range:

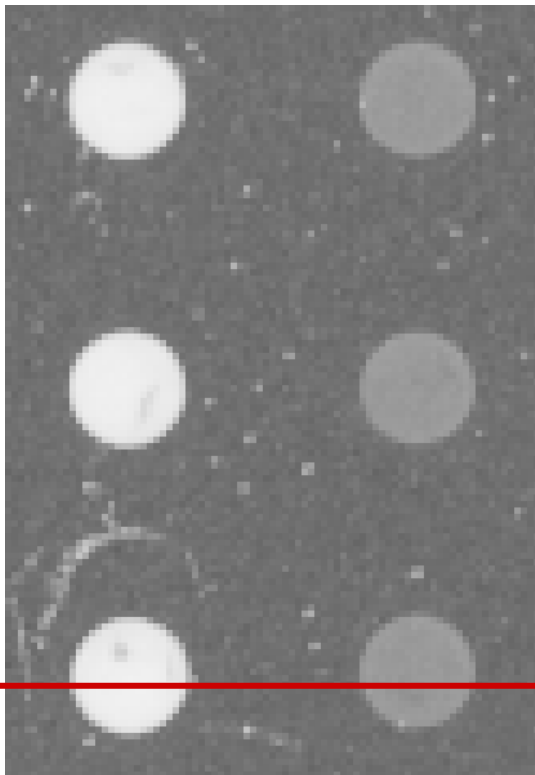
- Max spot intensity – background: 60 000 – 200
 - Dynamic range: $60\,000/200 = 300$
- If signal is linear we can only detect strains with relative concentration $> 1/300$

Challenges

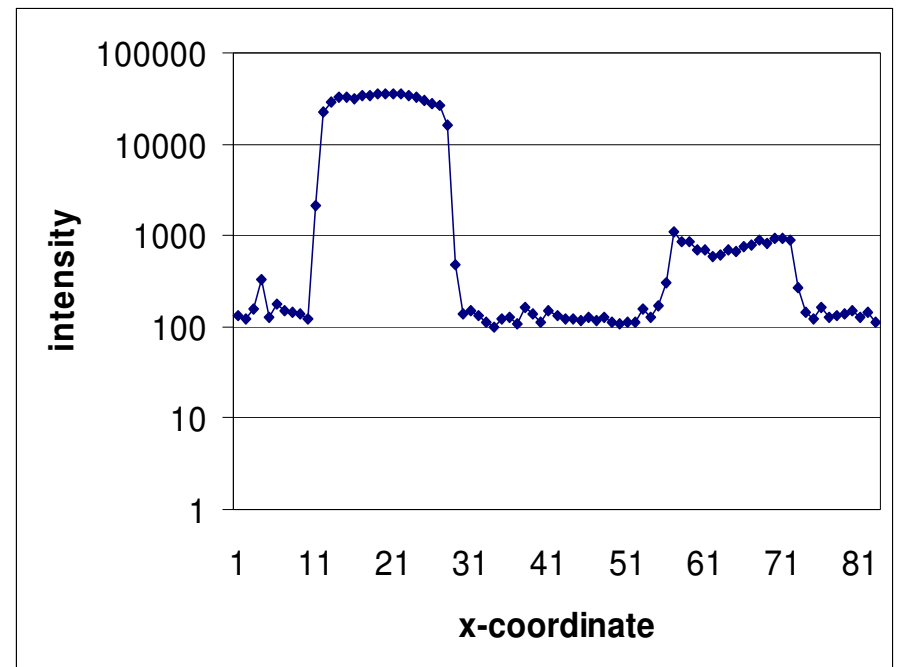
Limited Signal Range (ct'd)

Example:

Part of the array



Intensities at the red line



Challenges

Evolving Databases & Strains

The number of 16S rRNA sequences increases

New bacterial strains are discovered

Some bacterial strains are redundant

Many bacterial strains are not culturable

Data Analysis

Agilent Scanner

→ Image

GeneSpotter

→ Spot intensities and QC information

Preprocessing (R Scripts)

→ Relative abundance per probe

Strain & probe performance (R Scripts)

→ Identifiability of strains

→ Reliability of probes

Chip optimization (R Scripts)

→ Eliminate non-working probes

Data Analysis

GeneSpotter - QC

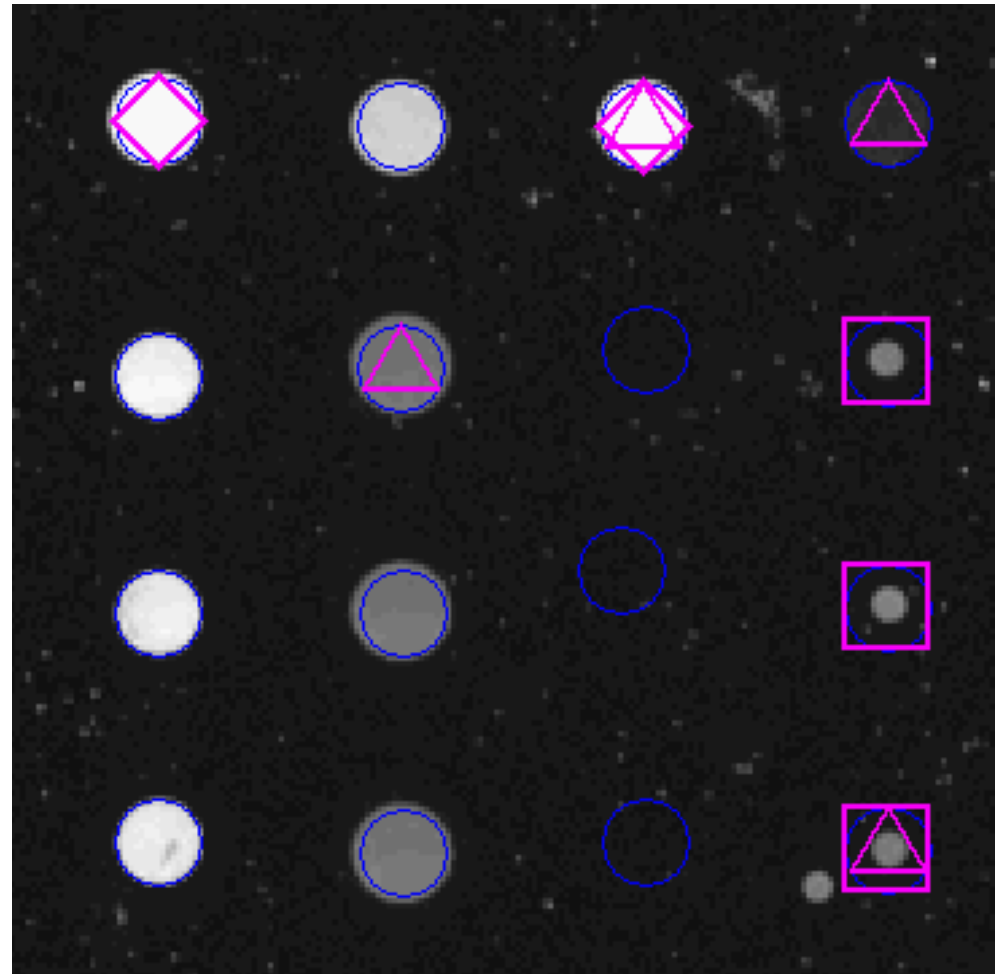
○ Spot outline; median of inner pixels is the **spot intensity**

◇ Saturated

△ Artefact

□ Poor Shape

Marked spots are excluded



Data Analysis

Quantification & Normalization

Raw spot intensity:

$$\text{Signal} = \text{Median spot intensity} - \text{background}$$

Normalized spot intensity:

$$\text{rel. abundance} = \text{Signal} / \text{Signal}(\text{EUB338I})$$

We use EUB338I as a normalization probe because it is predicted to match on all strains (see Sanguin et al. 2005; used EUB342)

Data Analysis

Performance Assessment

Performance of the array is assessed by comparing the measured abundances to ARB's **Probematchtable** (contains weighted mismatches)

Probes with no mismatches in a strain should give high signal.

Sequence mismatches of probes and strains

	Cn4Burk2	S000395087	BrkSp232	BrkSpe57	BrkSp168	BrkSp167	UncEu229	BrkSpe38	BrkSpe56
Probe 1	0	0	2	2	2	2	2	2	2
Probe 5	3.8	3.8	3.6	3.6	3.6	3.8	4.5	4.5	4.5
Probe 48	0.2	0.2	0	0	0	0	1.3	0	0
Probe 49	3.3	3.3	4.5	4.5	4.5	4.5	4.5	1.2	1.2
Probe 50	4	2.3	4.5	4.5	4.5	4.3	4.3	3.1	3.1
BETA1	0	0	0	0	0	4.5	0	0	0
EUB338	0	0	0	0	0	0	0	0	0
EUB338II	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9	2.9
EUB338III	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1	2.1
EUB342	0	0	0	0	0	0	0	0	0
MALLEI2	4.5	4.5	4.5	4.5	4.5	4.5	4.5	3.4	3.4
UNIV1389A	0	0	0	0	0	0	0	0	0
UNIV1389E	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
UNIV1389C	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

■ ■ ■ 1048 more strains

■ ■ ■ 146 more probes

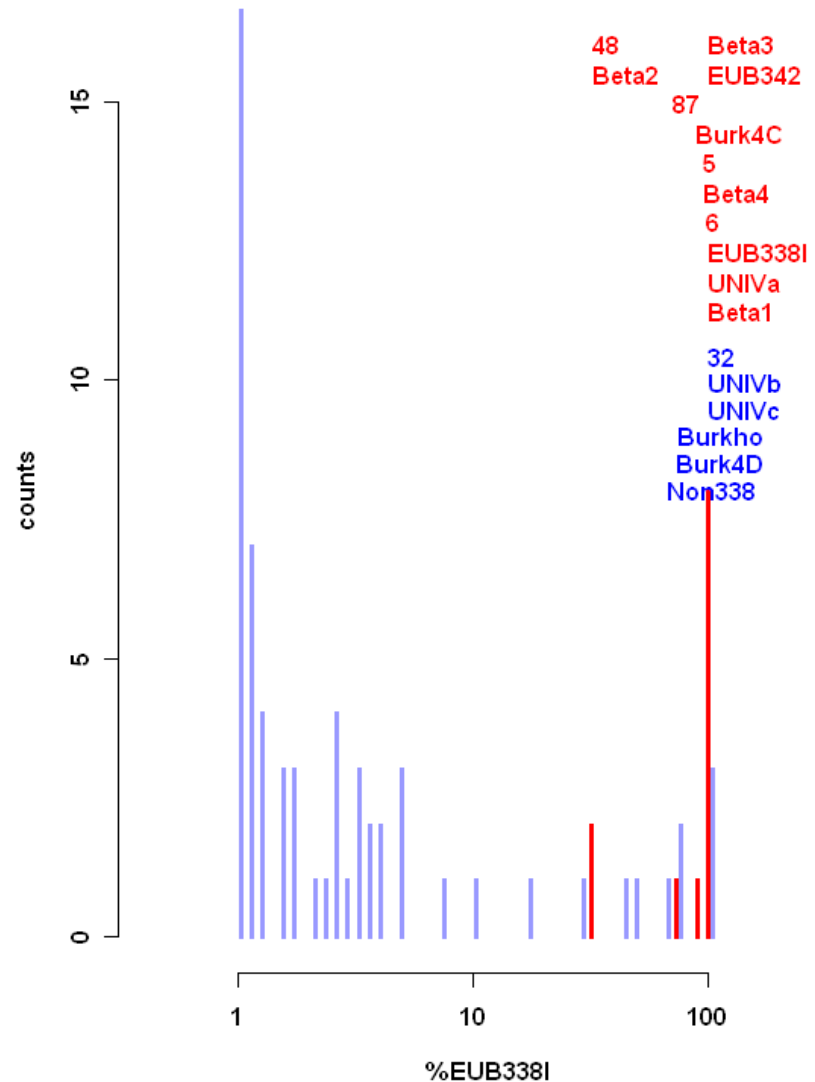
Data Analysis

Strain Identifiability

Histogram of abundances for Strain 1257.

Matching probes have high signal and are indicated by name.

Some **mismatching probes** yield more than 50% EUB; also indicated by name.

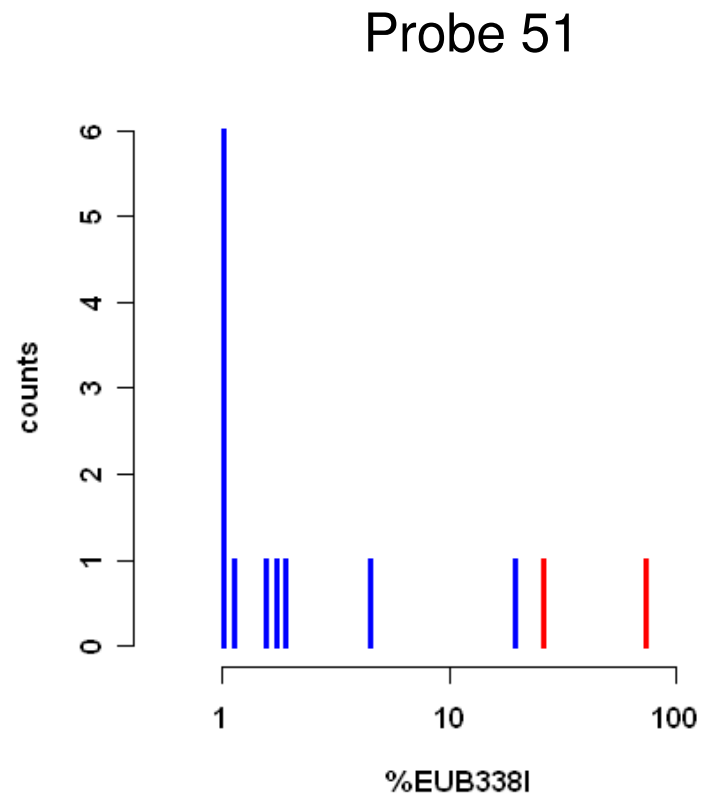


Data Analysis

Probe Performance

Probes should have high signal in **perfect matching strains** and low signal in **non-matching strains**.

Example: Relative abundance of Probe 51 measured on 2 matching strains and 12 non-matching strains.



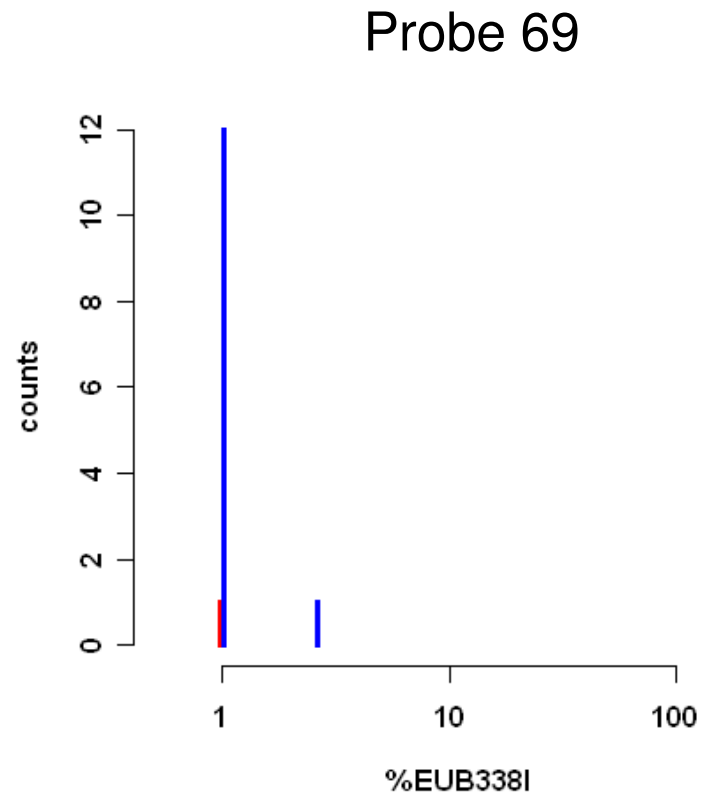
Data Analysis

Chip Optimization

Measure all available 34
typestrains and 8 outgroups
+ environmental samples

Iteratively eliminate non-working
probes

... work in progress



Optimization Details

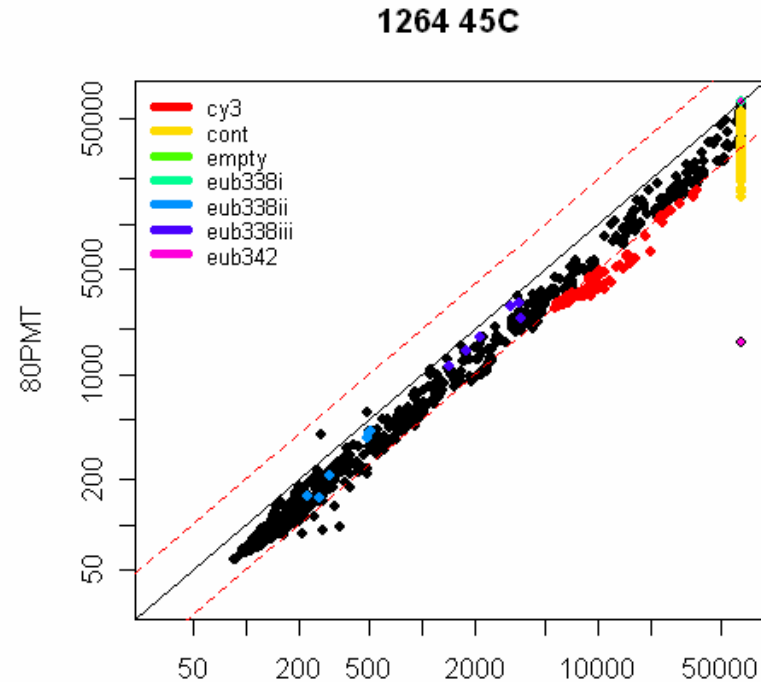
Scanner Gain

Rescanning the array with lower scanner gain

- shows good reproducibility of data processing
- reveals saturation effects

→ If saturation is observed, arrays are rescanned with lower gain

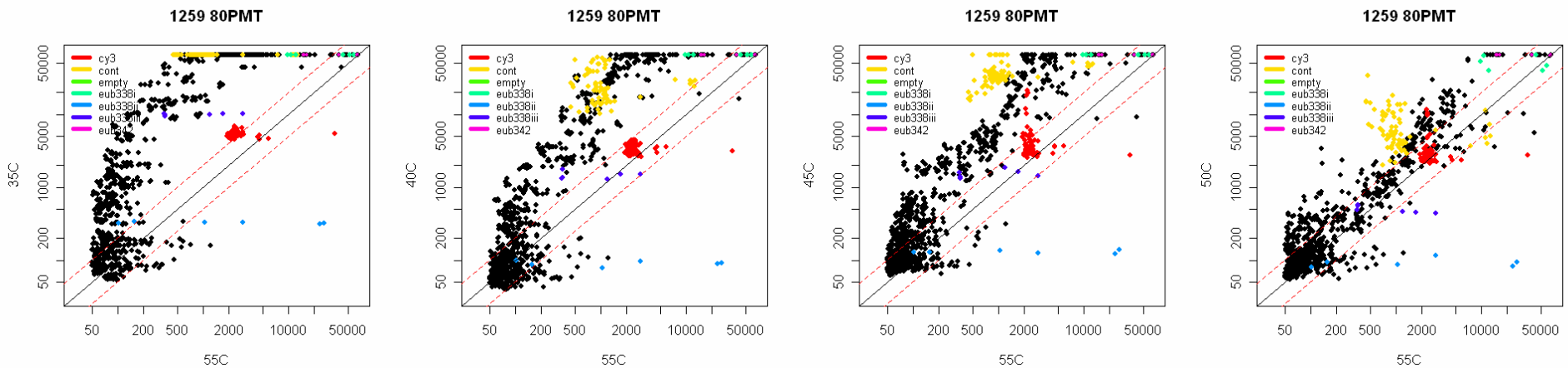
→ Use dilution series to identify optimal dye concentration



Optimization Details

Hybridization Temperature

At hybridization temperatures below 45C many probes are saturated independent whether they match or not.

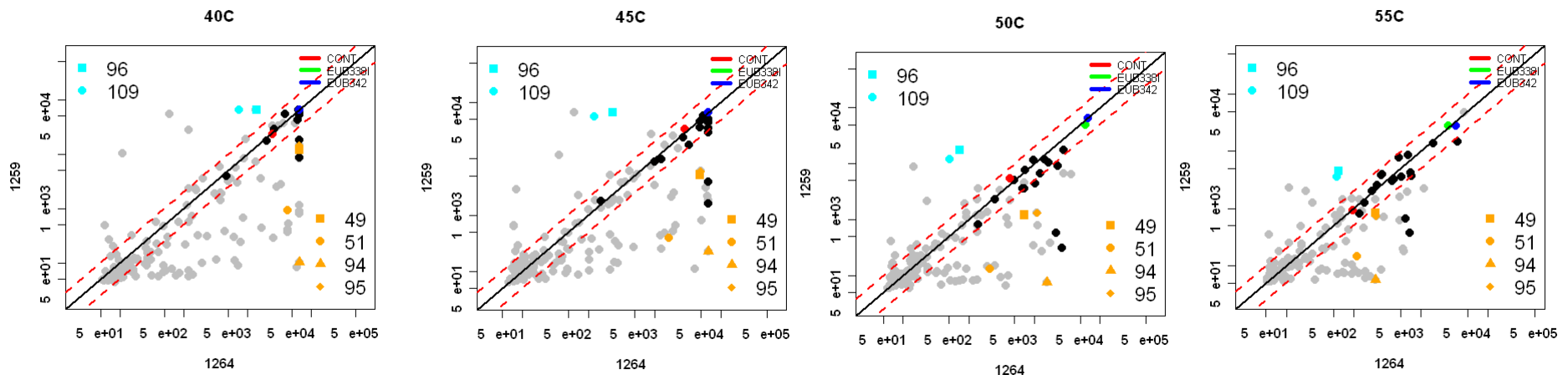


Optimization Details

Hybridization Temperature (ct'd)

Strain comparison at different hybridization temperatures.

- Strains can be discriminated at all temperatures
- Orange probes are specific for Strain 1264
- Cyan probes are specific for Strain 1259
- Probes 51 & 109 discriminate only by 1 mismatch
- Greatest discrimination is obtained by 45C and 50C

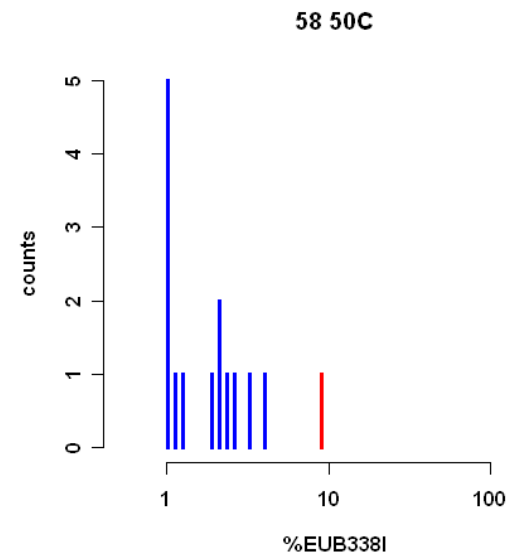
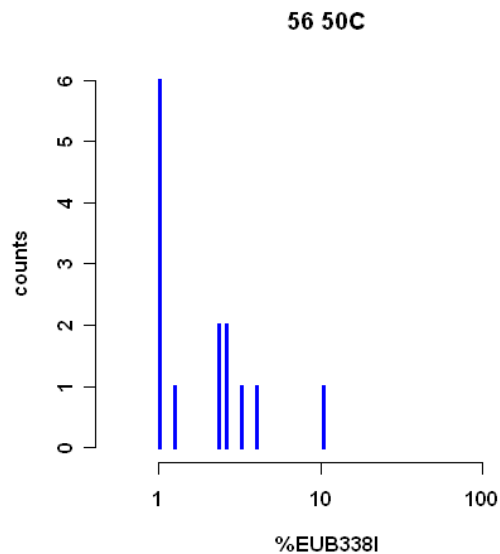


Preliminary Results

Probe-dependent Threshold

Example:

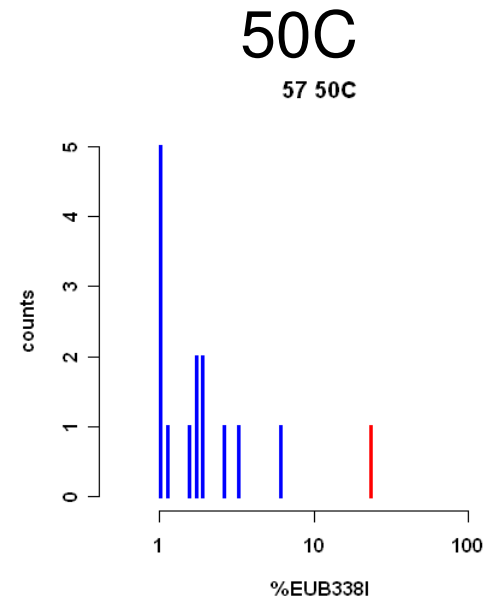
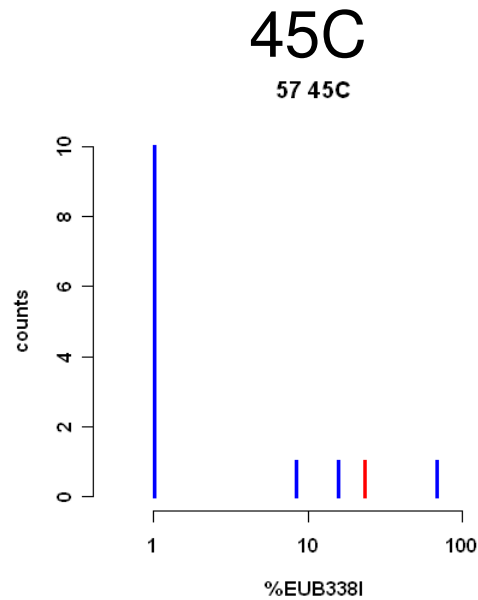
- Probe 56: highest false positive has 10%EUB
- Probe 58: true positive is below 10%EUB



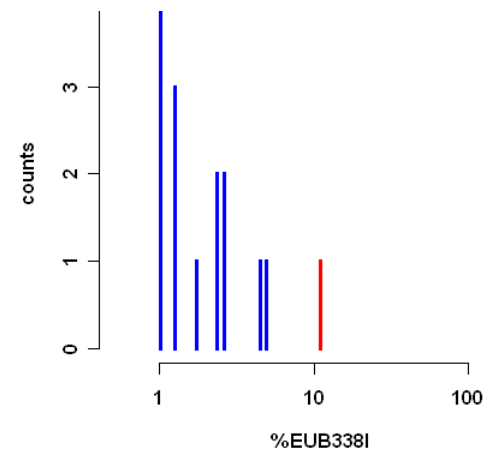
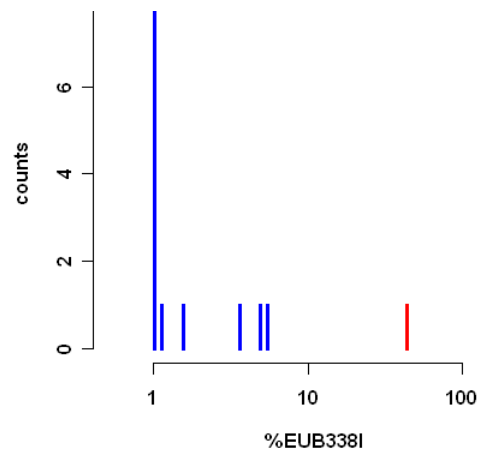
Preliminary Result

Probe-dependent Optimal Hyb. Temperature

Probe 57



Probe 59



Outlook

- Hybridize the remaining strains and eliminate non-working probes
- Strain mixture study:
 - Validate signal linearity
 - Assess detection limit
- Classification of unknown strains
- Differential bacterial biodiversity of natural samples from normal and polluted forest soil

Summary

- Designed *Burkholderia* Phylochip and demonstrated its function
- After elimination of faulty probes, the Phylochip can discriminate all 14 strains tested so far
- Probe performance depends on hybridization temperature; there is no single optimal hybridization temperature for all probes
- Optimal measurement conditions for strain detection and strain quantification still to be assessed

Acknowledgements

University Zurich – Institute of Plant Biology – Department
of Microbiology:

Susan Schönmann

Prof. Leo Eberl

Functional Genomics Center Zurich:

Catherine Aquino

Dr. Jens Sobek