

# **NEW ALGORITHMS DEVELOPMENT FOR MICROARRAY PROBE DESIGN TO IDENTIFY MICROORGANISMS AND CHARACTERIZE METABOLIC PATHWAYS .**

**P. PEYRET.**

**[pierre.peyret@univ-bpclermont.fr](mailto:pierre.peyret@univ-bpclermont.fr)**



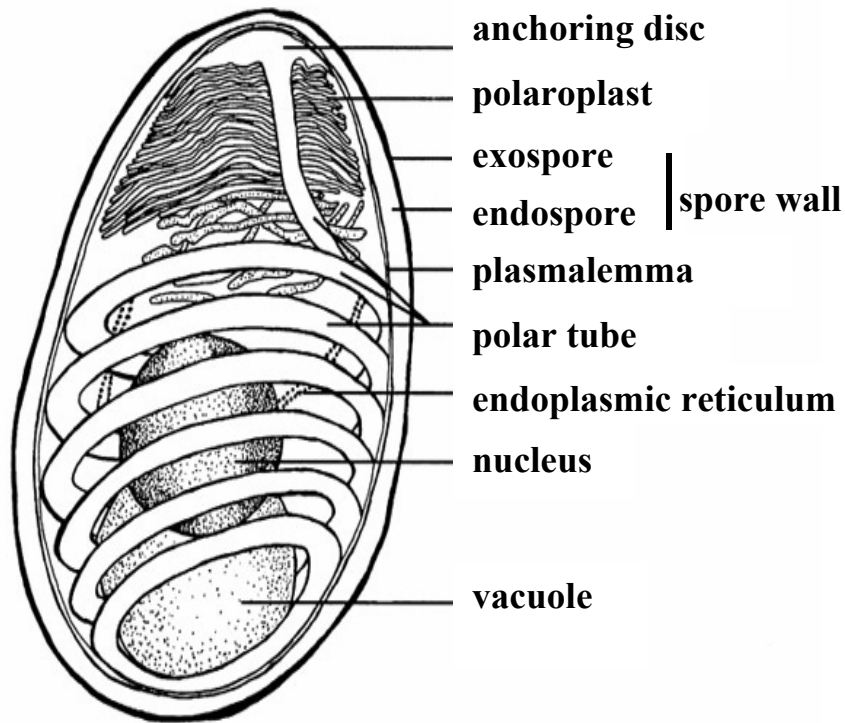
# Presentation

- **Biological Models**
- **Microarray Developments**
- **New Algorithms development**
- **Biological validation**
- **Conclusion and prospects**



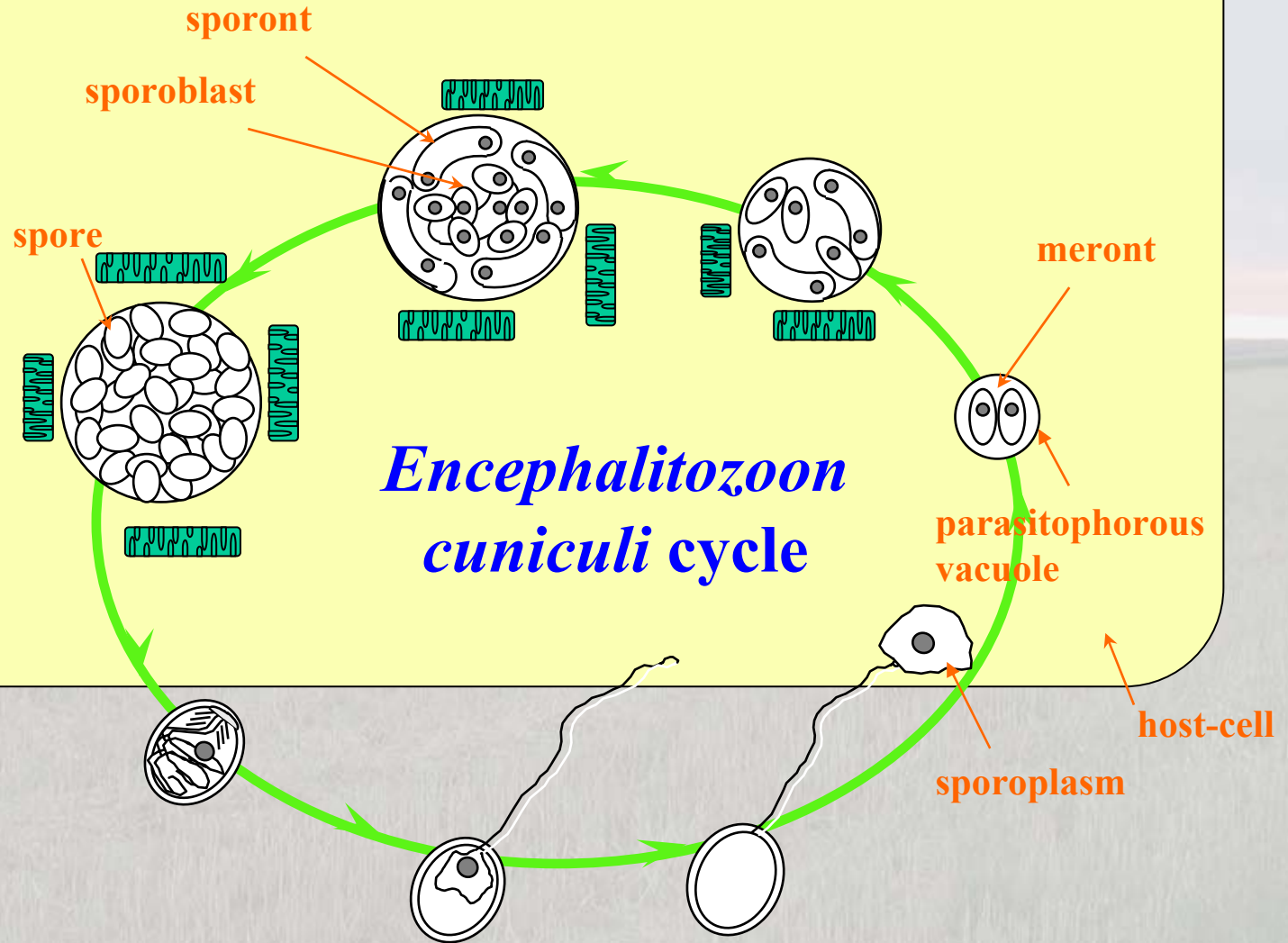
# Biological Models

# Microsporidia

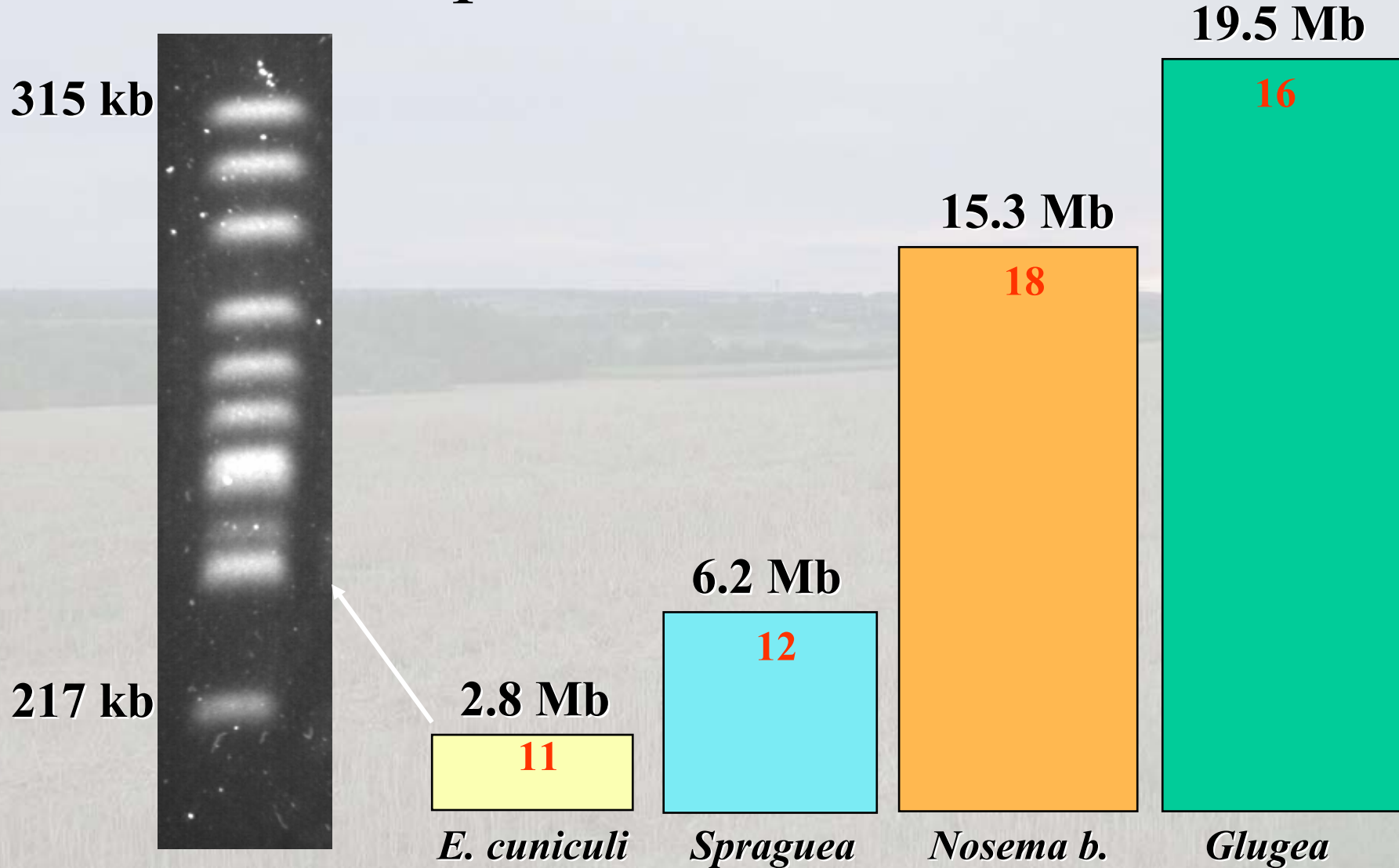


- Eukaryotic parasites (1200 species)
- Obligate intracellular parasite
- Original invasion of the host cell
- Pathogen for all the animal phylum (from man to mollusc and even sporozoa)

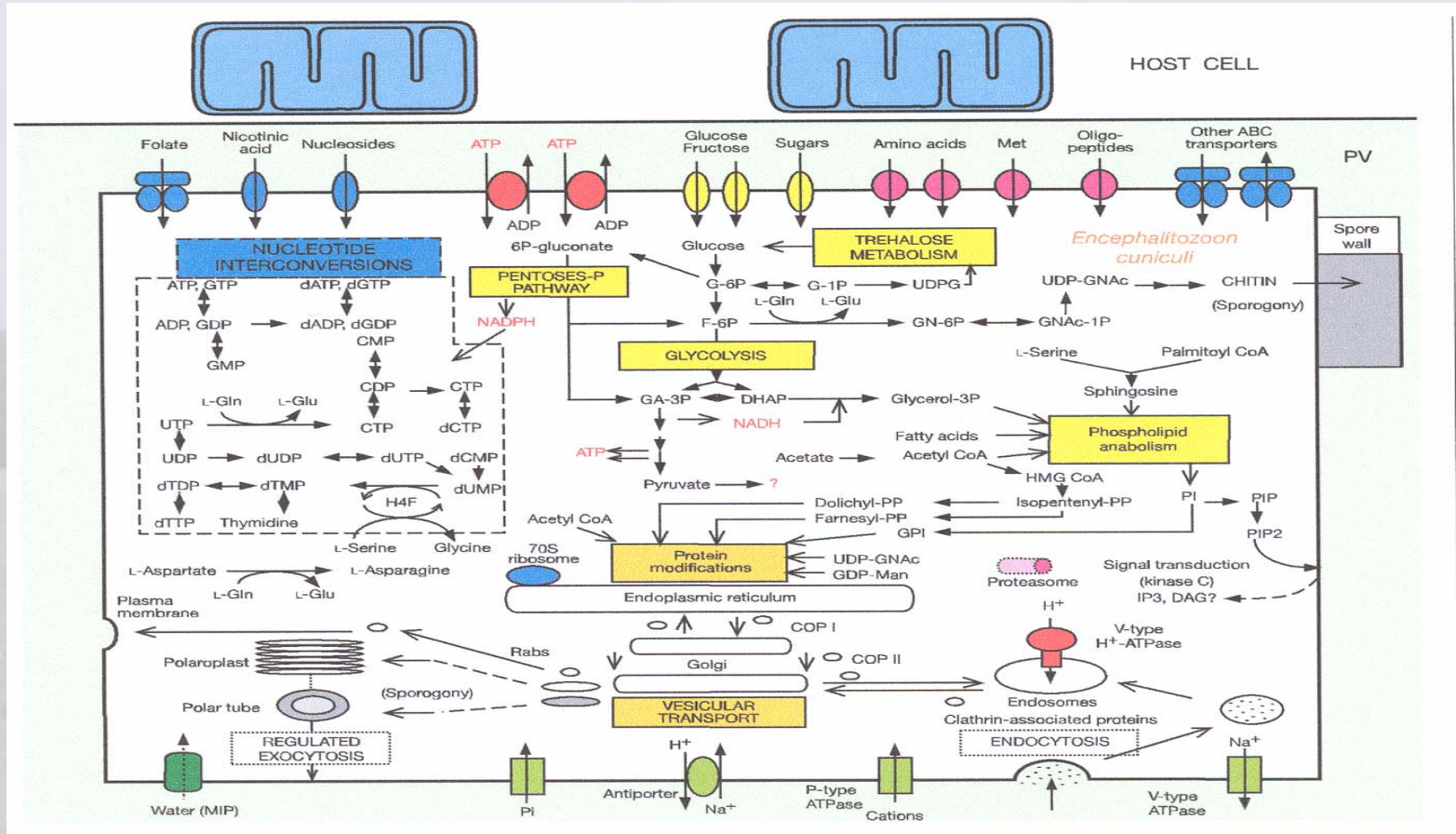
# Microsporidian Life Cycle



# Microsporidian Genome Size



# Complete Genome Sequence and Metabolic Reconstruction



Peyret et al., 2001, Genome Research

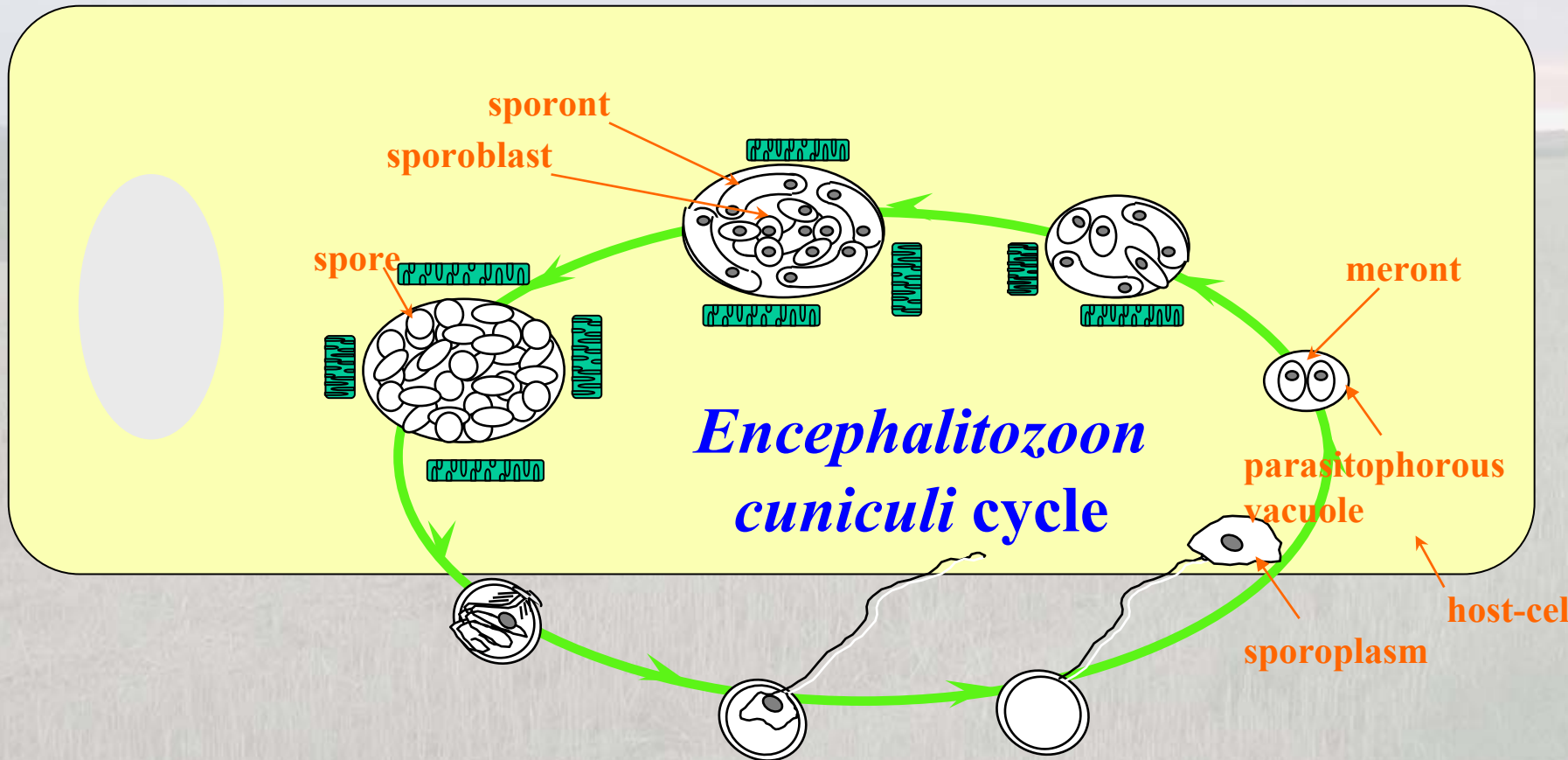
Katinka et al., 2001, Nature

# Microarray approach to identify specific gene expression during life style

Development of a specific *E. cuniculi* microarray

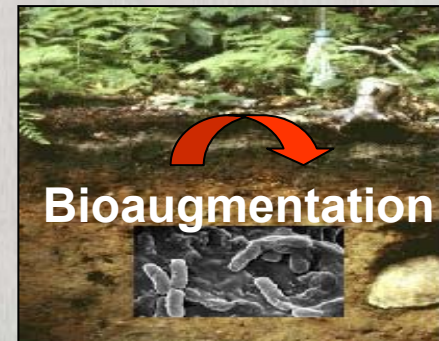
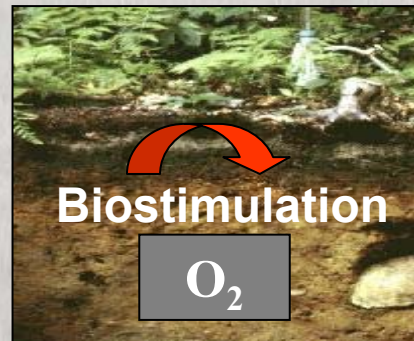
Hybridization with a mix of mRNAs from the parasite and the host cell

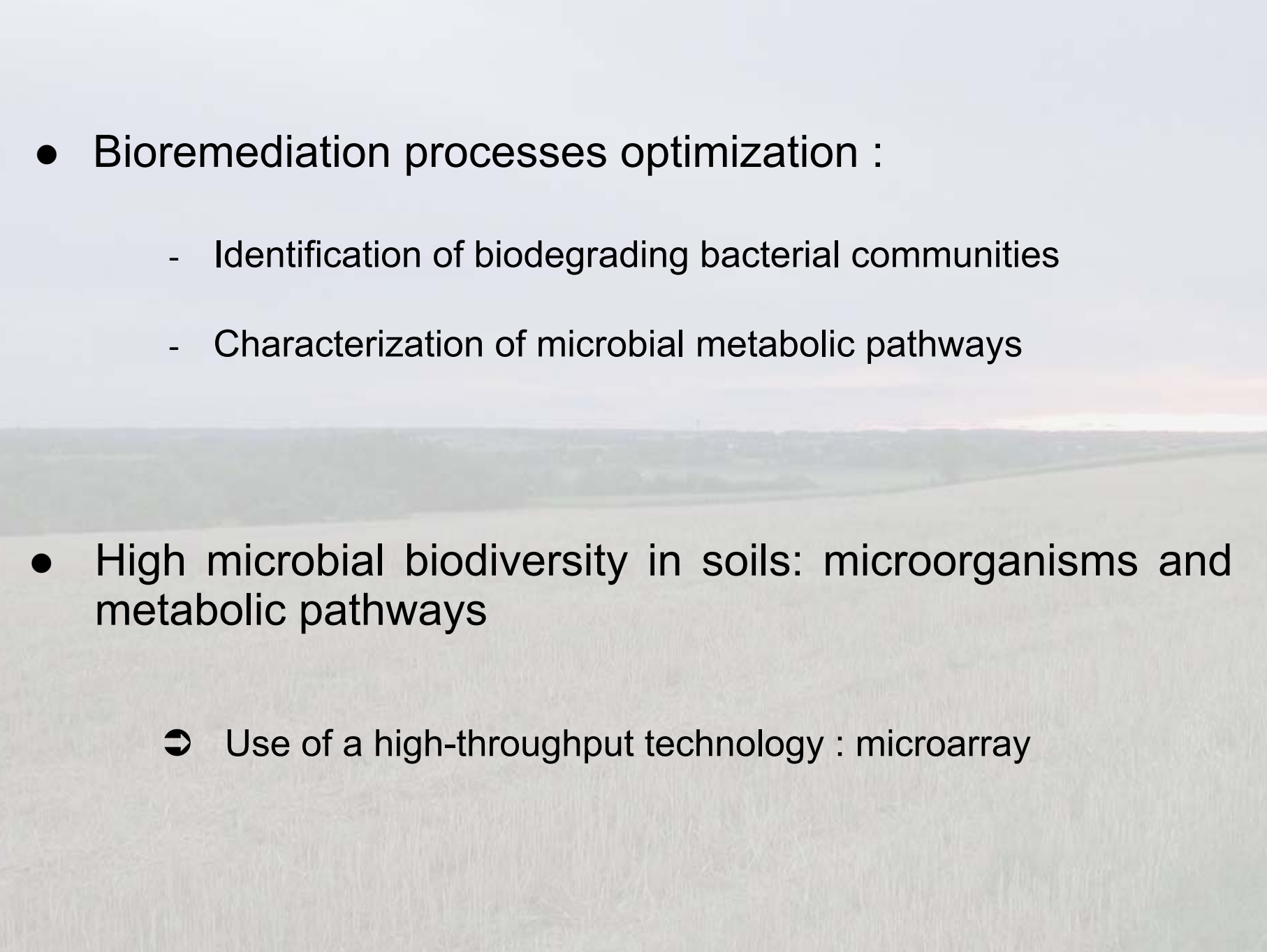
Highlight the stage specific genes to a better understanding of the parasitic life style



# Adaptative strategies of microbial consortia in polluted soils

- Two millions sites across Europe are potentially contaminated in which 100,000 requiring remediation (European Environment Agency data)
- Pollutants: essentially hydrocarbons & heavy metals
- Cleaning-up processes:
  - Physical and/or chemical processes (invasive & expensive)
  - Bioremediation processes

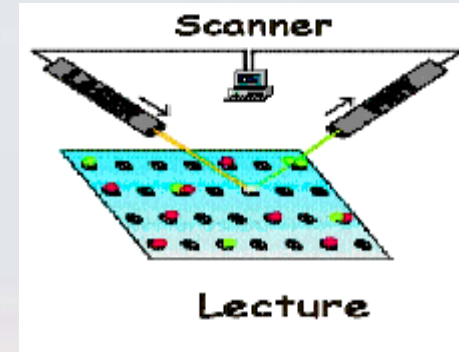
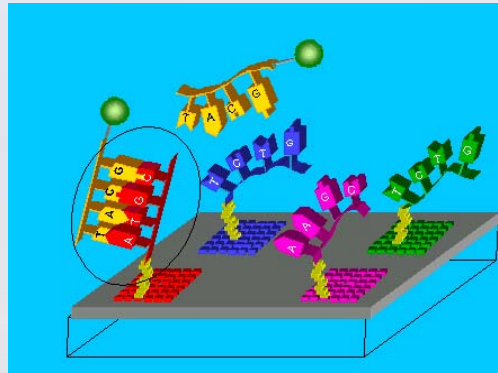


- 
- Bioremediation processes optimization :
    - Identification of biodegrading bacterial communities
    - Characterization of microbial metabolic pathways
  - High microbial biodiversity in soils: microorganisms and metabolic pathways
    - ➔ Use of a high-throughput technology : microarray



# Microarray Developments

# Microarray experiments

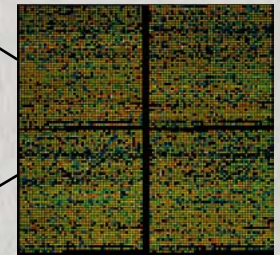
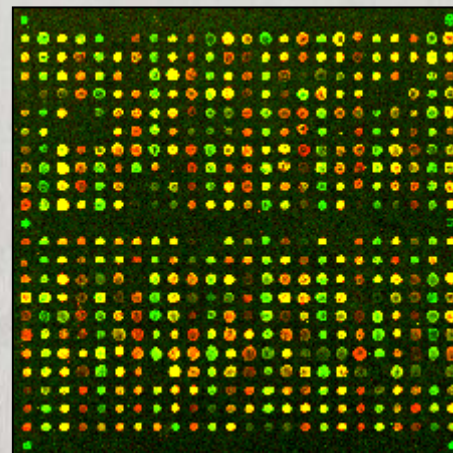


Microarray conception  
Probes design

Targets labelling  
Hybridization

Fluorescence detection

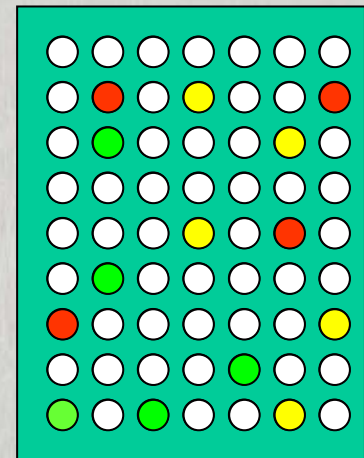
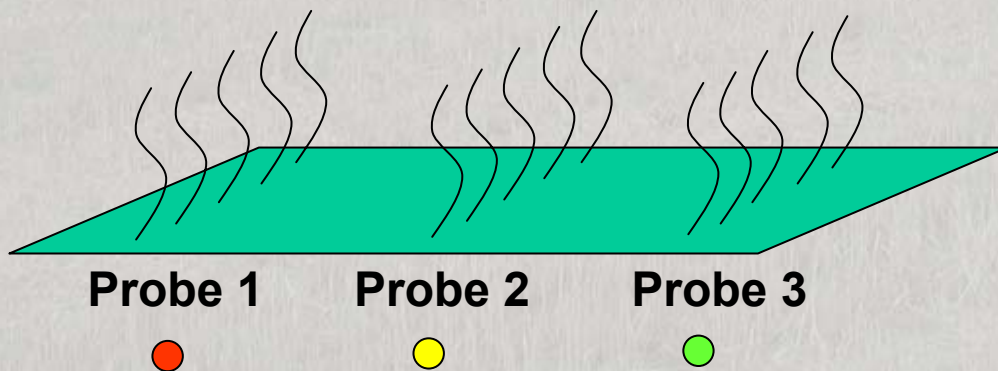
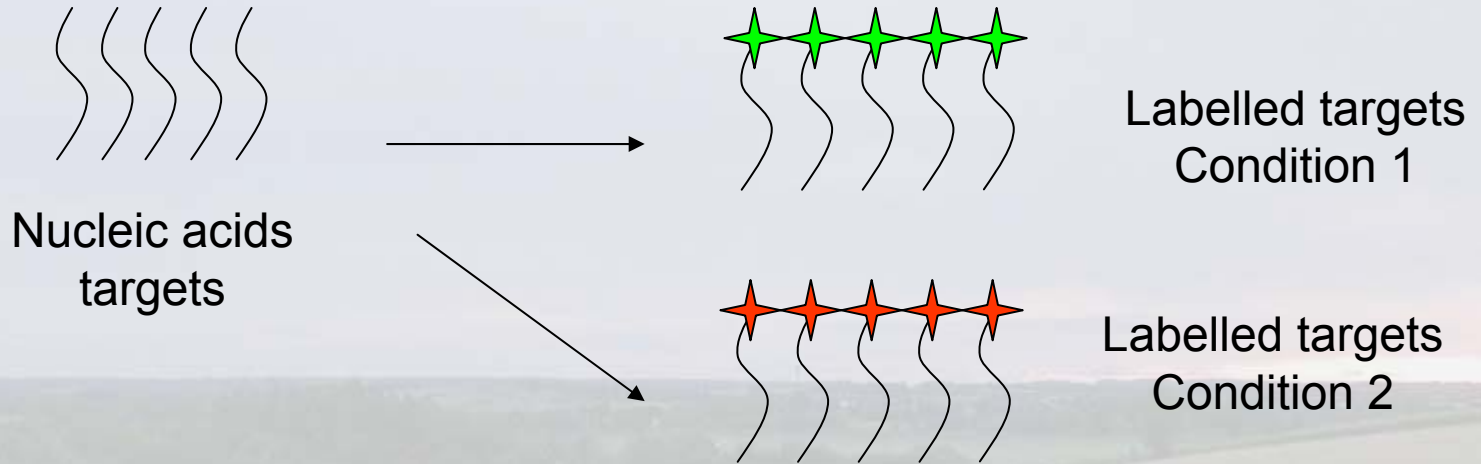
0.81	-0.62	0.57	...
-0.25	-0.65	0.87	...
0.81	-0.62	0.57	...
-0.25	-0.65	0.87	...
...			



Statistical analysis  
Biological interpretation

Image analysis

# Microarray principle



# Different applications for microarrays

Transcriptomic: genes expression,

Diagnostic (pathogens, diseases...),

Comparative Genomic Hybridization (comparaison of close genome to identify rearrangements like insertion or deletion; identification of virulence islands),

ChIP on chips (Chromatin Immunoprecipitation for the identification of DNA site fixation of transcriptional regulators),

Environmental microarrays:

- \* phylogenetic arrays
- \* functional gene arrays (metabolic arrays)
- \* community genome array

# Different kinds of DNA arrays

Macroarrays: few probes spotted on nylon membrane, targets labelled by radioactivity.

Microarrays: numerous probes (500 000 PhyloChip Affymetrix), targets labelled by fluorescence.

- \* cDNA, PCR products, genomic DNA
- \* oligonucleotides
  - short: 18-25 mers
  - long: 50-100 mers

Oligonucleotides fixation on the solid support :

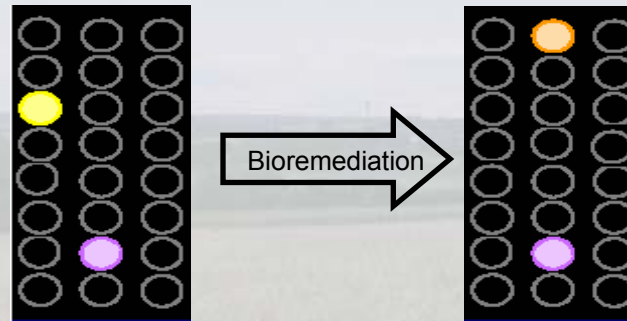
- Photolithographic (Affymetrix): *in situ* synthesis
- Inkjet (Agilent): *in situ* synthesis
- Spotting: *in vitro* synthesis



# Algorithms development

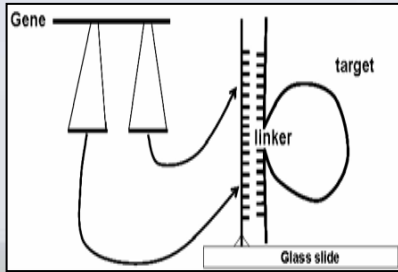
- Study approach

Follow up gene expression in complex biological system, evolution of bacterial communities and metabolic pathways dynamics using phylogenetic and metabolic microarrays in various ecosystems.



- Needs for accurate results:

- Highly efficient probes set (specific, sensitive & explorative)
- RNA targets extracted from polluted soils



1/ Development of new probe design algorithms



2/ Biological validation of algorithms with reference microorganisms



3/ Bioremediated soils study with the pilot microarray

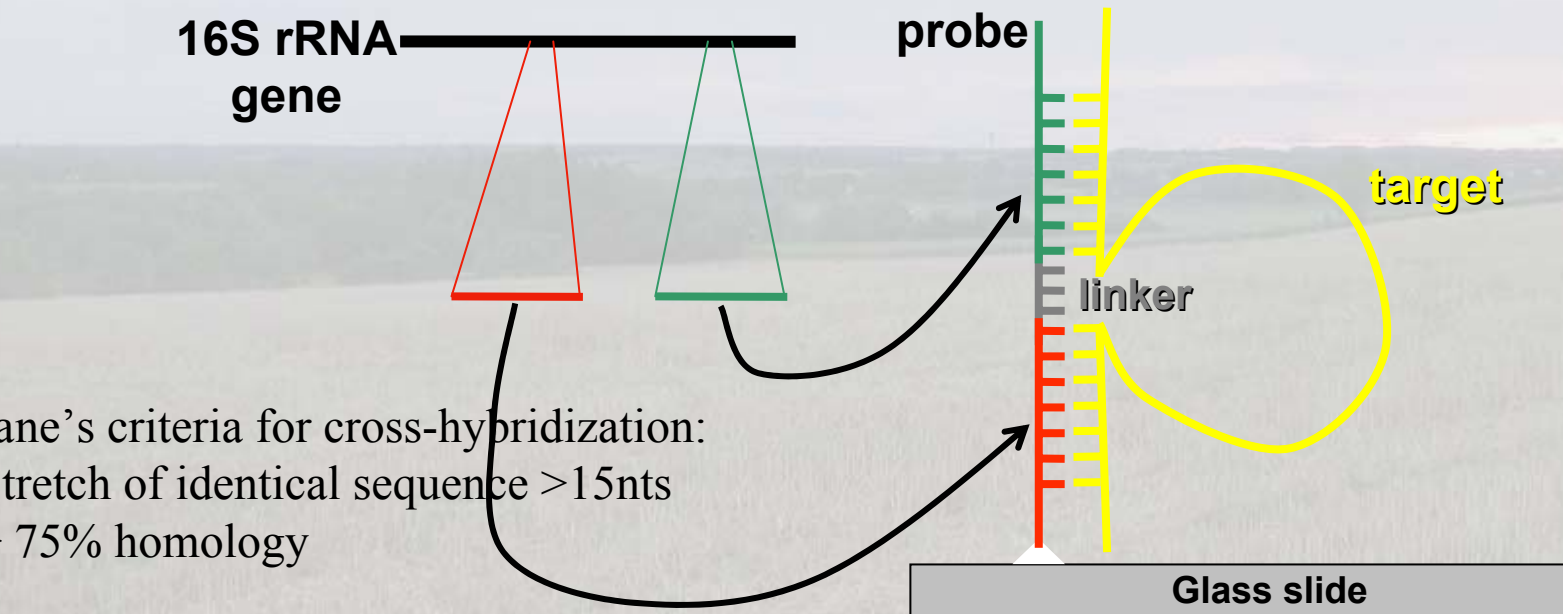
# The problem of oligonucleotide design

- Unique gene specific sequence
- Key point in the microarray manufacturing chain
- Multiple parameters are to be taken into account :
  - Cross hybridization
  - Melting Temperature
  - Distance from the 3'end
  - Secondary structure

# GoArrays strategy

Rimour *et al.* (2005) *Bioinformatics* 21:1094-1103.

- Concatenation of two specific short probes separated by a random linker

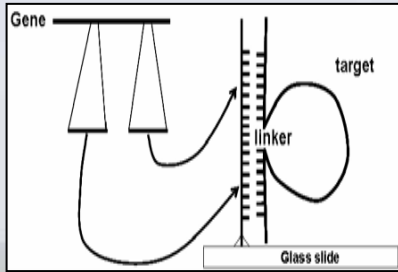


- Specificity of short sub-sequences with the potential high sensitivity of long probes

# GoArrays Interface

The screenshot shows the GoArray software interface with the following components:

- Window Title:** GoArray
- Tabbed Interface:** Four tabs are visible: "Sub Seq Oligos", "Tm", "Sec. structure", and "Cross Hyb.". The "Cross Hyb." tab is currently selected.
- Input File:** A text input field followed by a "Choose..." button.
- Output File:** A text input field followed by a "Choose..." button.
- Criteria used for selection:** Three checked checkboxes:
  - Melting Temperature
  - Secondary Structure
  - Cross hybridisation
- Database Parameters:** Two input fields:
  - DataBase Name:** An empty text input field.
  - Max length for identity:** A text input field containing the value "16".
- Buttons:** Two buttons at the bottom left: "Run !" and "Quit".



1/ Development of new probe design algorithms



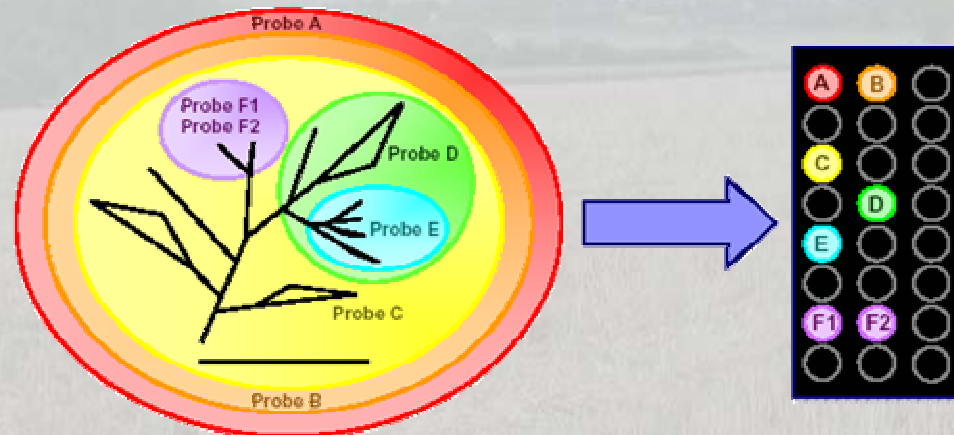
2/ Biological validation of algorithms with reference microorganisms



3/ Bioremediated soils study with the pilot microarray

# PhylArray

- Design of oligonucleotides potentially allowing the monitoring of known & unknown parts of bacterial communities
  - Taxonomic groups of higher levels detection



- Combination of degenerated and non-degenerated probes

# 16S rRNA biomarkers

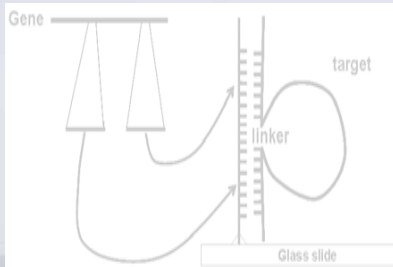
- used to infer phylogeny
- absence or very few lateral transfert
- less prone to homologous recombination
- conserved regions allowed PCR amplification
- divergent regions allowed determination of specific primers
- rapidly expanding gene database

# PhylArray algorithm

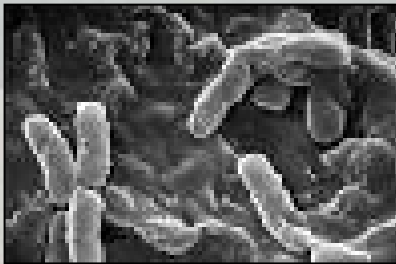
## Probes validation

- Potential probes comparison against all fungi and prokaryotic SSU rRNA sequences
- Evaluation of similarity results using Kane's criteria
- less than 75% of identity
- nucleotides stretch (no more than 15)

# Characterization of purifier microorganisms



1/ Development of new probe design algorithms

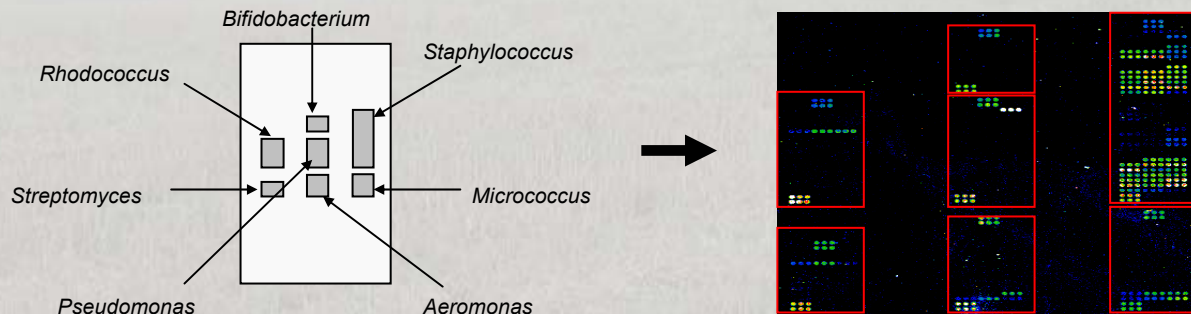


2/ Biological validation of algorithms with reference microorganisms



3/ Bioremediated soils study with the pilot microarray

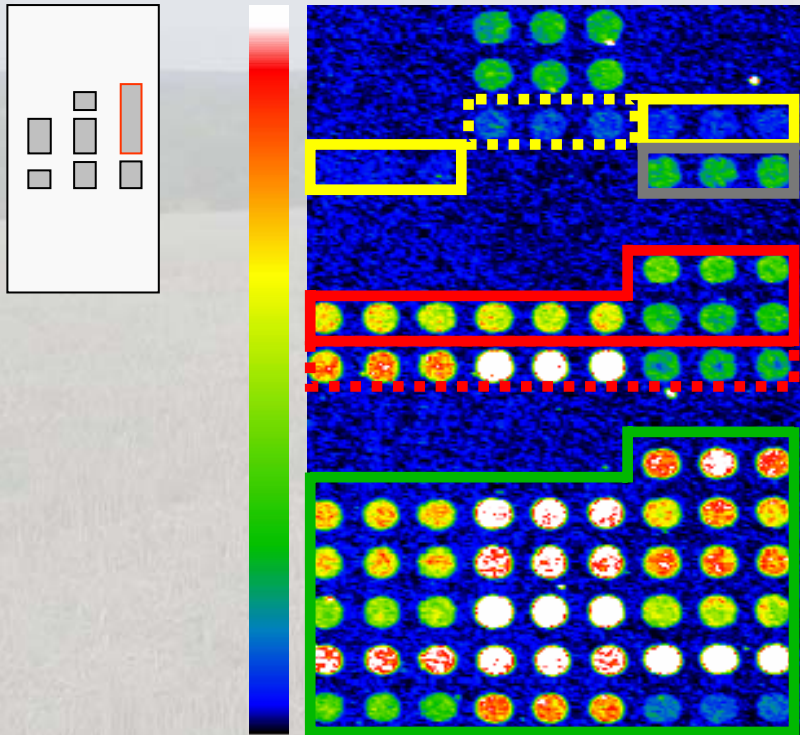
- Selection of 275 probes targeting 7 genera
  - Classical design algorithms: Primrose & ARB
  - Developed design algorithms: PhylArray & GoArrays
- Generation of a pilot microarray hybridized with 16S rRNA extracted from *Staphylococcus xylosus*



- Probes sensitivity & specificity analysis

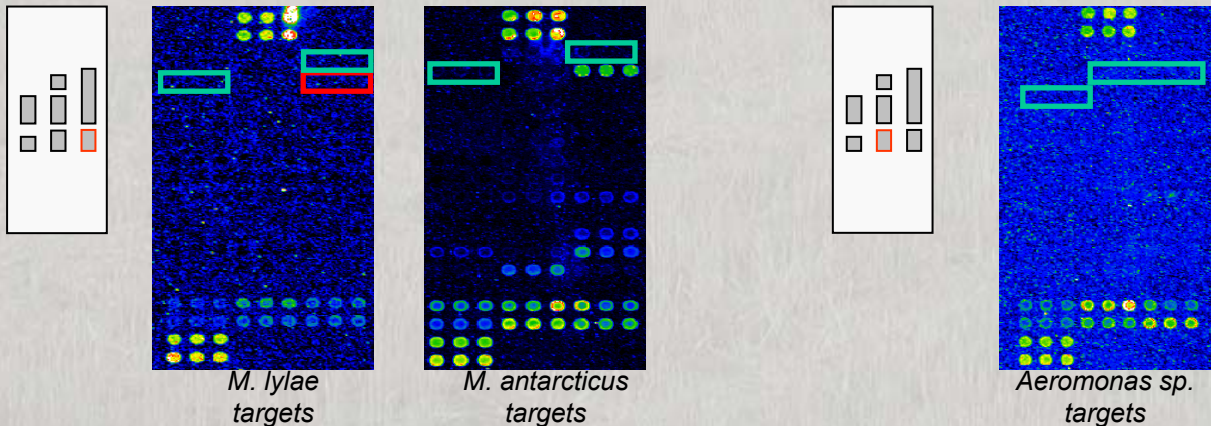
# Sensitivity and Specificity

- Fluorescences of targeting-*Staphylococcus* probes hybridized with *S. xylosus* targets

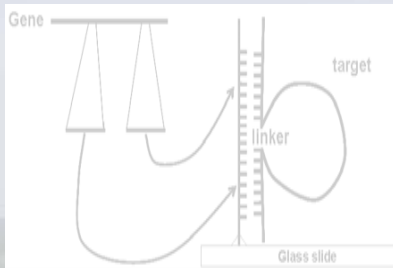


- ⇒ Short oligonucleotides
  - ARB probes
  - 25-mers PhylArray probes
- ⇒ Long oligonucleotides
  - Primrose probes
  - 50-mers PhylArray probes
- ⇒ GoArrays strategy on
  - ARB probes
  - 25-mers PhylArray probes

- PhylArray and GoArrays relevancy confirmed with others bacteria: *Micrococcus lylae* & *antarcticus*, *Enterococcus faecalis*, *Aeromonas sp.*, *Nesterenkonia sandarakina*
- Primrose and ARB probes have sensitivity failures or lacks in strain detection
  - *Micrococcus lylae* for Primrose probes (red)
  - *M.lylae*, *M.antarcticus* & *Aeromonas sp.* for Arb probes (green)



# Characterization of purifier microorganisms



1/ Development of new probe design algorithms



2/ Biological validation of algorithms with reference microorganisms



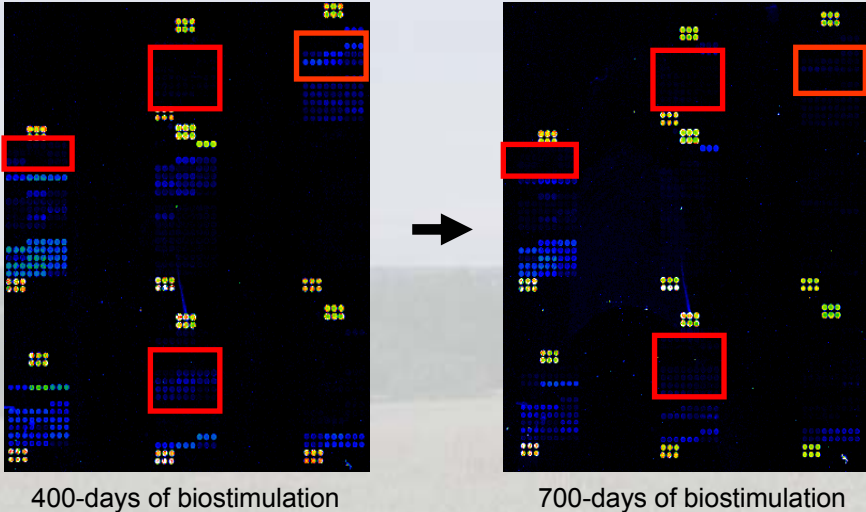
3/ Bioremediated soils study with the pilot microarray

# ● Soil decontamination follow-up

Biobasic Environnement<sup>®</sup>

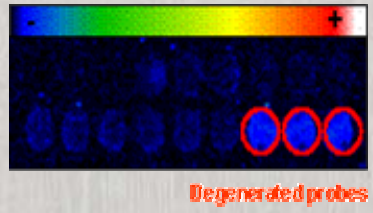
- “Z1” - Biostimulated soil from industrial site
- 10.000 ppm TPH (Total Petroleum Hydrocarbons)
- Identified pollutants: Laminated oils

- Soil microbial communities follow-up between 400 and 700-days of biostimulation



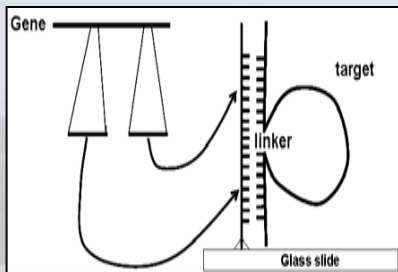
- Decrease of *Staphylococcus*
- Disappearance of *Rhodococcus* and *Bifidobacterium*
- Complex dynamic for *Aeromonas*

- Detection of new microorganisms with degenerated probes



Unknown species belonging or close to the targeted genus (non-indexed in databases)

# Identification of metabolic pathways



1/ Development of new probe design algorithm



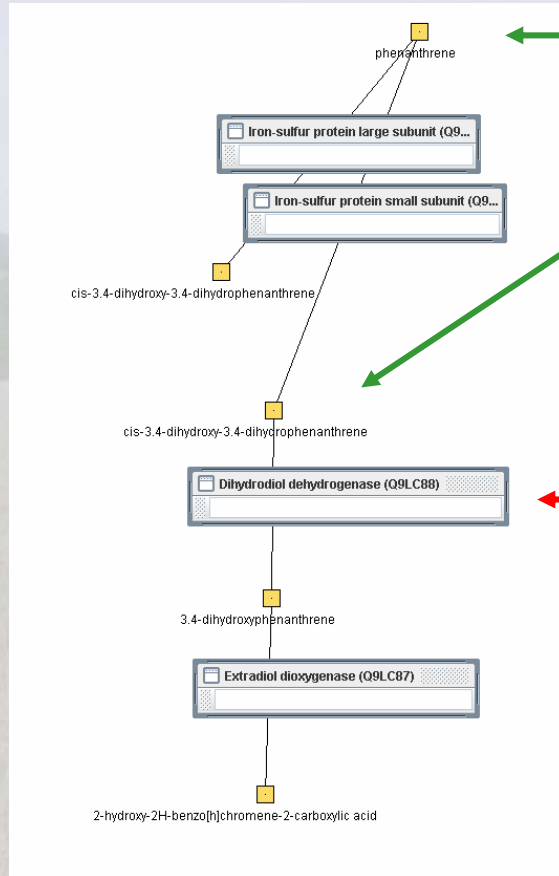
2/ Biological validation with a reference strain

# Metabolic Design

- Characterization of microbial metabolic pathways
  - Selection of probes targeting enzymes.
    - Degenerated probes to ensure the recognition of all genes encoding these enzymes (whatever the microorganism).

# MetabolicDesign algorithm

## 1) Metabolic pathways *in-silico* generation



**Metabolites**

**Reference enzyme  
searched in Swissprot**

# Metabolic Pathway search by species



The screenshot displays a metabolic pathway search interface with several search results windows. A red circle is present in the top right corner of the overall image.

**Search Results: Biphenyl dioxygenase alpha subunit**

e-value	Name	Id
0	Pseudomona...	Q50578
0	Pseudomona...	P95566
0	Pseudomona...	Q52383
0	Pseudomona...	Q9AEY5
6.E-89	Pseudomona...	Q7WTK6
4.E-88	Pseudomona...	Q7WTK5
2.E-87	Pseudomona...	Q7WTK3
2.E-87	Pseudomona...	Q7WTK4
3.E-74	Pseudomona...	Q7WTK7
5.E-57	Pseudomona...	P95564
7.E-56	Pseudomona...	Q6XUJ7
4.E-38	Pseudomona...	Q8KLZ1
4.E-38	Pseudomona...	Q9F5T0
2.E-36	Pseudomona...	Q8KLY8
2.E-36	Pseudomona...	Q8KLY9
2.E-36	Pseudomona...	Q8KLZ0
3.E-36	Pseudomona...	Q8KLY7
1.E-35	Pseudomona...	Q9F5S8
1.E-35	Pseudomona...	Q9F5T3

**Search Results: Biphenyl dioxygenase beta subunit**

e-value	Name	Id
3.E-69	Pseudomona...	Q79C36
2.E-67	Pseudomona...	Q7B113
1.E-64	Pseudomona...	Q9AEY4
2.E-56	Pseudomona...	Q52384
2.E-13	Pseudomona...	P95565
4.E-10	Pseudomona...	Q6XUJ6

**Search Results: 2 3-dihydroxy-1-phenylcyclohexa-4 6-diene d...**

e-value	Name	Id
2.E-85	Pseudomona...	Q9AEY1
2.E-77	Pseudomona...	P95570
1.E-69	Pseudomona...	Q52387
1.E-50	Pseudomona...	Q6XUJ5

**Search Results: 2 3-dihydroxybiphenyl dioxygenase**

e-value	Name	Id
0	Pseudomona...	Q7B8M5
0	Pseudomona...	Q7BP42
1.E-114	Pseudomona...	Q9AEY0
4.E-98	Pseudomona...	P95571
3.E-44	Pseudomona...	O32477
1.E-42	Pseudomona...	Q9RBT2
9.E-40	Pseudomona...	Q6XUJ3
4.E-39	Pseudomona...	Q52533
4.E-33	Pseudomona...	Q52534

**Search Results: 2-hydroxy-6-oxo-6-phenylhexa-2 4-dienoate ...**

e-value	Name	Id
6.E-47	Pseudomona...	Q8RT87
5.E-45	Pseudomona...	Q9RBT0
3.E-42	Pseudomona...	Q52532
2.E-36	Pseudomona...	Q7BJY2
7.E-34	Pseudomona...	Q6PKW3
2.E-33	Pseudomona...	Q6XUR7
2.E-29	Pseudomona...	Q84IA0
3.E-10	Pseudomona...	Q8VFP0

**Search Results: benzoate - cis-2-hydroxypenta-2...**

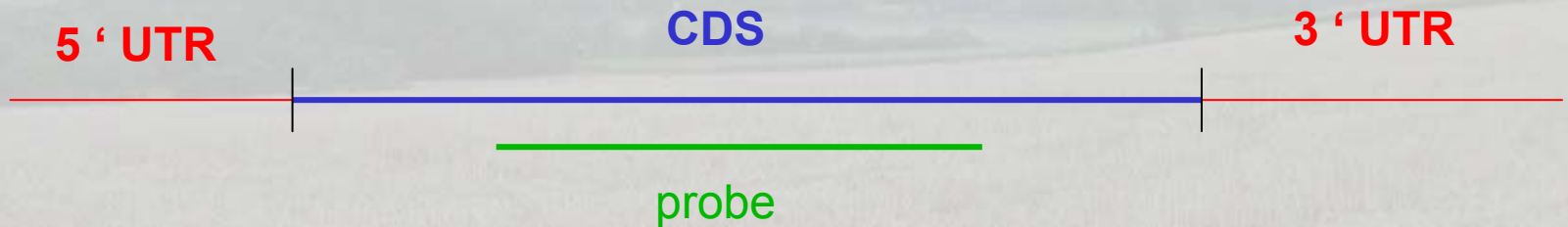
e-value	Name	Id
6.E-47	Pseudomona...	Q8RT87
5.E-45	Pseudomona...	Q9RBT0
3.E-42	Pseudomona...	Q52532
2.E-36	Pseudomona...	Q7BJY2
7.E-34	Pseudomona...	Q6PKW3
2.E-33	Pseudomona...	Q6XUR7
2.E-29	Pseudomona...	Q84IA0
3.E-10	Pseudomona...	Q8VFP0

Navigation buttons at the bottom: **- BLAST -**, **- Graphe des voies métaboliques -**, **- Graphe des especes impliqués -**

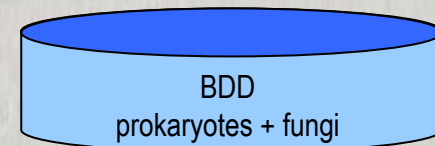
# MetabolicDesign algorithm

## Evaluation of candidate probes specificity

- Search of potential cross-hybridizations coding sequences belonging to prokaryotes and fungi

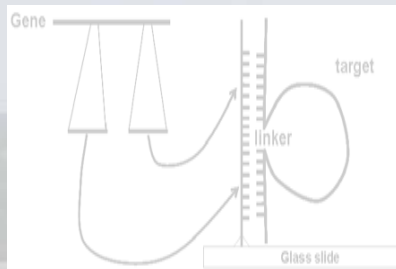


- Database generation :
  - CDS sequences + UTR regions (100 nt)
  - functional RNA sequences



- EMBL files from Prokaryote & Fungi databases exploitation

# Identification of purifier metabolic pathways



1/ Development of new probe design algorithm



2/ Biological validation with a reference strain

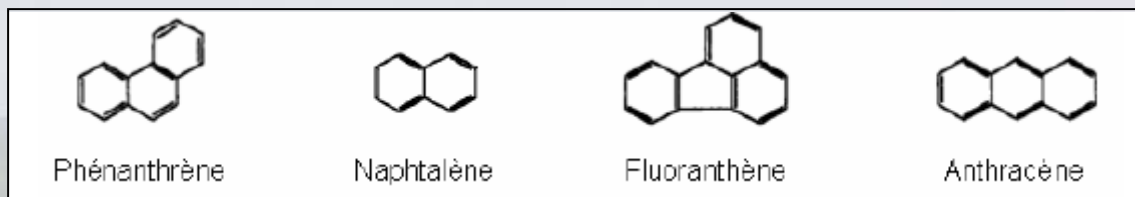
- Pilot metabolic microarray:

- *In-silico* generation of phenanthrene degradative metabolic pathway
- Probes selection for each metabolic pathways steps

Reference enzymes	Species	Probes
Ferredoxin SU (Q83VL0)	<i>Sphingomonas</i> sp. P2	0
Oxygenase SSU (Q83VL1)	<i>Sphingomonas</i> sp. P2	4
Oxygénase LSU (Q83VL2)	<i>Sphingomonas</i> sp. P2	2
1,2-dihydrodiol 1,2-dihydroxy dehydrogenase (Q9X9Q9)	<i>Sphingomonas xenophaga</i>	2
Biphenyl-2,3-diol-dioxygenase (P11122)	<i>Sphingomonas paucimobilis</i>	2
2-hydroxychromene-2-carboxylate isomerase (Q83VL3)	<i>Sphingomonas</i> sp. P2	0
2-hydroxy-benzylpyruvate aldolase (Q83VI8)	<i>Sphingomonas</i> sp. P2	2
Salicylaldehyde dehydrogenase (P0A390)	<i>Pseudomonas</i> sp. C18	4

- ➔ 16 specific oligonucleotides for the majority of the pathway
- GoArrays strategy used for probes design

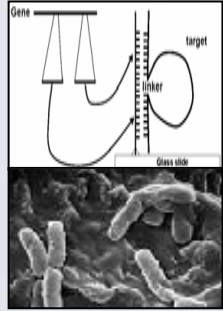
- Biological validation of the pilot microarray:
  - Reference strain: *Sphingomonas paucimobilis* EPA 505
  - Abilities to degrade several PAH (in which phenanthrene)



- Preliminary studies:
  - *S.paucimobilis* growth follow-up on naphthalene, fluoranthene and phenanthrene
  - Characterization of genes belonging to the *S.paucimobilis* phenanthrene metabolic pathway



# **Conclusions and prospects**



## ➔ Biological validation of probe design algorithms

- PhylArray probes (25-mers) are more sensitive & specific
- The GoArrays strategy decreases cross-hybridizations and trebles signals intensities



## ➔ Microbial dynamic during a biostimulation

- Microarray is usable for complex environments
- Monitoring of known & unknown parts of bacterial communities

## ● Prospects

- Complete phylogenetic microarray elaboration (potentially 1560 genera)
- Characterization of microbial strains with the complete microarray
- Characterization of metabolic pathways with the functional microarray

# Acknowledgements



## **Algorithms Conception and Microarray Experiments**

A.MONE, Y.KERRIOU, M.BONNET, B.CHEBANCE,  
G.GAGNE, E.PEYRETAILLADE, P.VEISSEIRE  
P. PEYRET.



## **Softwares Implementation and Microarray Data Analysis**

S. RIMOUR  
V. BARRA



## **Bioremediation Engineering**

C. VACHELARD  
J. TROQUET