Technical Note TN70.2
Part A: Review and trade-off of the technical methods to monitor genetic/proteomic evolution
June 2002
Number of pages including front page: 58

ESA/ESTEC
ESTEC contract number 15680/01/NL/ND

	Name	Company	Signature	Date
Prepared by	Larissa Hendrickx	SCK/CEN		
	Paul Janssen	ULg		
	Ruddy Wattiez	UMH		
	Sarah Baatout	SCK/CEN		
	Annick Wilmotte	ULg		
	Philippe Bertin	SCK/CEN		
	Max Mergeay	UMH		
Checked by				

CONFIDENTIAL

Distribution list

Quantity	Company/Department	Name
2	ESA	C. Paillé
		N. Dorval
2	UMH	R. Wattiez
3	ULg	P. Janssen
		A. Wilmotte
6	SCK/CEN	M. Mergeay
		S. Baatout
		L. Hendrickx
10	SCK/CEN	Stock

Document change log

Version	Issue	Date	Observation
0	1	05/07/02	Draft 1
1	1	30/08/02	Draft 2
2	1	14/01/03	Final version

Abbreviations

AFLP(analysis)	amplified (DNA) fragment length polymorphism
AP-PCR	arbitrary primed polymerase chain reaction (of DNA)
ARDRA	amplified ribosomal DNA restriction analysis
BOX-PCR	polymerase chain reaction of (DNA sequences between) BOX Elements
DGGE	denaturing gel gradient electrophoresis
DNA	desoxyribonucleinic acid
ERIC-PCR	polymerase chain reaction of enterobacterial repetitive intergenic consensus sequences
FAME	fatty acid methyl ester analysis
IGS(analysis)	intergenic spacer analysis (=ITS analysis =RISA when used for ribosomal spacers)
Inter-LINE-PCR	polymerase chain reaction of (DNA sequences between) long interspersed elements
ITS(analysis)	internal transcribed spacer analysis (=RISA)
MLEE	multilocus enzyme electrophoresis
ORFG	Oligonucleotide fingerprinting of rRNA genes
PFGE	pulsed field gel electrophoresis
PCR-RFLP	restriction fragment length polymorphism of polymerase chaincreaction-generated amplicons
RAPD-PCR	polymerase chain reaction of random-amplified polymorphic DNA
RC-PFGE	pulsed field gel electrophoresis (of DNA digested by) rare cutting endonucleases
REP-PCR	polymerase chain reaction of (DNA sequences between) repetitive extragenic palindromic elements
rep-PCR	polymerase chain reaction of (DNA sequences between) repetitive (REP, ERIC or BOX) elements
RISA	ribosomal intergenic spacer analysis
SCAR	sequence-characterized amplified (DNA) regions
SSCP	Single stranded conformation polymorphism
TGGE	temperature gradient gel electrophoresis
TRFLP	terminal restriction fragment length polymorphism

Table of contents

1. Sampling methods **7**
1.1. Proteomic analysis 7
   1.1.1. General precautions 7
   1.1.2. Sample processing 7
1.2. Genomic analysis 7
   1.2.1. General precautions 7
   1.2.2. Sample processing 7
1.3. Flow cytometry 8
   1.3.1. General precautions 8
   1.3.2. Sample processing 9

2. Proteomic analysis **10**
2.1. Outline and principle 10
2.2. Study of stress response and radiation sensitivity by mass spectrometry 13
2.3. Membrane or surface accessible proteins detection (SAPD) 14
2.4. General procedures 15
   2.4.1. Short term and long-term storage procedures 15
   2.4.2. Target selection/Validation 15

3. Description of biodiversity **17**
3.1. Outline and principle 17
3.2. Amplified 16S and 16-23S ribosomal DNA gene sequences and restriction endonuclease analysis (ARDRA) (Massol-Deya et al., 1995) 17
3.3. Analysis of bacterial communities using whole cell *in situ* hybridization (Amann, 1995) 19
3.4. Single stranded conformation polymorphism (SSCP) (Hawkins, 1997) 19
3.5. Terminal restriction fragment length polymorphism (TRFLP) 19
3.6. Denaturing gradient gel electrophoresis (DGGE) (Muyzer et al., 1998) or temperature gradient gel electrophoresis 20
3.7. Oligonucleotide fingerprinting of rRNA genes (ORFG) 20

4. Investigation of critical genes by DNA chip technology **22**
4.1. Outline and principle 22
4.2. Experimental approach 23

5. Genomic pattern analysis by PCR-based techniques **25**
5.1. Outline and principle 25
5.2. Genotyping methods 26
5.3. Polymorphism analysis by AFLP 27
   5.3.1. The AFLP concept 27
   5.3.2. Template preparation 28
   5.3.3. Selective PCR 29
   5.3.4. Fragment separation 30
   5.3.5. Pattern visualisation 31
   5.3.6. Data analysis 31
   5.3.7. Applications 31

6. Flow cytometry analysis **33**
6.1. Outline and principle 33
6.2. Analysis Equipment at SCK-CEN: the EPICS XL (Beckman Coulter) 34
6.3. Optimised detection of bacterial changes induced by exogeneous factors. 35
6.3.1. Optimisation of the flow rate of sample cell inside the flow cytometer 35
6.3.2. Discrimination of bacteria from the background 35
6.3.3. Basic detection of differences in bacterial size and granularity 35
6.3.4. Assessment of viability and survival 37
6.3.5. Use of flow cytometry to check DNA content changes in microbial populations 39
6.3.6 Preparation of samples to analyse DNA content from bacterial cultures 40

7. Bioinformatic approaches **43**
7.1. Outline and principles 43
7.2. Bioinformatic tools 43
7.3. Databases 44
7.4. Relevant analysis strategies 45

8. References **46**
8.1. Reference list for proteomic analysis 47
8.2. Reference list for bacterial community analysis 49
8.3. Reference list for DNA chip technology 52
8.4. Reference list for genomic analysis 53
8.5. Reference list for flow cytometry 55

1. Sampling methods

1.1. Proteomic analysis

1.1.1. General precautions

The proteome of a bacterial species could be modified by proteases present in the culture or in the supernatant. Wherever possible, a protease cocktail inhibitor will be used (e.g. Complete™ Mini, EDTA-free, Roche) and work at 4 °C. Isolated proteins (ex : filtrated supernatant) can be temporarily stored on ice or refrigerated (2-8 °C) or frozen (-80—20 °C) for long-term storage. Storing protein can cause protein precipitation, especially if the protein solution is repeatedly going through freeze-thaw cycles. Thus, it is recommended to freeze multiple aliquots of each preparation.

1.1.2. Sample processing

Number and volume of samples required for each time point, frequency of sampling

- **TBP and SP techniques:** TBP :3 samplings (1mL) for each stress conditions. The volume of each supernatant culture has to be evaluated primarily.
- ICM-MS: **3 samplings (1ml) for each stress conditions.**
- **SAPD technique:** 3 samplings (5-10mL) of the axeny or inoculums cultures (especially for the compartment III and the *Arthrospira* compartment).

Special considerations when taking the sample

- **TBP-SP ; ICM-MS and SAPD techniques :**
  Ideally, all experiments are to be performed in sterilized polypropylene test tubes. Liquids must be transferred using sterilized graduated plastic pipettes.

Sample preparation just after the sampling

- **ICM-MS and SAPD techniques:** for the intact cell analysis and the surface protein detection, the samples do not need to be submitted as a specific preparation. A sterile storage of the samples at 4 °C is recommended.
- **SP technique:** The samples will be sterilised with specific filtration units (0.2um) (ex: Nalgene disposable filter ware: 150mL filter unit with sterile receiver) just after the sampling. The filtration has to be done in the sterile hood. Before the filtration, a protease cocktail inhibitor (1 tablet for 25 mL of supernatant) is added to the supernatant (in the sterile hood). After filtration, samples are kept at –20 °C.

1.2. Genomic analysis

1.2.1. General precautions

It is an absolute prerequisite that growth of experimental and reference cultures and sampling from the reactors for PCR reactions are carried out in a controlled and sterile way, as even minute quantities of contamination, in the order of nanograms of free DNA or as few as 10-100 individual cells, may give false signals in the PCR results. For cultures, one should work in a laminar flow hood, using sterilised (e.g. autoclaved) material. For taking samples from the MELISSA reactors, different

strategies need to be deployed, depending on the compartment, but great care must be taken that at no time there is contact with personnel, and as little as possible contact with the open environment (e.g., by treating reactor outlets and recipient openings with a blue flame during the sample taking).

1.2.2. Sample processing

Samples must be fixed in order to enable the preservation of the nucleic acids (containing the target sequence for PCR). Several methods are possible.

1) The best method is freezing, an easy method when the subsequent steps of analysis are carried out in the same laboratory. The practical drawback of this method when the DNA extraction and PCR steps are performed in a different country (as for the samples of the pilot plant of Barcelona) is that transportation should be done in dry ice, a costly and not very flexible procedure.

2) A practical alternative is the fixation with 70% ethanol. This is also the concentration where nucleic acids are precipitated and the ULg ‘s lab has good experience with that method for preserving the nucleic acids of cells and sending them by normal post at ambient temperatures.

3) Molecular biology companies have designed reagents for stabilizing RNA, as RNAlater (Ambion Inc., USA), of secret composition, that should be able to preserve the nucleic acids when sampling in places where there is no possibility to freeze (field trips, etc). One can use it for DNA but it depends on the precise method of extraction (e.g. not compatible with the use of phenol).

4) It should be noted that the use of fixatives like formaldehyde should be avoided because they cross-link all cell materials and will inhibit PCR.

DNA quality. There are numerous methods available for the preparation of microbial genomic DNA. The most commonly used methods are based on cell lysis through addition of cell wall degrading enzymes (e.g., lysozyme, lysostaphin, proteinase K, etc.) and ionic agents such as SDS or sarcosyl, followed by phenol and chloroform extractions. Methods that aim to obtain high molecular weight DNA from organisms known to release high levels of DNAses often make use of nucleic acid stabilising agents such as cetyltrimethylammonium bromide (CTAB) or guanidinium thiocyanate (GTC). For AFLP it is strongly recommended to use DNA of adequate purity and high molecular weight. The simple reason lies in the fact that the activity of restriction enzymes may be sensitive to co-purified impurities leading to inspecific cleavage ('star-activity') or incomplete digestion. If this occurs, partial fragments of aberant size will be amplified with serious consequences for data interpretation and reproducibility.

1.3. Flow cytometry

1.3.1. General precautions

Ideally, samples should be first analysed fresh. Alternatively, if fresh analysis is not possible, a simple method for the preservation of the samples, that interfere minimally with the cellular properties of cells can be used (see preservation and storage of samples). The combined analysis of the forward and the side scatters can

allow the identification of different groups that differ in terms of size and granularity, respectively. Several aspects are critical to succesful analysis of bacterial samples :

- careful sample preservation (if necessary)
- good discrimination of populations from noise
- accurate identification of populations
- careful determination of flow sample rate

Best results should be obtained on fresh samples run immediately after collection. Fresh samples can be stored at 4°C for up to 12 hours with minimal effect. Fixation (in paraformaldehyde solution) will always result in cell loss (about 10%) and in change of scatters signals.

1.3.2. Sample processing

Retention of fully replicated chromosomes . To obtain bacteria that contain only fully replicated chromosomes, different substances such as rifampicin (around 150 µg/ml) or cephalexin (around 10 µg/ml) will have to be added to the culture grown to an optical density The molecular mechanism of rifampicin is an inhibition of bacterial RNA polymerase that forms a complex with rifampicin. Rifampicin prevents the ribosomes from binding to messenger RNA, and thus proteins are not produced.

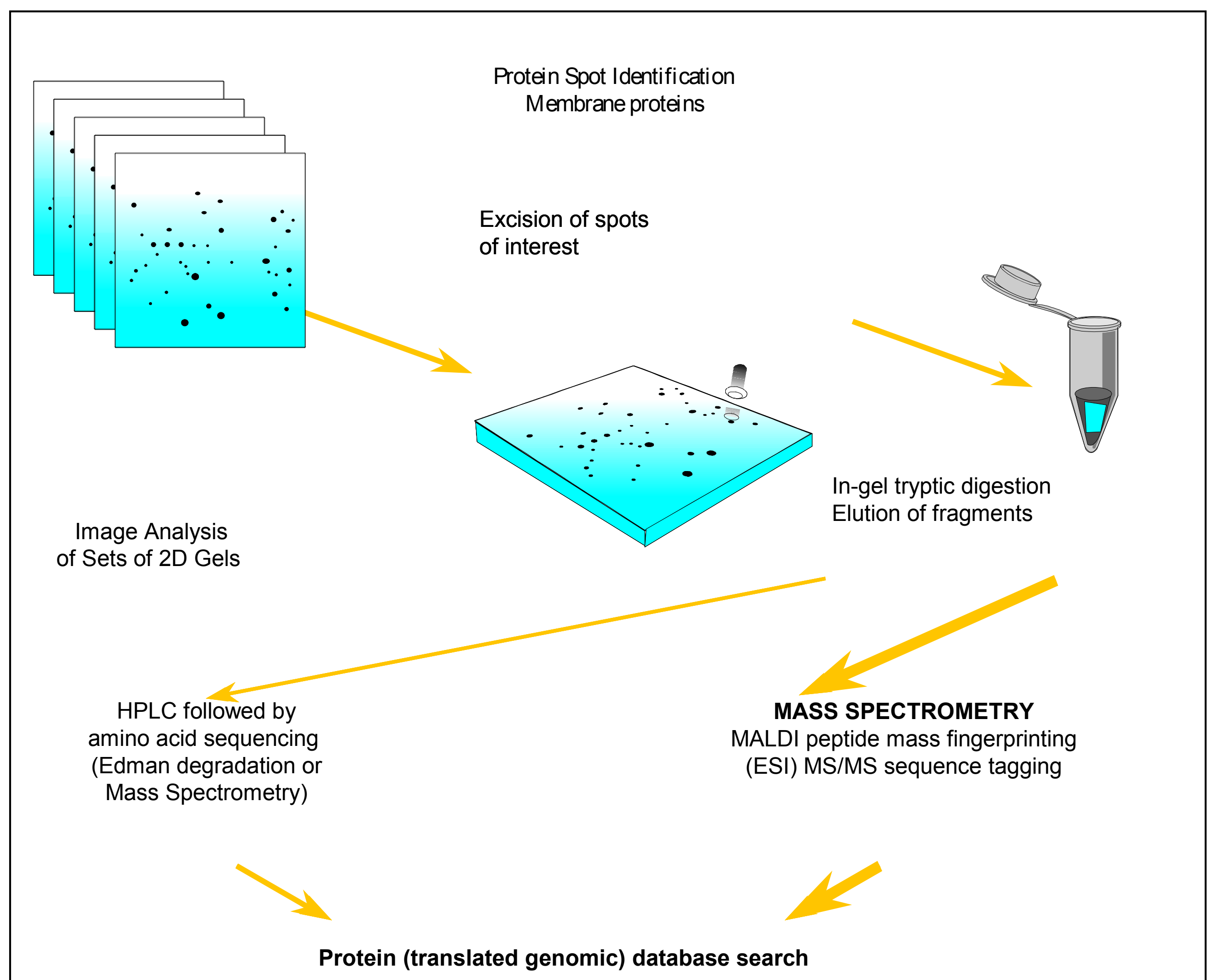
Flow cytometry is useful in order to study the DNA distribution of an exponentially growing culture. When DNA of an exponentially growing *E. coli* is investigated in the absence of antibiotics, it is observed that the DNA distribution resembles that of the age distribution, there are more cells with the low DNA content corresponding to a little more than 2 fully replicated chromosomes, than cells with twice the amount of DNA. When rifampicin or cephalexin is added, the DNA occurs in discrete peaks, most of the cells have 4 fully replicated chromosomes. The number of fully replicated chromosomes corresponds to the number of origins present in the cell at the time of drug addition. Cultures will have to be preincubated for around 3 hours (time to optimise) in the presence of rifampicin and cephalexin in order to allow complete runout of replication, such that all cells contain only fully replicated chromosomes.

2. Proteomic analysis

2.1. Outline and principle

Proteomics (proteome is defined as the expressed protein complement of a genome) holds a key position in the new biology. This discipline recently emerged from the decades-long work on comprehensive protein visualization on two-dimensional gels, which was revitalized by developments in biological mass spectrometry and the growth in searchable sequences databases. Proteomics adds value to these databases by providing tools for the parallel separation and **large-scale identification of proteins** (i.e. high-throughput bacterial membrane protein identification) (figure 2.1). Over the past ten years, mass spectrometry has become the technique of choice for protein or molecules characterization. The reason is the development of new methods for the ionization of proteins and peptides, especially matrix-assisted laser desorption-ionization and electrospray ionization. In the field of proteomics, the technique of MALDI- TOF MS is particularly suited to **high throughput identifications of low femtomole level protein** digest samples (figure 1). The samples are previously excised and processed from 2D electrophoresis gels or high-pressure liquid chromatography.

Figure 2.1.: Strategy for high throughput protein identification



The technique routinely provides high mass measurement accuracy (< 10 PPM RMS) which leads to high specificity for identification of proteins from databases. The new reflectron based instrument has been optimized primarily for the function of peptide mass fingerprinting of large arrays of protein digest samples. Now, the MALDI-TOF technique can be used also to **characterize microorganisms**.

Proteome analysis is a particularly powerful tool for study **stress response or radiation sensitivity** of different bacterial species as shown in numerous examples (review, proteomics a trends guide). Moreover, the proteomic approach allows to **detect and identify specific surface proteins.**

A variety of characteristics of microorganisms (both phenotypic and genotypic) are used to differentiate between strains and often to determine the relationship between strains of bacteria present on the bacterial cell wall are surface components, which give rise to a unique pattern of biomarkers. These surface components are important since they mediate the contact between the cell and the environment. This unique population of molecules can be rapidly desorbed from the cell surface; ionized and analysed by matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) resulting in a mass fingerprint. This mass fingerprint is characteristic of the particular species, and in some cases differences at the strain level can be observed. Automated collection of mass spectral data from 96 well plates,in less than 30 seconds per sample necessitates the implementation of customised software to analyse and report the data produced in real time. Innovative spectral matching software, which employs a probabilistic search algorithm, is used to match these 'bacterial mass fingerprints' against the database of quality-controlled reference spectra. The best match, with a probability score, and the genus, species and strain of the organism is reported.

A schematic of the instrument is shown in Figure 2.2.. The reflectron system has an effective path length of 2.3 meters. The target plate is held at ground potential and the flight tube and detector assembly are held at - 20kV. “Time lag focusing ”(TLF) is applied to the source to enhance resolution by velocity focusing, this allows the instrument resolution to be almost independent of the laser fluence - an essential element of automated control. The TLF method has been further modified so that the optimum pulse voltage is the same for both high and low mass peptides . The detector system consists of a 2 inch dual MCP (Burle, MA). The signal from the detector is acquired using a PC based 8 bit, 2GS/ sec A/ D converter. The target plate is positioned on an X- Y stage using stepper motors 4 microns resolution. Either 96 or 384 well plates may be accommodated. The ionising laser is the OEM VSL- 337i (3 nsecs pulse width (FWHM) at a UV wavelength of 337nm) . The laser fluence is computer controlled using a motorised iris assembly. One of the most contentious issues in MALDI (matrix assisted laser desorption ionization) is the role of the matrix in the ionization process. When the laser strikes the matrix crystals, energy deposition is though to cause rapid heating of the sample-crystals brought about by matrix molecules emitting absorbed energy in the form of heat. Laser ionizes sample by 3-ns-wide pulses. Ions are accelerated into the flight tube by accelerating voltage. Since they are allowed to drift through the field-free region to the detector, **they separate according their mass-to-charge ratios**. Lighter ions drift more quickly, heavier more slowly (time of flight -TOF) (figure 2.3.).

Figure 2.2. : Reflectron M@ LDI Instrument Description

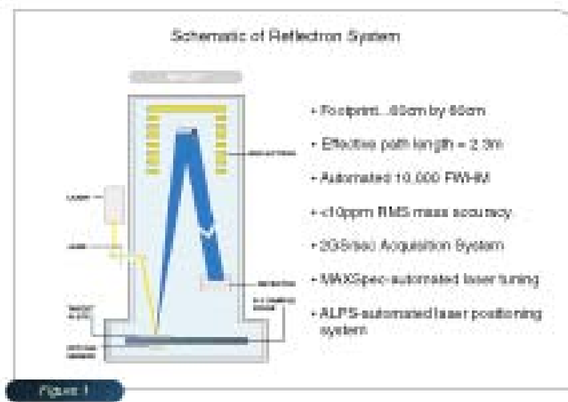
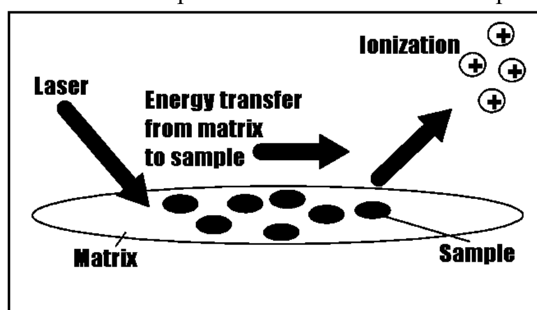
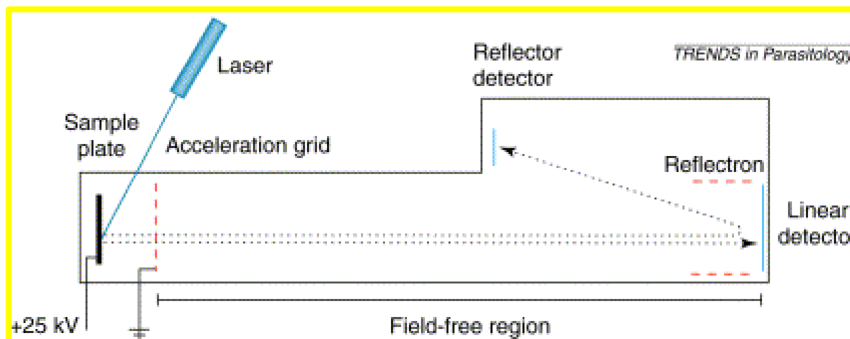


Figure 2.3. :

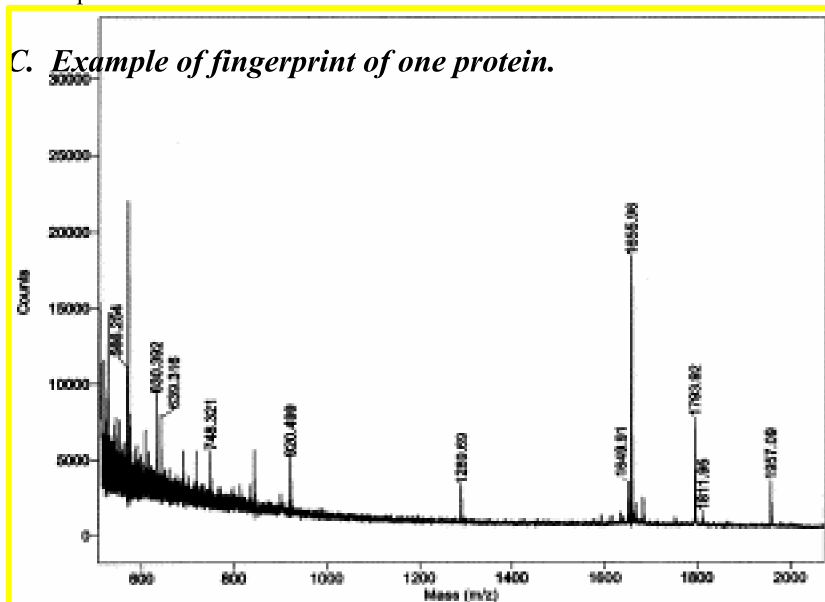
A. Schematic representation of the ionization process.



B. Schematic representation of MALDI-TOF



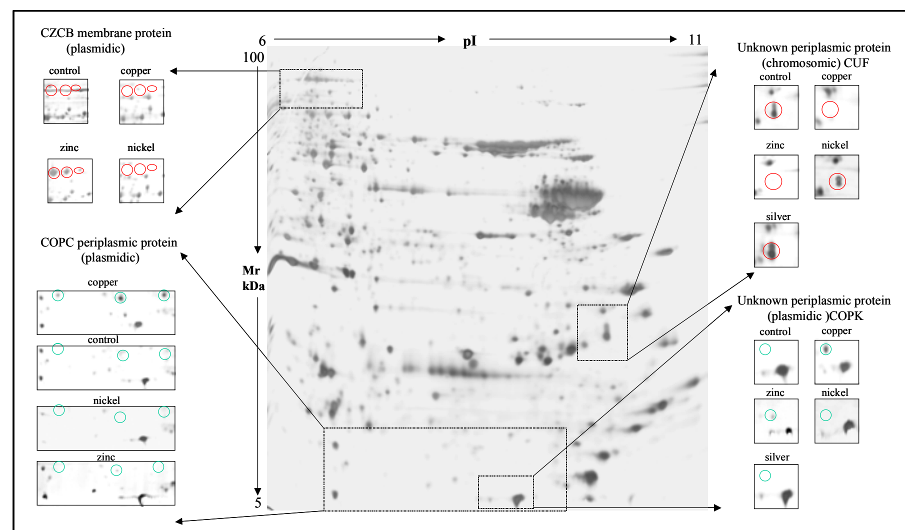
C. Output of MALDI-TOF/MS



2.2. Study of stress response and radiation sensitivity by mass spectrometry

The word phenotype was coined in the field of genetics, where it refers to the observable properties of an organism that result from the interaction between its genetic state (genotype) and its environment. The concept of phenotype can usefully be extended to proteomics. A proteomic phenotype is a phenotype that can be observed using proteomic methods such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry. A proteomic phenotype is determined by the state of the biological system under analysis ; genotype and environment can both modify proteomic phenotypes. Using proteomic methodology, several components of proteomic phenotypes can readily be measured: induced proteins, repressed proteins, modified proteins, unresponsive proteins. To date, various proteomics techniques exist (reverse phase capillary HPLC, electrospray ionisation mass spectrometry, gas chromatography-tandem mass spectrometry, MALDI-TOF) that offer new and refined developments to study and characterize stress response or radiation sensitivity of different bacterial species.

The analysis of a proteome involves the resolution of the proteins in a sample followed by the identification of the resolved proteins. 2D PAGE followed by mass spectrometry is the most widely used method of protein resolution and identification. However, now multidimensional chromatography must be used to achieve the resolving power of 2D PAGE. In 2D PAGE, proteins are separated in one dimension by isoelectric point (pI) and in the other dimension by molecular weight. As a result, a single 2D-PAGE system (2D map) can resolve more than 5000 proteins. This allows differential expression studies at the protein level by comparison between two different 2D-Maps (figure 4). In this context, analyze proteins may originate from **total bacterial pellet (TBP)** or **from secreted proteins (secretome SP)**.



Figue 2.4.: Example of a differential expression study from total bacterial pellet.

Mass spectrometry combined with novel bio-informatic approaches provides a powerful new strategy for the rapid speciation and typing of microorganisms. This revolutionary concept of Bacterial "Mass-Fingerprinting" offers greater sensitivity,

selectivity and speed of analysis as compared to classical identification methods that are currently used in clinical microbiology, food science, biotechnology, water quality and pharmaceutical analysis. Recent works show that this technology could be used to possible modifications after environmental variations.

Intact cell MALDI-TOF-MS (ICM-MS) applies established biopolymer Mass Spectrometry technology to the analysis of intact bacteria. This method allows the rapid sampling of a population of macromolecules expressed on the surface of bacteria and characterisation of these molecules by molecular weight. The resulting mass spectrum provides a unique physico-chemical fingerprint for the species tested. Mass-Fingerprints of unknown species can be reliably matched against databases of quality controlled reference mass spectra leveraging this simple analytical method into a powerful new tool for real-time detection and sub-typing of bacteria.

Mass-Fingerprinting is very rapid - the entire process from sample preparation to result takes only a few minutes for each test microorganism. Sample preparation is quick and easy - intact cells from primary culture are smeared across a stainless steel target plate and allowed to co-crystallise with a UV-absorbing matrix. After drying, the target is placed into the MALDI-TOF mass spectrometer. The microorganisms in the matrix are illuminated with a pulse from a nitrogen laser (337 nm). The matrix absorbs energy from the laser and macromolecules from the surface of the microorganisms are desorbed and ionised. The resulting ionised macromolecules are mass analysed and the results reported as a mass spectrum, i.e. a plot of mass (X axis) versus abundance (Y axis).

The Mass-Fingerprint of the test microorganism is then submitted to the MicrobeLynx™ search algorithm, which challenges an appropriately selected database from a range of quality controlled bacterial reference mass spectra. This technique could be used to characterize possible modifications after environmental variations (oxidative stress, low radiations,….) of the different Melissa bacteria's (Bright et al., 2002). Each bacterial Melissa compartment will be characterized by a specific fingerprint .

2.3. Membrane or surface accessible proteins detection (SAPD)

Flow cytometry has great potential as a rapid, automated tool for ecological studies of microorganisms. It is used for detection and characterization of aquatic bacteria, analysis of bacterial populations present in soils and sediments, detection of bacterial food contamination, analyses of intestinal flora and drinkable water. Flow cytometry can detect and separate specific populations of bacteria by using the cell sorting. Using an antibody against a specific surface bacterial protein allows to detect contaminants in a specific culture, example a Melissa compartment. The proteomic approach is one of the best approaches to identify a typical surface protein for each Melissa bacterial culture. In the first step, bacterial surface proteins are labeled by a specific surface probe (Sabarth, 2002). After extraction, these proteins are separated by two-dimensional gel electrophoresis and identified by mass spectrometry.

2.4. General procedures

The protein concentration will be estimated by the classical Bradford technique using BSA as standard. This procedure will be used to determine the required volume to the SCM-MS technique.

2.4.1. Short term and long-term storage procedures

- ICM-MS and SAPD techniques: A short-term storage is available at 4 °C.

- TBP-SP techniques: Short and long-term storages are available at –20 °C or –80C if the samples are prepared as indicated above.

2.4.2. Target selection/Validation

Which molecules are being targeted?

Flow cytometry can detect and separate specific populations of bacteria by the application of cell sorting. Using an antibody against specific surface bacterial protein would improve the detection of contaminants in a specific culture, for instance in a Melissa compartment. The proteomic approach is one of the best approaches to identify a specific surface protein for each Melissa bacterial culture. In the first step, bacterial surface proteins are labelled by a specific surface probe (Sabarth, 2002) . Bacteria were suspended in ice-cold PBS (PBS : Phosphate buffer saline), 1mM $CaCl_2$, 0.5mM $MgCl_2$ at an optical density at 600nm of 2.5-3.5. The bacteria were surface-labeled by incubation with 200uM (final concentration) sulfosuccinimidyl-6-(biotinamido)-hexanoate [s-NHS-LC-biotin (Pierce)] for 30 min on ice. The reaction was stopped by adding two volumes of TNKCM (50mM Tris , pH 7.4, 100mM NaCl, 27 mM KCl, 1mM CaCl2, 0.5 mM MgCl2). After 10 min incubation at room temperature, the bacteria were sedimented by centrifugation by centrifugation at 3500 g for 10 min and washed three times with TNKCM. The biotin surface proteins are purified by affinity chromatography on reversibly binding avidin-agarose (Boehringer). After extraction, these proteins are separated by two-dimensional gel electrophoresis and biotinylated proteins were detected on the blots using NeutrAvidin-peroxydase staining (Pierce) and chemiluminescent visualization (ECL, Amersham). Proteins will be identified by mass spectrometry. The specificity of these surface proteins can be further analysed by comparison with the genes included in the different genomic and proteomic databases.

Validation of analytical methods

a) TBP and SP techniques:
So far, this technique has been validated to characterize and identify different proteins up or down regulated after different stress conditions on many bacterial species. However, in order to fulfill the demanding requirements of proteome analysis (resolution, reproducibility, ….), several conditions will be tested for each bacterial . These conditions will be : the protein extraction method , the pH gradient in the first dimension and the gel reticulation of the second dimension.

b) ICM-MS technique:

A possible study by this technique will be validated pilot study. In the first step, a specific MS fingerprint of *Arthrospira* cultures will be characterized. After centrifugation (5000 rpm during 30 min at 4 °C), the pellet will be laid directly with a Maldi-tof matrix on the sample plate and the measure will be done. A characterisation of intact bacterial fingerprints will be done in different culture time points.
In the second step, the *Arthrospira* cultures will be submitted to different environmental stress (heat shock, oxidative stress,…) and a MS fingerprint will be characterized. This study will allow us to assess the sensitivity and the reproducibility of this method.

c) SAPD technique :
This technique has been validated for a large number of bacteria (Sabarth, 2002). Briefly, In the first step, bacterial surface proteins are labelled by a specific surface probe (Sabarth, 2002). After extraction, these proteins are separated by two-dimensional gel electrophoresis and identified by mass spectrometry. In this context, a specific surface membrane two-dimensional gel electrophoresis from the different Melissa strains will be realized. After separation, surface proteins will be characterized by Edman degradation or mass spectrometry techniques.

3. Description of biodiversity

3.1. Outline and principle

The classic approach to enumerate bacteria from the environment have been culture-dependent techniques combined with a simultaneous or subsequent differentiation of the isolates based on batteries of physiological and biochemical tests. Together with the description of the morphology of the bacteria, this could not offer a straightforward strategy to identify and classify micro-organisms. Next to the ambiguous character of the extracted data, culture-dependent methods do not accurately reflect the actual bacterial community structure (Theron and Cloete, 2000). Therefore it was necessary to develop various recombinant DNA techniques for the identification of bacteria that are independent of cultivation methods.

Two main types of molecular techniques have been developed. The first molecular approach investigates parts of DNA by focusing on genome sequences which are targeted and amplified by PCR (f.e. 16S rRNA). Another approach will try to investigate all the genetic information in the extracted DNA (Ranjard et al., 2000). In this project, however, we will concentrate on the approach based on the identification of micro-organisms via 16S rRNA gene sequences, determined directly from the biomass without further cultivation. The rRNA molecules comprise highly conserved sequence domains interspersed with more variable regions. Consequently comparative analysis of rRNA sequences can identify so-called signature sequence motifs on various taxonomic levels that are targets for an evolutionary-based identification. rRNA molecules offer a number of advantages to be used as targets for several reasons: they are functionally conserved molecules in all organisms; the primary structures of 16S and 23S rRNA molecules are composed of sequence regions of higher and lower evolutionary conservation, 16S rRNA sequences have already been determined for a large fraction of bacterial species, and their natural amplification with high copy numbers per cell greatly increases the sensitivity of the method (Amann, 1995). Furthermore, their sequences can be retrieved from the samples of interest without prior cultivation (Theron and Cloete, 2000).

3.2. Amplified 16S and 16-23S ribosomal DNA gene sequences and restriction endonuclease analysis (ARDRA) (Massol-Deya et al., 1995)

ARDRA can be used for analyzing mixed bacterial populations. If the rDNA fingerprints for individual bacteria in a community are sufficiently different, then one can examine the amplified products for a series of distinct patterns resulting from the different populations that make up the community. This method for ribosomal DNA fingerprinting of communities could hence be used for a quick assessment of genotypic changes over time (Massol-Deya et al., 1995).

This method involves the use of a pair of universal priming sequences for the PCR amplification of either the 16S rRNA genetic loci or the intergenic regions of the 16S and 23S rRNA genes. The latter regions exhibit a large degree of sequence, length and frequency variation because the spacer region is not well conserved. ARDRA using the 16-23S rDNA region will provide a more diverse template for restriction analysis resulting in more diverse template for restriction analysis resulting in more complex band patterns and potentially higher resolution of community members. This version of ARDRA may be useful in communities that are composed

of closely related populations. In highly diverse communities, the pattern may turn out to be too complex or there may be too little resolution of the distinct bands. If this is the case, an alternative may be to use group specific primers for a given bacterial group or to carry out the DNA amplification under more stringent conditions. Further, resolution can also be obtained by hybridization with DNA probes specific for the target groups (Massol-Deya et al., 1995). The limiting step is the fractionation step. The separation of DNA fragments derived from different populations requires that they differ in sequence at the restriction endonuclease sites, or differ in length of DNA flanked by common restriction sites. The separated DNA fragments may also be transferred to filters for hybridization with probes specific for an organism of interest, subjected to Southern transfer (Theron and Cloete, 2000).

Advantages of the method include: detection of community members present in very small numbers (1-10 cells); monitoring population dynamics without drastic disturbance by sampling; free of bias of culture dependent methods. The **disadvantages** are: ARDRA detects only the dominant species; cannot unequivocally distinguish between two closely related strains (Massol-Deya et al., 1995).

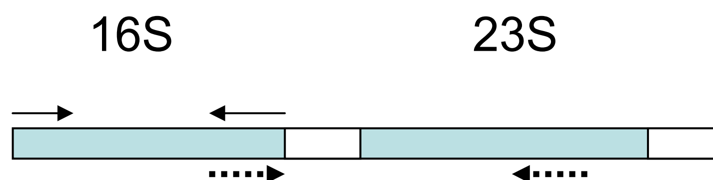


Figure 3.1. General schematic representation of the rRNA operon showing the localization of primer binding sites for PCR amplification. White fields indicate spacer regions; filled arrows indicate the annealing sites of the 16S primer pair; dotted arrows indicate the annealing sites for the primers amplifying the 16-23S rDNA intergenic region.

Highly conserved flanking sequences are used for primer binding sites (arrows) for the amplification of 16S ribosomal genes (Fig. 5.). For the 16-23S rDNA intergenic region, amplification is carried out by using a pair of an opposite highly conserved 16S and a 23S ribosomal primer (dotted arrows). Afterwards, the intragenic PCR product will be digested and the pattern will be analysed by gel electrophoresis. In order to ameliorate the ability to distinguish among different populations separate digestions with three different restriction enzyme digestions is recommended (Massol-Deya et al., 1995).

3.3. Analysis of bacterial communities using whole cell *in situ* hybridization (Amann, 1995)

Analysis by whole cell *in situ* hybridization does not only reveal the identity of the bacteria at single cell level, it can also visualize cellmorphology or the spatial distribution of the cells within the bacterial community (Theron and Cloete, 2000). During whole cell *in situ* hybridization morphological integrity has to be maintained under hybridization conditions. Cell walls and membranes have to be permeabilized to allow penetration of the oligonucleotide probe. After hybridization of the probe to the target DNA and several washing steps, the sample can be viewed by epifluorescence microscopy. Probes will only bind correctly under defined and optimized hybridization and washing conditions (Theron and Cloete, 2000).

A set of group-specific oligonucleotide probes for major phylogenetic sublineages of micro-organisms allows a crude characterization of the community structure in a certain ecosystem (Amann, 1995). Using top down *in situ* hybridization more and more detail of the bacterial community can be revealed. Although this is an indirect approach to identify the members of a bacterial community, it can provide quantitative as well as qualitative information of the different groups of bacteria very quickly. On the other hand does the method require expensive tools like the confocal laser scanning microscope connected to PC's programmed to analyze the obtained images automatically and subsequently process the output data in order to be use the method of whole cell *in situ* hybridisation most advantageously.

3.4. Single stranded conformation polymorphism (SSCP) (Hawkins, 1997)

The mobility of single-stranded or double stranded DNA electrophoretically run through a gel matrix depend on their size as well as on their tertiary structure. Small molecules pass through the pores in the matrix more easily than large molecules and hence migrate faster. In order to run single-stranded DNA a denaturant is needed that maintains the single-strandedness of the DNA fragments. Running single stranded DNA on a nondenaturing gel will permit intramolecular interactions to occur. In other words, the single-stranded DNA is able to bind partially to itself. As DNA is not running as a linear molecule, the mobility of the DNA is governed by both its size and tertiary structure. This tertiary structure of a single-stranded DNA fragment is dependent on the sequence of the entire fragment. Any sequence differences in a given fragment will result in a different conformation and hence a different pattern. Alas, the method is very sensitive to the gel temperature, cross-linker concentration, acrylamide concentration, presence of glycerol and the type of the gel matrix. It is therefore difficult to guarantee that a replicate of the same SSCP investigation will result in an identical banding pattern.

3.5. Terminal restriction fragment length polymorphism (TRFLP)

The TRFLP method is based on the ARDRA method. But while ARDRA is of limited use for demonstrating the presence of specific phylogenetic groups or for estimating species richness, the technique of TRFLP can give much more specific insight in the bacterial community structure. The initial steps of DNA isolation, PCR amplification and restriction are similar to ARDRA. However, in TRFLP one of the primers used is labeled with a fluorescent dye so that when the preparation was analysed with an automated DNA sequencer, the sizes of only the terminal restriction

fragment (T-RF) could be determined and the amount can be quantified. Using TRFLP on 16S rDNAs, the method can be adopted for analyzing complex microbial communities (Liu et al., 1997). However, the method does not directly identify the members of the bacterial community as it is possible for the DGGE method. Still, a growing database of T-RFLP's and the use of controls can give a clue on the identity of the microorganisms residing in the community.

3.6. Denaturing gradient gel electrophoresis (DGGE) (Muyzer et al., 1998) or temperature gradient gel electrophoresis

In DGGE and TGGE DNA fragments of the same length but with different base-pare sequences can be separated. Separation in DGGE is based on decreased electrophoretic mobility of a partially melted DNA molecule in polyacrylamide gels containing a linearly increasing gradient of DNA denaturants (a mixture of urea and formamide), while separation of the fragments in TGGE is accomplished by a temperature gradient. Melting of the DNA fragments proceeds in discrete so-called melting domains (base pairs with an identical melting temperature). Once a melting domain reaches its melting temperature at a particular position in the denaturant gradient gel, a transition of helical to partially melted molecules occurs, and migration of the molecule will stop. Sequence variation within such domains causes their melting temperature to differ. Sequence variants will thus stop migrating at different positions in the denaturing gradient and be separated, thereby enabling the analysis of a mixed microbial community (Muyzer et al., 1998).

Although DGGE analysis of PCR-amplified 16S rDNA fragments provide a rapid method to characterize community population structure, more specific information of population composition can be obtained by secondary analysis of the DGGE banding pattern. Individual bands may be excised and sequenced. Alternatively, DNA can be transferred to nylon membranes and hybridized with group- and species-specific oligonucleotide probes to identify specific populations within the microbial community (Theron and Cloete, 2000).

This method of describing bacterial diversity is straightforward and will give a rough idea of what bacterial species are present in the sample. Therefore this project will concentrate on DGGE as the primary method to reveal the most dominant species in the first compartment.

3.7. Oligonucleotide fingerprinting of rRNA genes (ORFG)

The recent advances in DNA microarray technology provide for a method were thousands of genes can be simultaneously assessed by using a large set of probes, miniaturized on one glass slide. It is likewise an ideal format to assess the sequence diversity of 16S rRNA in natural environmental samples (Koizumi et al., 2002). In this approach labeled rRNA or rDNA from environmental samples are analyzed by hybridization to oligonucleotide probes attached to a substrate. This method, however, had only limited success. The most significant problem was based on probe design next to and in conjunction with hybridization conditions.

The method of oligonucleotide fingerprinting of rRNA is an alternative array based approach in which the rDNA, not the oligonucleotide probes, are attached to a solid substrate. ORFG is an adaptation of a method used for gene expression profiling, where clone libraries are constructed from rDNA molecules that have been PCR amplified from environmental DNA. The rDNA clones are then arrayed on nylon membranes and subjected to a series of hybridization experiments, each using a single DNA probe 10 nucleotides long. For every probe, the signal intensities reflect negative, positive, or uncertain hybridization events. This creates a hybridization fingerprint for each clone. Subsequently the clones are identified by clustering their hybridization fingerprints with those of known sequences, as well as by nucleotide sequence analyses of representative clones within a cluster (Valinsky et al., 2002).

4. Investigation of critical genes by DNA chip technology

4.1. Outline and principle

A significant technology that has emerged is the use of miniaturized DNA hybridization systems: DNA microarrays or DNA chips. They allow simultaneous detection of expression of thousands of genes, and reveal the manner in which an organism uses its genetic arsenal under various conditions (Stephens, 2001).

An entire microbial genome can be easily represented in a single array, making it feasible to perform genome-wide analysis (DeRisi et al., 2001). Two of the most used applications for DNA chips are the investigation of genome-wide transcriptional profiles and the measurement of the similarities or differences in genetic contents among different microbes (Ye et al., 2001).

DNA microarrays are basically a miniaturized form of dot blot in a high-throughput format. There are two major types of DNA chips: the oligonucleotide-based array; and the PCR product-based array. The outline of an DNA-chip based experiment is shown in fig. 4.1..

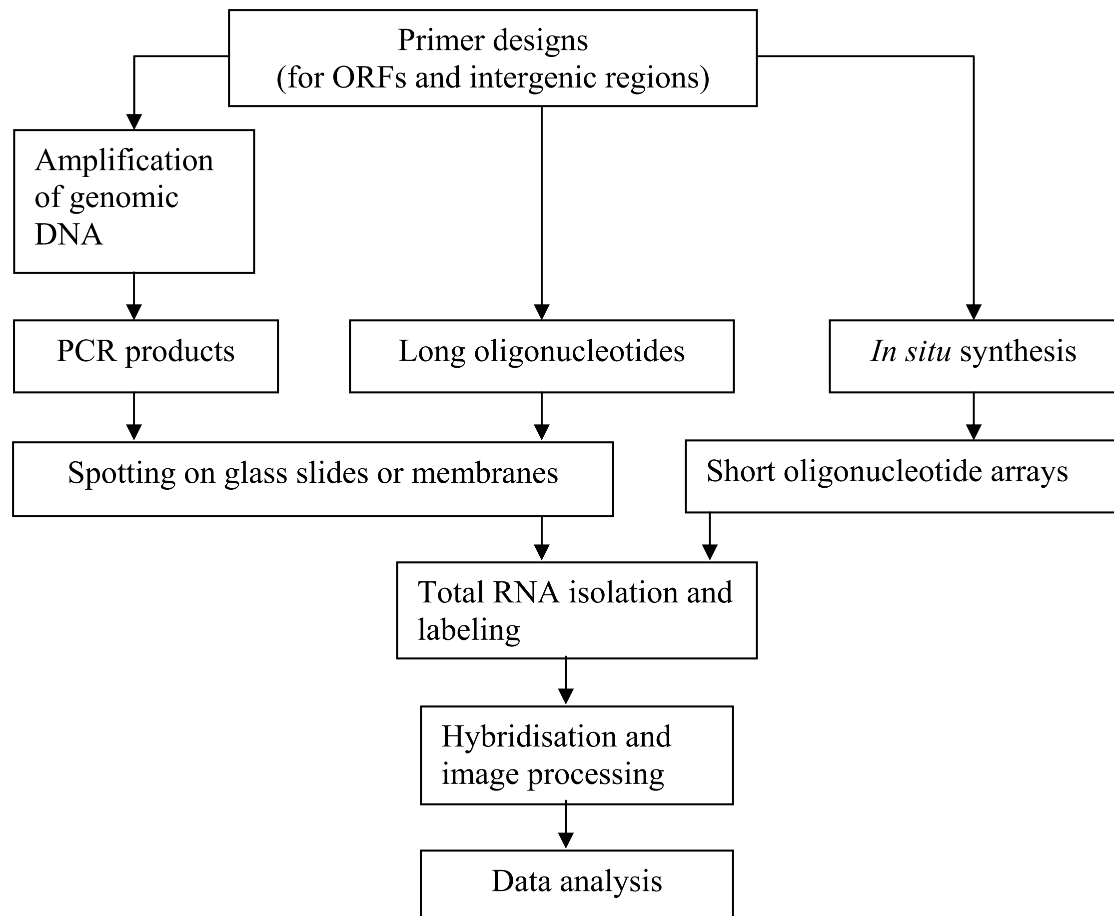


Fig.4.1. Construction of DNA microarrays. (Ye et al., 2001)

4.2. Experimental approach

The possible stress resulting from both space conditions and effluents from the various compartments will be evaluated on the bacterial microflora of MELISSA loop. Indeed, these conditions can have an effect on the physiology of the strains and alter the functioning of the whole system. With this aim, the expression level of genes known to play an important role in the resistance to stressful conditions.

The expression of genes in bacteria is regulated by environmental conditions. Hence variations in environmental stimuli will result in differential expression of the genes. Until recently only a limited number of genes could be analyzed. Therefore it was impossible to obtain a complete overview of the induction or repression of gene expression under certain environmental conditions. Now, simultaneous detection and quantitative estimation of all transcripts in a micro-organism or tissue was made possible by the invention of arrays with DNA fragments representing an entire genome or a subset thereof (Rimini et al., 2000).

This procedure will require the prior identification of such genes in the strains of interest. First, the protein sequence of homologous genes present in the databases and identified in various microorganisms, including *Rhodospirillum rubrum, Rhodobacter sphaeroides* and *Synechocystis* spp. that are closely related to some MELISSA strains will be aligned to identify conserved domains in all these proteins. Second, the nucleotide sequence of the corresponding genes will be aligned to allow the design of degenerate oligonucleotides on the basis of the conserved regions in their coding parts.

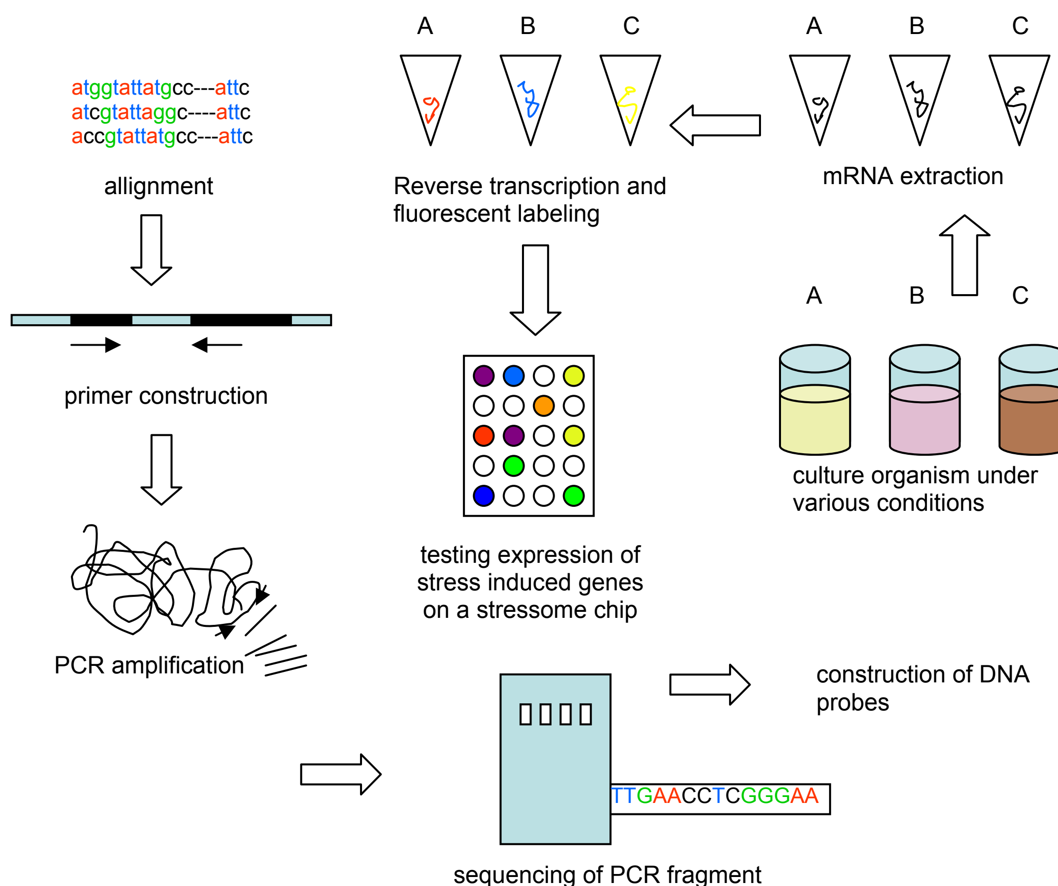


Figure 4.2. Overview of the approach to investigate induction of stress related genes.

Third, these oligonucleotides will be used to amplify by PCR the genes of interest in *Arthrospira platensis* and in *Rhodospirillum rubrum*. These PCR amplified DNA fragment will then be used as probes to monitor the level of gene expression from mRNA preparations of the various strains. Alternatively, the physiological alterations resulting from specific stresses can be investigated in a global way with the so-called DNA chips, which are glass slides where all the genes of an organism can be spotted. For this purpose, DNA chips will be prepared from DNA genomic libraries of MELISSA strains.

The procedure for transcriptome (i.e. the whole transcript set of an organism) analysis is as follows: messenger RNAs are extracted and DNA complementary to these RNAs are synthesized by a reverse transcriptase enzyme. During this process, cDNAs are labelled with two different fluorescent compounds. For each strain, cDNA pools corresponding to two different conditions can be hybridized on the same support and analysed in one step by specific software and statistical methods. This allows a direct comparison of the differential expression rate of all the genes of the organism of interest.

5. Genomic pattern analysis by PCR-based techniques

5.1. Outline and principle

For prokaryotic species that can be cultivated, numerous molecular methods exist to obtain genetic information. These methods largely depend on four levels: i) the genome (DNA/RNA), ii) proteins, iii) cell components (fatty acids, lipids, etc.), and iv) morphology and behaviour. The complexity of interactions of gene product increases from level i) to iv), and accordingly the genetic relatedness of microorganisms is more difficult to deduce (Fig. 5.1.).

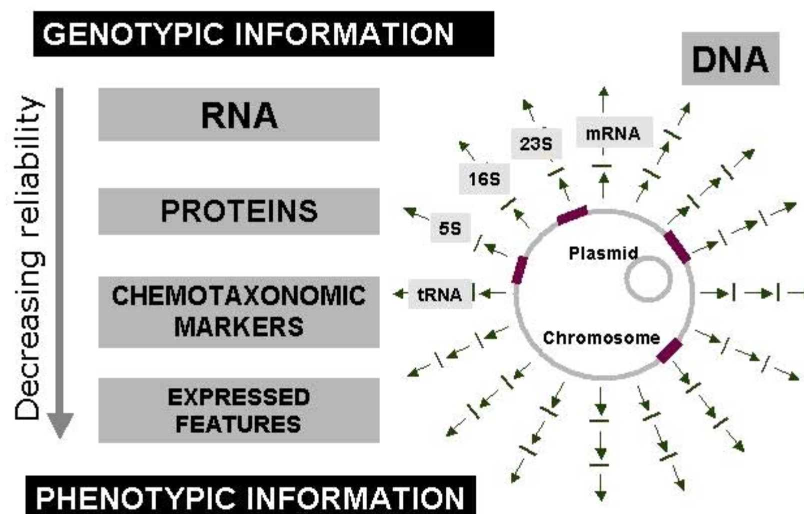


Fig. 5.1.: An overview of taxonomic information provided by biological molecules

Nucleic acid-based analysis methods, often referred to as 'genotypic methods', have become increasingly important to address the actual state of genetic relatedness within groups of prokaryotic organisms (reviewed by Schloter *et al.*, 2000). Consequently, the scientific community has been confronted with an avalanche of new genotypic fingerprinting techniques, often with confusing or overlapping names and terminologies (reviewed in detail by Vaneechoutte, 1996). The common theme in all these methods is the detection of naturally occuring DNA polymorphisms (point mutations, insertions, inversions, deletions) which are a result of evolutionary processes that take place in the cell. In the past few years, general approaches have been established that allow the detection of single nucleotide polymorphisms (SNPs), but these approaches require high density microarrays or novel sequencing procedures and it is not our immediate aim to detect SNPs, but rather to study overall genetic stability in terms of differences in DNA banding patterns using restriction enzymes and PCR.

5.2. Genotyping methods

The development of genotypic methods can be divided in a number of phases. In the 1970s, analysis of plasmid content and the restriction patterns of plasmid DNA was used, but this was quickly superseded in the early 1980s by the realisation that plasmids have a potential transient nature, and methods that examined the whole chromosome emerged. These included total chromosomal restriction digests, RFLP with probes (i.e. on rRNA encoding genes – 'ribotyping'), followed by more advanced methodologies such as PFGE, RAPD, REP-PCR, and AFLP (see also section 5.5). In the last few years, primarily owing to a leap in the efficiency, throughput, and automatisation of various DNA sequencing technologies, methods such as multilocus sequencing (MLST) and microarray analysis have gained ground. The majority of these genotypic methods are able to address, with various degrees of reliability, reproducibility, and resolution, genetic variability at the strain level (Fig. 5.2.).

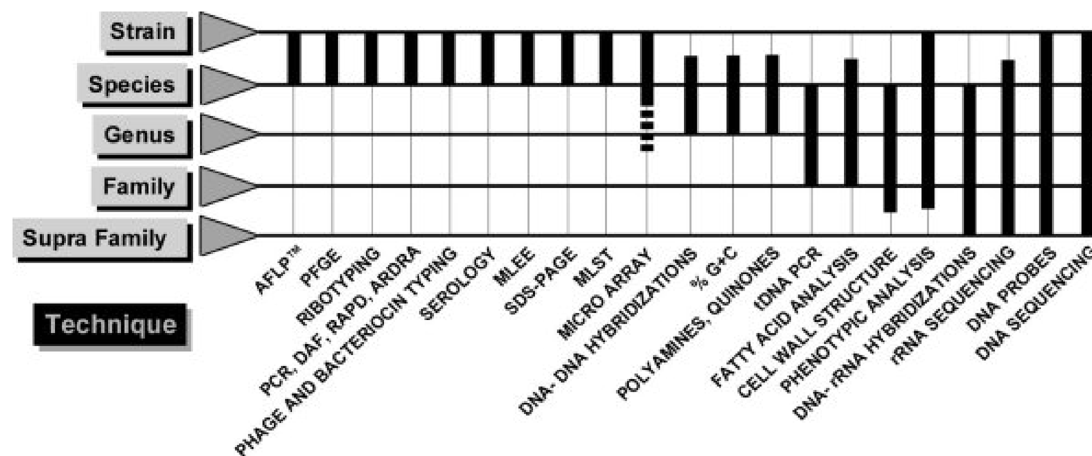


Fig. 4.2.: Taxonomic resolution of current genotypic and phenotypic methods (abbrevations as in section 4.5.)

In addition, whole-genome sequencing has provided a wealth of genetic information and has enabled full-sequence comparison between a number of strains of the same species (e.g. *E. coli*, *S. aureus*, *B. anthracis*, etc.). As of June 2002, 64 eubacterial and 15 archaeal genomes have been sequenced and published (Janssen *et al.*, 2002).

From the above overview, it can be seen that none of the common phenotypic typing methods offer an ideal approach for the sub-dividing of microbial species, let alone the monitoring of genetic variance in microbial populations or the study of genetic stability. In contrast, genotypic methods have the potential to be used for studying diversity in any microbial species, with some genotypic methods also offering the possibility of providing a 'universal' approach (i.e. through DNA sequence independency). One of the most promising techniques is selective restriction fragment analysis (SRFA), in which the entire genome is digested with restriction enzymes and a subset of tagged fragments is selectively amplified using universal PCR primers. The best known SRFA method is AFLP™ , originally developed at Keygene International (Wageningen, The Netherlands) for the monitoring of traits in plant and animal breeding (Vos *et al.*, 1995). The technology was successfully adapted for microbial analysis (Janssen *et al.*, 1996), and since then has been used for

the genotypic characterisation of fungi, yeast, and a large variety of prokaryotic species (reviewed by Janssen, 2001).

5.3. Polymorphism analysis by AFLP

5.3.1. The AFLP concept

Strictly speaking, and based on the technical layout, AFLP should be classified under the category of Selective Restriction Fragment Amplification (SRFA) techniques employing short double stranded 'adaptors' or 'indexers' that are ligated to genomic restriction fragments and which serve as primer binding sites for amplification (Vaneechoutte, 1996). The AFLP concept can be largely divided into three steps (Fig. 5.3):
(1) digestion of total cellular DNA with two restriction enzymes and ligation of restriction halfsite-specific adaptors to all restriction fragments,
(2) selective amplification of some of these fragments with two PCR primers that have corresponding adaptor- and restriction-site sequences, and
(3) electrophoretic separation of amplicons on a gel matrix followed by visualisation of the banding pattern.

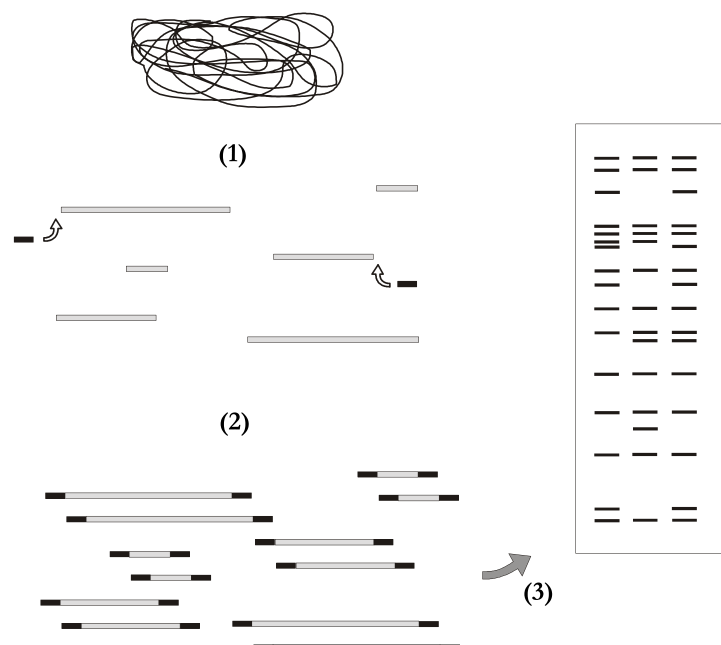


Fig. 5.3.: Main principles of the AFLP method: digestion of total genomic DNA (1), ligation of adapters (2), and amplification with adapter-specific primers and electrophoretic separation of amplicons (3) (Janssen *et al.*, 2001).

The simultaneous use of two restriction enzymes is beneficial. First, heterologous ends are produced, allowing ligation of adaptors with different sequences so that many primer pairs can be used on the same template. This facilitates the analysis of highly monomorphic genomes to a great extent (Keim *et al.*, 1997)

(note: AFLP has been used for strain identification of *Bacillus anthracis* in the recent anthrax scare in the US). Second, by using a mix of enzymes, only a limited number of fragments (due to the 6-base cutter) are obtained that are relatively small (due to the 4-base cutter) and amplification is thus straight forward. In addition, the resulting amplicons are in an optimal size range (50 - 1000 bp) for high resolution separation on denaturing 5 - 6 % (w/v) polyacrylamide gels.

5.3.2. Template preparation

Ligation of adaptors. After digestion of the genomic DNA, fragments are tagged with short double stranded DNA molecules that bind in a complementary fashion to the restriction half-sites. In the presence of T4 ligase and ATP, a phosphodiester link is created between the 3'-OH and 5'-PO_4 groups at the end of each adaptor and fragment (Fig. 5.4., step 1). Note that, because unphosphorylated oligonucleotides are used for the adaptors, there is only one phosphodiester link for each adaptor-to-fragment ligation (Fig. 5.4., step 2), meaning that if DNA was denatured prior to the addition of *Taq* DNA polymerase and dNTPs, no amplification would occur (Vos *et al.*, 1995). The PCR in AFLP works because *Taq* polymerase fills in the 3'-recessed ends or displaces the non-ligated strand at ambient temperatures (Vos *et al.*, 1995), thereby ensuring the presence of the originally intended primer binding site (Fig. 5.4., step 3).

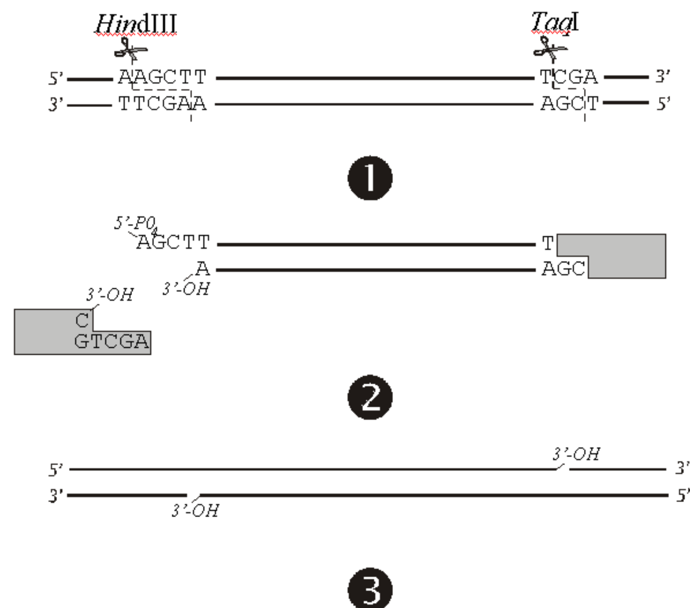


Fig. 5.4.: Preparation of template DNA for selective amplification (see text for details).

Adaptor features. When two restriction enzymes are used, adaptors have different sequences allowing a combination of PCR primers. Obviously, the core sequences of the adaptors are dictated by proper primer design, i.e., inverted repeats and long nucleotide runs should be avoided. Also, although the sequences must be sufficiently different to ensure specificity, the T_m of both primers should be in the

same range. Once the primer sequences are chosen, adaptors can be designed. An essential feature of AFLP adaptors is that they contain a base change so that the original restriction site is not reconstructed during ligation. In the example of Fig. 5.4., a *Hin*dIII-adaptor with a 3'-cytosine is used for this purpose. This has the advantage that the ligation can be done in the presence of restriction enzymes, thereby preventing fragment-to-fragment ligation and simultaneously stimulating adaptor-to-fragment ligation. In addition, because adaptors are made of unphosphorylated oligonucleotides, they are unable to ligate to each other. The combination of these features ensures that virtually all restriction fragments are tagged with the appropriate adaptor sequence.

5.3.3. Selective PCR

Principles. AFLP primers consist of the adaptor-derived core sequence, including the 3'-part of the restriction half-site, and an extension sequence made up by a number of so-called selective bases. These selective bases are complementary to nucleotides flanking the restriction sites. For a given template, elongation of the DNA chain will only take place if the primer binds with high specificity, e.g., if the corresponding complementary nucleotide is present in the fragment (Fig. 5.5.). In addition, exponential amplification is only achieved when DNA synthesis occurs from both ends. The stringency of primer binding is essential for the succes of reproducible AFLP (for review, see Janssen, 2001).

Primer structure. Obviously, proper primer design is very important. Hairpin structures should be avoided and the formation of self-dimers and cross-dimers should be kept to a minimum as much as possible to prevent the loss of reaction components. Also, it is preferred that a primer sequence does not contain long stretches of As or Ts, because this may cause local instabilities in the primer-template hybrid, and the G+C-content should range between 40 and 60%, allowing highly specific annealing between primer and template at 56 - 60°C. For AFLP primers, another feature is the obligatory presence of a 5'-guanine residue. This prevents the generation of so-called 'double bands' that were observed occasionally in older AFLP gels due to incomplete addition of an extra nucleotide to the synthesised strands. This terminal transferase activity, also known as 'extendase' activity, of the DNA polymerase is 3'-base dependent: if the 3'-base is a cytosine, extendase activity is quite strong and mainly adenines are added. Using AFLP primers with a 5'-guanine thus guarantees that most amplicons will be identical and have the same electrophoretic properties, leading to the formation of discrete bands and improving comparative analysis of banding patterns.

Selection criteria. Assuming random base distribution, one out of four fragments will be amplified for each selective base used. However, in reality, the final complexity of the AFLP pattern is determined not only by the number of selective bases used, but also by the choice of the selective bases and the base composition of the genome (reviewed by Janssen, 2001). In general, when properly executed, AFLP methodology deploying +1 or +2 primers (i.e. primers with 1 or 2 selective bases, respectively) ensures a proportional reduction in the number of amplicons for each selective base used.

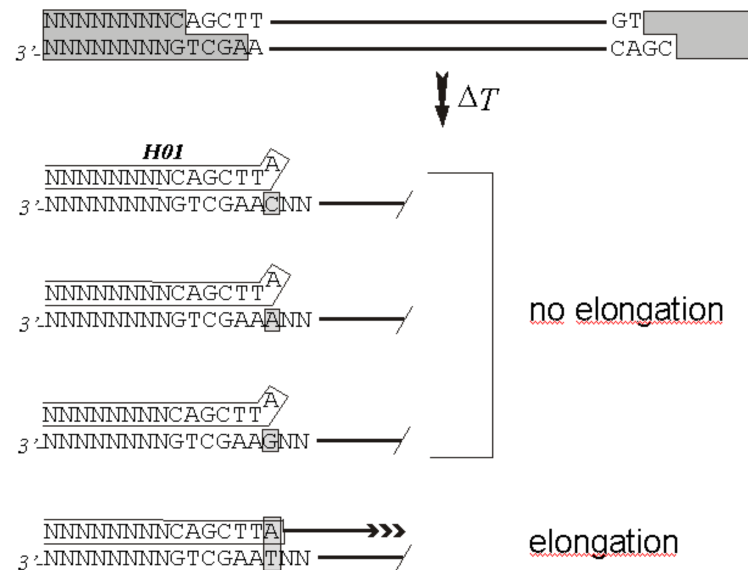


Fig. 5.5.: Selective amplification with AFLP primers; only a perfect match at the 3'-end between primer and template results in DNA chain elongation and subsequent PCR amplification.

5.3.4. Fragment separation

The original AFLP procedures described by Vos *et al.* (1995) made use of conventional sequencing gels containing 7 - 8 M urea and 4.5 - 6% cross-linked polyacrylamide. Typically, detection of AFLP amplicons required the labelling of one of the primers with either P-32 or P-33, followed by visualisation of the banding patterns by autoradiography or by the use of a phosphorimager (Fig. 5.3.). However, many research institutes prefer to limit the use of radioactivity, and methods for nonradioactive detection of nucleic acids have been put forward, e.g. by silverstaining (Chalhoub *et al.*,1997) and chemiluminescense (Lin *et al.*, 1999). A much better, but also much more expensive, approach involves the use of Cy5-labelled primers in combination with on-line laser detection of fluorescent amplicons while they pass through the gel (Fig. 5.6.). This now has been done for AFLP analysis of a large number of bacterial species (reviewed by Janssen, 2001).

If an automated apparatus is equipped with a multiple fluorescence detection system e.g. the 4200 NEN Global IR2 DNA Analyzer (LI-COR, Lincoln, NE) that is present in the group of Dr. Wilmotte (ULg), a separate fluorescent label can be used as an internal standard for addition to each sample, thus greatly improving the normalisation of the gels. Improved resolution is also achieved because the fragments are detected at a fixed distance from the origin, leading to a more uniform spacing between fragments. General advantages of 'automated' fluorescent AFLP (fAFLP) include the large simultaneous throughput of samples, a rapid turnaround, and the direct processing of raw data and subsequent error-free storage of results through linkage with a computer. The Li-Cor instrument that will be dedicated to our AFLP genotyping analysis has the added advantage that it is connected to our Local Aerea Net (LAN), facilitating remote analysis of the generated data using the ISIX software package.

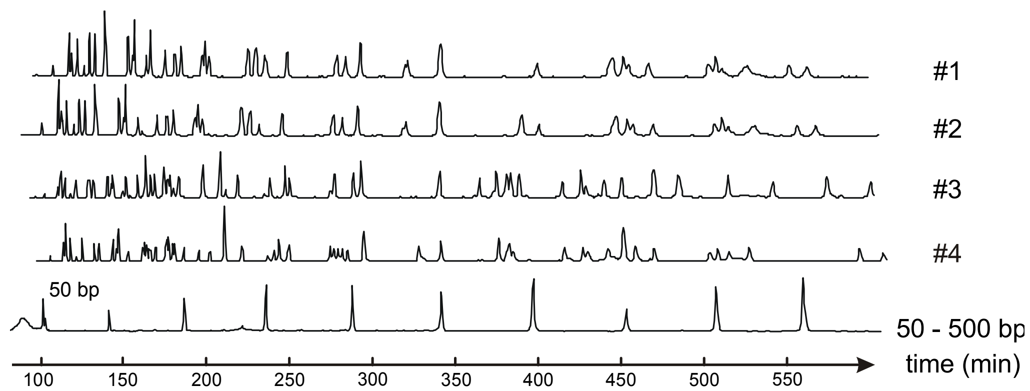


Fig. 5.6.: fAFLP electropherogram obtained using fluorescently labelled primers and an automatic sequencing apparatus.

5.3.5. Pattern visualisation

With fAFLP technology, fluorescently labelled amplicons are detected with a laser while they migrate downwards through the gel. The detection signals are collected, digitised and sent to the computer for storage and processing. These raw data are converted to the TIFF format with software that is included in the automatic sequencing apparatus package. The actual size of the image file depends on the conversion settings used, with a typical 7- to 10-fold reduction of the original vertical resolution (as defined by the run-time and signal detection interval).

5.3.6. Data analysis

Data analysis is done using the GelCompar Software package of Applied Maths, Kortrijk, Belgium. Digital images are normalized by aligning the internal AFLP reference patterns. Background signals are substracted using the curve-fitting algorithm supplied with the Gelcompar software. Different digital images (e.g. from different electrophoresis runs) can be combined by assigning one particular reference track as a standard and alligning all other reference tracks versus this standard (Fig. 5.6.). Bands will be scored and a similarity matrix will be created using the band-based Dice coefficient, S_D, equal to the ratio of twice the number of common bands in two compared patterns to the sum of all bands in both patterns (Sneath & Sokal, 1973).

5.3.7. Applications

AFLP as a taxonomic tool. The AFLP method has been used with great success in numerous taxonomic studies of various bacterial and fungal species, as well as plants. The use of AFLP as a novel taxonomic tool for bacterial species has been evaluated in detail (Janssen *et al.*, 1996).

Epidemiological typing of bacteria. AFLP has also been used in conjunction with other typing methods for the epidemiological typing of a wide scala of medically important bacteria, including *Acinetobacter*, *Aeromonas*, *Campylobacter*,

Mycobacterium, *Helicobacter*, *Staphylococcus*, *Streptococcus*, etc. (reviewed in Janssen, 2001).

AFLP for studying the molecular evolution of microbes. In nature, genetic diversity in prokaryotes is driven largely by a number of dynamic processes that enable them to react swiftly to changes in their environment. To accomplish this 'adapt-to-survive' strategy, microbes have a plethora of routes at their disposal to acquire beneficial, or eliminate superfluous, genetic material, and to 're-shuffle' genes that need to be expressed at short notice (Casjens, 1998; Arber, 2000). This structural plasticity of microbial genomes has been the subject of numerous investigations, although these studies focused on only one particular gene or set of genes, and reports on whole genome analysis in the context of evolutionary studies on microbes are very scarce (Arber *et al.*, 1995; Naas *et al.*, 1995). The AFLP method has great flexibility in that many different primer pairs (i.e. up to 16 for +1 primers) may be used on the same template. This means that large numbers of nucleotides distributed over the entire genome can be surveyed simultaneously. For instance, with an average of 60 bands for each of the 16 patterns, and given that 12 nucleotides (6 + 4 of the restriction sites and 2 of the +1 selective bases) are associated with each band, 16 x 60 x 12 or 11,520 nucleotides are examined for point mutations and, assuming an average fragment length of 250 bp, 240,000 nucleotides are surveyed for length mutation. This approach allows the detection of rare polymorphisms, is particularly interesting for the analysis of highly related genomes, and owing to its supurb resolution, is perfectly suited to follow the genetic stability of prokaryotic organisms.

6. Flow cytometry analysis

6.1. Outline and principle

In flow cytometry, the cells may be alive or fixed at the time of measurement, but must normally be in monodisperse (single cell) suspension. They are passed single-file through a laser beam by continuous flow of a fine stream of the suspension. Each cell scatters some of the laser light, and also emits fluorescent lights excited by the laser which are collected by photomultiplier tubes positioned at different angles relative to the incident light and 90° relative to the jet of the cells (figure 6.1.).

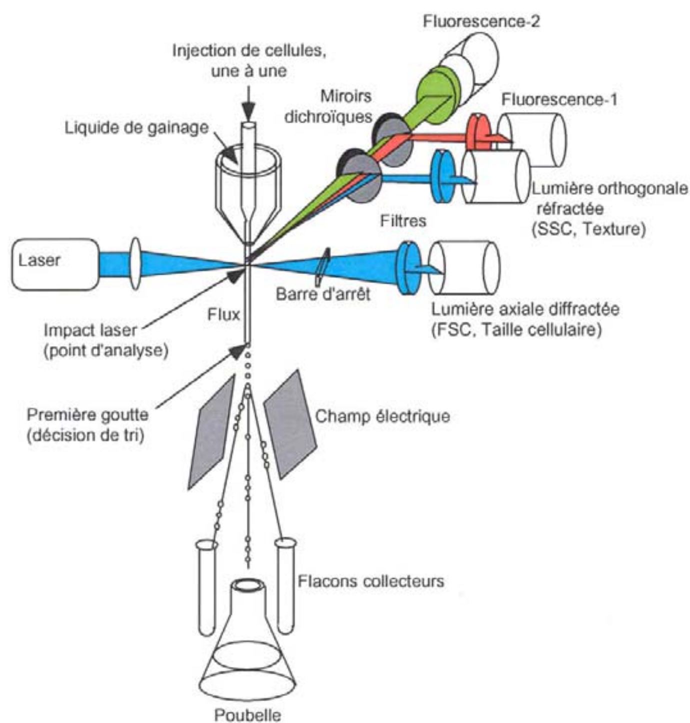


Fig. 6.1.: Individual cells confined within a rapidly flowing jet of water pass a measuring window, at which several parameters can be simultaneously measured (Fig 6.2).

The flow cytometer typically measures several parameters simultaneously for each cell (figures 1 and 2):

- low angle forward scatter intensity, approximately proportional to cell diameter
- orthogonal (90 degree) scatter intensity, approximately proportional to the quantity of granular structures within the cell

- fluorescence intensities at several wavelengths

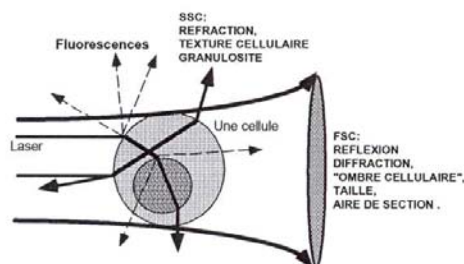


Fig. 6.2.: measurement of various parameters in flow cytometry

Light scatter alone is often quite useful. It is commonly used to exclude dead cells, cell aggregates, and cell debris from the fluorescence data. Fluorescence intensities are typically measured at several different wavelengths simultaneously for each cell. Different cell components are stained with fluorescent dyes and, as the cells pass the window, the dyes are excited with a high-intensity light source. The emitted fluorescence from the dye is proportional to the cellular content of the stained substance. In this way, DNA content (measured as fluorescence from stained DNA for example) can be measured at rates of more than 1,000 cells per second. The method yields unique information about the distribution of a parameter between the individual cells in a population, which provides a different and more informative result than methods in which average values are measured. Fluorescent antibodies are often used to report the densities of specific surface receptors, and thus to distinguish subpopulations of differentiated cell types. By making them fluorescent, the binding of viruses or hormones to surface receptors can be measured. Intracellular components can also be reported by fluorescent probes, including total DNA/cell (allowing cell cycle analysis), newly synthesized DNA, specific nucleotide sequences in DNA or mRNA, filamentous actin, and any structure for which an antibody is available. Flow cytometry can also monitor rapid changes in intracellular free calcium, membrane potential, pH, or free fatty acids.

Flow cytometers involve sophisticated fluidics, laser optics, electronic detectors, analog to digital converters, and computers. The computer records data for thousands of cells per sample, and displays the data graphically.

6.2. Analysis Equipment at SCK-CEN: the EPICS XL (Beckman Coulter)

The facility at the laboratory of radiobiology at SCK-CEN provides a Beckman Coulter EPICS XL for analysis of cell samples. This instrument cannot separate and thus sort cells into different containers based on their properties; the samples are thus consumed and discarded during analysis. The EPICS XL is a closed fluidic system, so use with biohazardous samples (such as human blood samples, bacteria) is possible with appropriate precautions.

The EPICS XL uses an air-cooled argon gas laser, 15 mW output, with a fixed wavelength emission of 488 nm. It has four fluorescence detection channels which simultaneously detect green, yellow-orange, and red light. Fluorescein is used extensively for the green channel, and phycoerythrin or propidium iodide (a DNA stain) for the yellow-orange channel. Dyes are also available which can be excited at 488 nm yet emit in the red.

The EPICS XL can analyze cell suspensions at the rate of several hundred cells per second. Typically, investigators acquire 5,000 to 15,000 cells per sample. Data are saved to the hard disk of a dedicated computer, where they can later be analyzed with graphics software.

6.3. Optimised detection of bacterial changes induced by exogeneous factors.

6.3.1. Optimisation of the flow rate of sample cell inside the flow cytometer

The quality of the results depends on the cell flow rate expressed as cells per minute that passes through the flow chamber inside the flow cytometer. When the rate is too high, the sharpness of the peaks decreases. When the rate is too low, the experimental time is too long. An optimal cell flow rate needs to be determined emperically.

6.3.2. Discrimination of bacteria from the background

To decrease to the minimum the level of background noise, the media that are used in order to analyse bacterial populations will have to be carefully filtered (0.22 µm, 1, 2 or 3x). Then, the discriminator has to be set on size (if cells are not stained), on DNA fluorescence (if the DNA of the cells is stained) in order not to visualise background.

6.3.3. Basic detection of differences in bacterial size and granularity

Some of the most important work in marine microbiology, using FCM includes the ability to detect, count, sort, and identify free living bacteria to help better understand their role in the marine ecosystem. FCM has shown that is has the ability to identify pico-phytoplanktonic cells. These particular organisms present a great challenge, due to their tiny size 0.2 to 2 µm, even the answers to basic questions about the organism are elusive. However, with natural seawater samples flow cytometry enables the discrimination of pico-organisms from larger phytoplanktonic cells, photosynthetic prokaryotes, bacteria, and nanoheterotrophs. The interest in these organisms is founded by their role as primary producers and as a food source. Using rRNA oligonuclotide probes, the organisms can be subjected to *in situ* hybridization and based on their different degrees of side scatter the organisms are readily detected. The FCM histograms can be directly examined to aid in bacterial determination.
There is a relationship between the forward scatter (size) and cell mass. In addition, the scatter signal from bacteria depends on its shape and on the range of scatter angles detected. As bacterial populations are morphologically heterogeneous, we will investigate to which extend results may vary according to the composition (possibly taking into account age and density) of reference populations. Furthermore, a study

will be performed in order to analyse whether a change in the bacterial culture (due to viral infection, specified stress,..) could be correlated with a change in size and granularity. Finally, a possible relationship between the side scatter (SSC) and bacterial morphology will be investigated. Visual evaluation is to be performed with phase contrast microscopy. All the changes will be quantified by the use of calibrated microspheres of known diameter size (0.1, 0.2, 0.5, 1, 2, 5 and 10 µm diameter) and a calibration curve of the mean channel versus the size will be established. If there is heterogeneity of the bacterial size, it should be reflected by a high variation coefficient.

Below are two tables taken from Environmental Monitoring of Bacteria that display the different types of fluorescent dyes that may be used for flow cytometry labeling. Flow cytometry can a mass a variety of data types. The possibilities for analysis types are innumerable, with different fluorochromes available to do a variety of experiments (tables 1 and 2). Investigations include many environmental samples especially aquatic samples but flow cytometry is also applicable to other fields as shown in table3.

Table 1
General Fluorochromes Used to Label Bacteria for FCM

Fluorochrome	Target molecules
Fluorescein and tetramethylrhodamine	General fluorescent labels, via isothiocyanate group; e.g,. total cell protein, antisera, oligonucleotides
Phycoerythrin	Conjugated to protein, usually for immunofluorescence
Ethidium bromide	Double-stranded nucleic acid, often with mithramycin
Propidium Iodide	Labeling double-stranded nucleic acid
Mithramycin	G/C rich DNA, often used with ethidium bromide
Chromomycin A_3	G/C rich DNA
Hoechst 33342 or 33258	A/T rich DNA
4',6-Diamidino-2-phenylindole	Nucleic acid dye
SYBR Green I	Nucleic acid
YOYO-I, YO-PRO-I	Cyanine dye-based nucleic acid
PicoGreen	Double-stranded nucleic acid

Table 2
Fluorescent Probes Used to Assess Bacterial Viability by FCM

Fluorochrome	Cell function measured
Rhodamine 123	Membrane potential
Dihexylocarbocyanine dyes	Membrane potential (several forms exist)
Fluorescein diacetate	Enzyme activity, membrane integrity
Carboxyfluorescein diacetate	Enzyme activity, membrane integrity
Bis-carboxyethyl-carboxyfluorescein acetoxymethyl ester	Enzyme activity, membrane integrity
Calcein acetoxymethyl ester	Enzyme activity, membrane integrity
Chemchrome B	Enzyme activity, membrane integrity
Cyanoditolyl tetrazolium chloride	Respiratory activity
Ethidium bromide	Double-stranded nucleic acid
Propidium iodide	Double-stranded nucleic acid
BacLight	Commercial kit, nucleic acid
Oxonol dyes	Accumulate in dead cell membranes
Calcofluor White	Nucleic acid
PoPro	Nucleic acid
SYTOX Green	Nucleic acid

Author/Year/Journal	Study Preformed	Sample Type	Dye/Method Used
Andreatta et al. 2000 Cytometry	Image analysis to detect subgroups	Heterotrophic Bacterioplankton	DAPI/ TransFloroSpheres
Bernard et al. 2000 Microb. Ecol.	Relationship cell size, productivity, and genetic diversity	Aquatic Mediterranean	SYTO/Cell Sorting
Bernard et al. 2001 Cytometry	Genetic diversity of viable and active bacteria	Seawater cultures	DVC/ CTC/Cell Sorting
Bouvier et al. 2001 Cytometry	Estimate Bacteria Biovolume	Bacterial strains of aquatic env.	DAPI/ Light Scattering
Button et al. 2001 AEM	Determination of DNA Content	Seawater/marine org.	DAPI/Triton X-100
Davey et al. 1999 Cytometry	Identification of Microorganisms	*B.subtilis, E.coli, M. luteus*	TinopalCBS-X/SYTO 17/Oxonal V/ TO-PRO-3
Dubelaar et. al. 1999 Cytometry	*In situ* analysis of marine and fresh water	Marine and fresh water	CytoBuoy: a wireless Flow cytometer
Edwards et al. 1999 Cytometry	Rapid sequential flow cytometry	Chinese hamster ovary cells	GFP/hydroethidine/Plug flow
Gasol et al. 1999 AEM	Significance of size and nucelic acid content	Planktonic bacteria	Live/Dead kit/ epiflourescence microscopy
Gregori et al. 2001 AEM	Resolution of viable mebrane-compromised bacteria	Fresh and marine water samples	Live/Dead SYBR Green I & II and PI
Huage et al. 1999 Cytometry	Bacterial genome fingerprinting	*S.aureus* and λ-phage	T0T0-1/ PFGE
Jacoberger et al. 1999 Cytometry	Cell-Cycle-regulated gene expression	Human prostate cancer cell line	FITC-anti-cyclin B1
Larson et al. 2000 Cytometry	Rapid DNA fingerprinting of Pathogens	*S. aureus, E.coli*	PFGE / PicoGreen dye
Robertson et al. 1998 AEM	Determination Biomass of small bacteria in low concentrations in mixed species	*C.oliotrophus, Marinobacter sp., P.diminuta*	DAPI/Triton X-100
Simon et al. 1995 AEM	FISH using rRNA probes to identify Small Phytoplankton	*Phylas Chlorophyta, Heteroconta, Dinophyta, and Cryptophyta*	
Schmid et al. 1999 Cytometry	Live cell growth and phenotype in population with low viability	Molt-4f t-cell leukemia	7-AAD/ Leu-3a FITC
Zubov et al. 1999 AEM	Determination of total protein content of bacterial cells	Seawater sample isolates	SYPRO

Table 6.3: Selected Studies Exhibiting Multiple Uses Of Flow Cytometry

The above table displays in utmost brevity the a few applications of flow cytometry. However, this table does not give a full understanding of technique and results.

6.3.4. Assessment of viability and survival

In the last few years, various studies for the rapid assessment of bacterial viability and survival have been undertaken. In addition, numerous applications of

flow cytometry to analyse drug susceptibility of eukaryotes and prokaryotes have been reported. Efficient dyes for the assessment of viability and survival are **Rhodamine 123, oxonol, and propidium iodide**. The signals from these fluorochromes detect changes in membrane potential or envelope integrity of the cells. The level of accumulation of Rhodamine 123 and oxonol inside the cell depends on its membrane potential or the difference in charge between both sides of the plasma membrane. A study will be performed in order to find a rapid method to assess bacterial survival and viability among the bacterial populations. This is based on the hypothesis that viral contamination or stress may disrupt or alter the outer membrane and cell membrane in bacteria, resulting in changes in the integrity of cell envelopes and membrane potential.

Rhodamine 123 is a cationic lipophilic dye (accumulated cytosolically by cells with an inside negative transmembrane electrochemical potential) that has been used extensively to study bacterial viability.

Oxonols [i.e. bis-oxonol DiBaC(4)3] are anionic lipophilic dyes which, unlike Rhodamine 123, are not extensively accumulated cytosolically by cells with an inside negative transmembrane potential, and which have also been used to assess bacterial viability. Therefore, their fluorescence response is opposite that of cationic Rhodamine 123. When membrane potential increases, the fluorescence response of Rhodamine increases, but the fluorescence response of oxonol decreases. Conversely, when membrane potential decreases, Rhodamine 123 fluorescence decreases but oxonol fluorescence increases.

Propidium iodide (a phenanthridinium dye) is a membrane-impermeant dye that stains by intercalating into nucleic acid molecules stains nucleic acid molecules. It binds RNA and DNA and exhibits no sequence specificity (one dye per 4-5 base pairs of DNA). On binding to nucleic acids, the fluorescence yield is enhanced ~ 25-fold, and the excitation maxima is shifted ~30-40 nm to the red and the emission maxima is shifted ~15 nm to the blue. In the DNA bound form, propidium iodide excites at 536 nm and fluoresces at 617 nm. Propidium Iodide is cell membrane impermeant dye and can be used for evaluating viable cells (will stain negative). It is also useful in double-label experiments to detect DNA and a fluorescein-labeled antigen simultaneously.

The overall set-up of the viability protocol based on Rhodamine 123, bis-oxonol DiBaC(4)3, and propidium iodide stainings is determined by the following steps that will each have to be optimised :

- In order to permeabilise the bacterial outer membrane, EGTA (at a final concentration of 1 mM) will be added to an overnight culture (or maybe more depending on the length of the cell cycle of the bacterial populations) previously diluted 1/1000 in NaCl 0.9%.
- Rhodamine 123 will be added to a final concentration of around 0.2 µg/ml (concentration to be optimised) from a stock solution in ethanol.
- Propidium iodide will be added to a final concentration of around 10 µg/ml (concentration to be optimised).

♦ Oxonol (around 4 µl of 250 µM solution of oxonol in ethanol) will be added to 1ml of bacterial suspension and incubated for a few minutes (between 2 and 5 minutes) at room temperature.

Remark : Oxonol and Rhodamine 123 staining at different culture times will have to be determined in order to check whether Rhodamine 123 and oxonol are taken up at different rates during the exponential and stationary phases. A progressive decline in Rhodamine uptake is usually observed during the growth phase. Bacteria showing high Rhodamine 123 fluorescence should exhibit little or no fluorescence with oxonol. Because the fluorescence intensity of Rhodamine 123-stained populations is highly dependent upon their physiological state, standardisation of the inoculum is an important factor that has to be chosen for flow cytometrical experiments.

6.3.5. Use of flow cytometry to check DNA content changes in microbial populations

Flow cytometry is uniquely suited for cell cycle analysis, since information about different cell cycle stages can be obtained without the need for synchronization of the cell culture. Also, flow cytometry offers the advantage that it can be applied to organisms whose genetics and physiology are largely unknown; as long as a DNA-specific stain can be made to enter the cells and reach the nucleic acid, flow cytometry can be performed on most species from which single cell suspensions can be obtained.

In cell cycle analyses of eukaryotic organisms, flow cytometry is commonly employed, and a variety of instrument configurations, light sources and stains are in use. In contrast, flow cytometric analysis of bacterial cells is not widely used. Previously, the reason for this was that the resolution and sensitivity of conventional flow cytometers was near or below the detection limit for bacterial analysis; the DNA content of a bacterial cell is about 1,000-fold less than that of a human cell, and the cell mass difference is also considerable.

With the development of a flow cytometer with a different configuration, flow cytometry of bacteria became feasible and several laboratories have used this instrument for bacterial cell cycle analysis. Unique information has been gathered, including measurements of lengths of different cell cycle periods in various *Escherichia coli* strains subjected to different growth conditions, the degree of coordination of initiation at multiple origins of DNA replication, the variation in the DNA content and cell size distributions of bacterial populations at different stages of growth in batch cultures as well as other cell cycle studies. Characterisation of DNA replication and cell division patterns in different mutant strains has been carried out, as well as characterisation of strains in which chromosome replication is controlled by integrated plasmid replicons. Other uses include studies of antibiotic effects as well as analysis of clinical material.

The quantitative measurement of the DNA content became possible by specific fluorescent dyes which bound stoichio-metrically to the double helix of nuclear DNA. The fluorescence inside the cell-nucleus can easily be detected using flow cytometry. Thus, the amount of DNA for cell populations can be studied regarding: a) the cell cycle; b) the ploidy level

In a growing population the DNA amount per cell is correlated to certain cell cycle phases of replication of chromosomes. Three different phases can be distinguished: G1, S, G2M (G1: normal DNA content = 2c, S: DNA synthesis, G2M: double DNA content = 4c and mitosis).

The peaks of the G1 and G2M compartments in a DNA histogram show Gaussian distributions (see below). Different mathematical models are available to integrate peak areas in order to determine cell counts (in percent) of the different cell cycle phases.

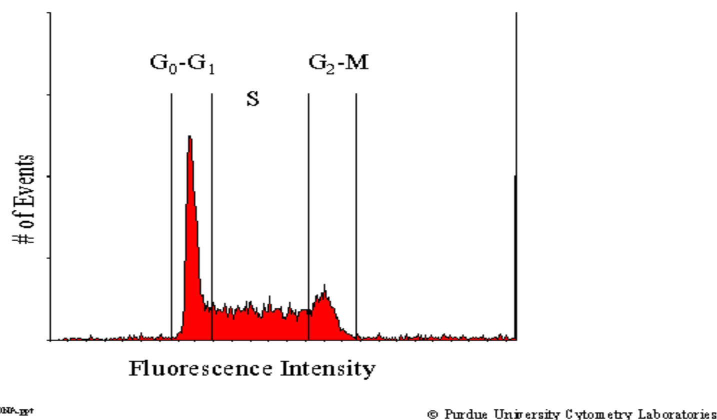


Figure 6.4 : DNA profile of a bacterial population.

6.3.6 Preparation of samples to analyse DNA content from bacterial cultures

Retention of fully replicated chromosomes . To obtain bacteria that contain only fully replicated chromosomes, different substances such as rifampicin (around 150 µg/ml) or cephalexin (around 10 µg/ml) will have to be added to the culture grown to an optical density of 0.1 to 0.3. The molecular mechanism of rifampicin is an inhibition of bacterial RNA polymerase that forms a complex with rifampicin. Rifampicin prevents the ribosomes from binding to messenger RNA, and thus proteins are not produced.

In figure 5.5, flow cytometry is useful in order to study the DNA distribution of an exponentially E. Coli growing culture. Note that the DNA distribution resembles that of the age. distribution, there are more cells with the low DNA content corresponding to a little more than 2 fully replicated chromosomes, than cells with twice the amount of DNA. Fig. 12B shows the DNA distribution of a sample from the same culture incubated for 3 hours in the presence of antibiotics inhibiting protein synthesis and cell division. The DNA now occurs in discrete peaks, most of the cells have 4 fully replicated chromosomes. The number of fully replicated chromosomes corresponds to the number of origins present in the cell at the time of drug addition. Cultures will have to be preincubated for around 3 hours (time to optimise) in the presence of rifampicin and cephalexin in order to allow complete runout of replication, such that all cells contain only fully replicated chromosomes.

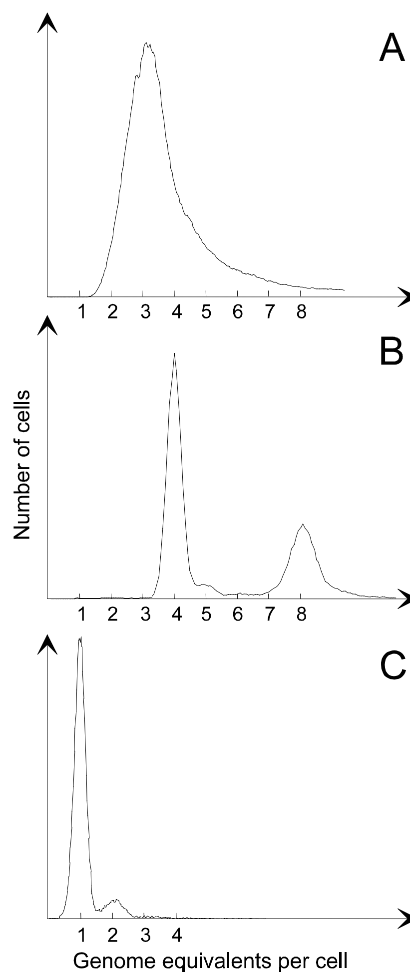


Figure 6.5 Flow cytometric DNA distributions. A. A sample from an exponentially growing culture of *E. coli* was fixed and stained. B. A sample from the same culture was incubated for 5 hrs with rifampicin and cephalexin and then fixed and stained. C. A sample from a 30°C culture of a dnaA46 mutant was incubated at 42°C for 2 hrs and then fixed and stained.

DNA staining:

Samples will be chilled in 1.5 ml of chilled Tris buffer (Tris 10 mM buffer + 1 mM EDTA) pelleted and washed in Tris buffer, pelleted again, and resuspended in the same buffer before fixation in 70%-80% cold ethanol during vortexing.Different DNA stains will have then to be compared and the optimal concentration range will have to be determined from dilution experiments:

- Propidium iodide (around 50 µg/ml)
- Mithramycin A (around 200 µg/ml)
- 7AAD (around 50 µg/ml)
- Ethidium bromide (around 40 µg/ml)

RNAse treatment:

Since DNA stains (propidium iodide, 7AAD) generally bind double strain RNA, the sample had normally to be treated with RNAse for DNA measurements to

be correct. Otherwise, the fluorescence from RNA bound dye molecules may broaden the DNA distributions. However, some authors do not include any RNAse treatment. A comparison between the presence or the absence of RNAse treatment will have to be performed.

Coefficient of variation :

The coefficient of variation (CV) of a peak area is given by the term standard deviation divided by mean channel number in percent. A low CV value indicates a high resolution analysis due to a very homogenous DNA staining and excellent instrument performance in detecting the fluorescent signal. With the EPICS XL instruments, very low CV values (lower than 1,5 %) can be obtained for various cell samples.

The coefficient of variation (CV; the standard deviation of a normal distribution divided by the average value) for the peaks obtained will be used for comparison of stains and concentrations of rifampicin and cephalexin.

7. Bioinformatic approaches

7.1. Outline and principles

Bioinformatics is an emerging discipline, which combines efforts and advances in computer science, computational science, and various aspects of life science. While there is currently no generally agreed upon definition of 'bioinformatics,' one view is that bioinformatics is concerned with the creation and development of advanced information and computational technologies to solve problems in biology. That is, bioinformatics typically relies on techniques from informatics, involving efficient algorithms for manipulating large amounts of data stored and processed on high-performance computers, in order to obtain information about genomic or protein sequence data. Bioinformatics is often applied to problems in *computational chemistry*, *functional genomics*, *neural imaging (brain mapping)*, *pharmacogenetics*, *pharmacogenomics*, *pharmacometrics*, *proteomics*, and *structural biology*.

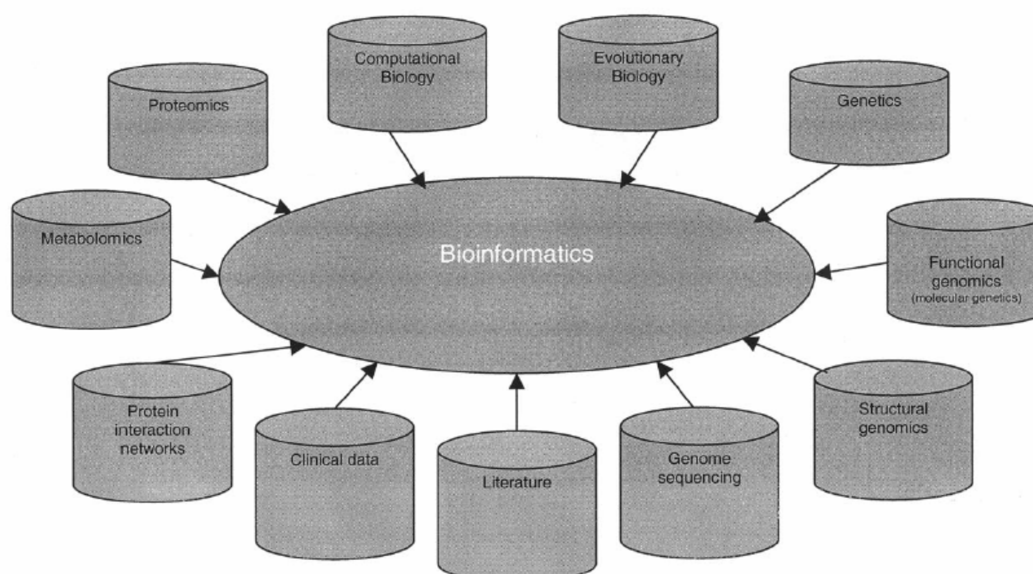


Fig. 7.1.: In the genome era, bioinformatics is strategically placed to exploit information directly from a multifold of science fields (from Attwood and Miller, 2001)

7.2. Bioinformatic tools

Predicting functions. Central to bioinformatic approaches is to predict putative gene functions by finding a similar sequence that has been experimentally studied in another organism. For example, a quarter of cloned human genes were found to have matches among *Sacharomyces* sequences with known functions.

For molecular biologists it is thus extremely important to have access to public databases with the most recent information (see further). In addition, there are numerous specialized databases covering a diverse array of areas. Taken together the databases provide valuable collections of organized data that is of broad benefit to the

molecular biologists. However, the number and diversity of information resources makes efficient data-resource discovery as important as effective data-resource use. Thus, **tools and systems** to assist the researcher in navigating the biological data are increasingly important. BLAST and FASTA are programs for finding sequence similarities between a selected piece of DNA or amino acid sequence and the DNA and protein sequences contained in the databases. These programs report the hit in the database along with the estimated statistical signficance of the hit. According to probability theory, the similarity score required for a given level of statistical significance is proportional to the logarithm of the database size. Therefore, as the database grows, biologically significant matches of distantly related sequences may have smaller similarity scores than random matches, and may be lost in the noise. One approach to this problem is to simplify the database by eliminating redundant sequences or by reducing families of similar sequences to a single representative sequence in the database that can be used for the initial searching.

The result of the sequence similarity search provides initial information about the putative function of the gene product of interest.

7.3. Databases

Because of the high rate of data production and the need for researchers to have rapid access to new data, public databases have become the major medium through which genome sequence data are published. Public databases and the data services that support them are important resources in bioinformatics.

EMBL (Heidelberg, Germany) and GenBank (Bethesda, USA) are the two major nucleotide databases. Both sources collaborate and synchronize their databases so that the databases contain the same information. The rate of growth of DNA databases has been following an exponential trend, with a doubling time now estimated to be 9-12 months. The two databases are updated on a daily basis.

General Databases

GenBank - DNA and RNA sequences, National Center for Biotechnology Information, USA
EMBL - DNA and RNA sequences; the European Molecular Biology Laboratory, Cambridge
PIR - Protein sequences, Protein Identification Resource, USA
SWISS-PROT - Protein sequences, Switzerland
NRL-3D - Predicted protein structures
OWL - Non-redundant collections of protein sequences
PROSITE - Protein sequence motifs
PRINTS - Protein sequence motifs
BLOCKS - Protein sequence motifs
SCOP - Proteins classified according to strutural similarities
PDB - Macromolecular structures

The principal requirements on the public data services are:

- *Data quality* - data quality has to be of the highest priority. However, because the data services in most cases lack access to supporting data, the quality of the data must remain the primary responsibility of the submitter.
- *Supporting data* - database users will need to examine the primary experimental data, either in the database itself, or by following cross-references back to network-accessible laboratory databases.
- *Deep annotation* - deep, consistent annotation comprising supporting and ancillary information should be attached to each basic datat object in the database.
- *Timeliness* - the basic data should be available on an Internet-accessible server within days (or hours) of publication or submission.
- *Integration* - each data object in the database should be cross-referenced to representation of the same or related biological entities in other databases. Data services should provide capabilities for following these links from one database or data service to another.

Access to bioinformation

Biological databases are built by different teams, in different locations, for different purposes, and using different data models and supporting database-management systems. However, biological databases are most valuable when interconnected than when isolated. One approach to database integration is construction of either *a data warehouse* or a database containing a combination of datasets from a variety of primary databases. Annotations and connections may be added, either automatically (algorithmically) or manually by experts (curators). Some examples of integrated bioinformation resources are:

- **SRS** (Sequence Retrieval System)
- **Entrez** Browser
- **ExPASy**
- **Integrated genome database**

The popularity of these services indicates the need for querying interrelated datasets, rather than isolated databases. The advantages of physical integration are that queries can be executed rapidly because all data are located in one place, and the user sees a homogenous, integrated data source.

7.4. Relevant analysis strategies

For *Nitrosomonas europea* the strategy would involve the search for relevant and critical genes via BLAST before designing a stressome DNA array to investigate the induction of specific genes. For *Rhodospirillum rubrum* the same strategy will be possible when the sequence will be completely annotated and published on the JGI (Doe Joint Genome Institute) DNA sequence site. However, because the published strain (ATCC 11170) is not the strain used in the MELiSSA loop (ATCC 25903), it will still be necessary to sequence (parts of the) genes to be able to know the exact sequence necessary to build the stressome chip.

To build the stressome chips for *Arthrospira platensis*, it will be necessary to design degenerate primers for extracting the interesting genes from the genome (see 2.3.) based on BLAST analysis of the closely related *Synechocystis* PCC6803.

8. References

8.1. Reference list for proteomic analysis

Albrecht, J; Schmid, E W; Sssmuth, R. (1986) Some remarks about laser-induced mass spectrometry of bacteria Zeitschrift Fur Naturforschung 41:337-342

Anderson NL, and Anderson NG. (1998) Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis 19:1853-1861.

Berkenkamp S, Menzel C, Karas M, and Hillenkamp F. (1997) Performance of infrared matrix-assisted laser desorption/ionization mass spectrometry with lasers emitting in the 3mm wavelength range. Rapid Commun Mass Spectrom 11:1399-1406.

Berndt P, Hobohm U, and Langen H. (1999) Reliable automatic protein identifcation from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. Electrophoresis 20:3521-3526.

Blackstock W and M. Mann. Proteomics . A trends guide. Elsevier Science London

Blackstock WP, and Weir MP. (1999) Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol 17:121-127.

Boyle MD, Romer TG, Meeker AK, Sledjeski DD. (2001) Use of surface-enhanced laser desorption ionization protein chip system to analyze streptococcal exotoxin B activity secreted by Streptococcus pyogenes. J Microbiol Methods. 46:87-97.

Bright JJ, Claydon MA, Soufian M, Gordon DB. (2002) Rapid typing of bacteria using matrix-assisted laser desorption ionisation time-of-flight mass spectrometry and pattern recognition software. J Microbiol Methods. 48:127-138.

Bundy J, Fenselau C. (1999) Lectin-based affinity capture for MALDI-MS analysis of bacteria. Anal Chem. 71:1460-3.

Bumann D, Aksu S, Wendland M, Janek K, Zimny-Arndt U, Sabarth N, Meyer TF, and Jungblut PR. (2002) Proteome analysis of secreted proteins of the gastric pathogen Helicobacter pylori. Infect Immun. 70:3396-3403.

Evason DJ, Claydon MA, Gordon DB. (2001) Exploring the limits of bacterial identification by intact cell-mass spectrometry. J Am Soc Mass Spectrom. 12:49-54.

Fenselau C, Demirev PA. (2001) Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev. 20:157-71.

Hughes MJ, Moore JC, Lane JD, Wilson R, Pribul PK, Younes ZN, Dobson RJ, Everest P, Reason AJ, Redfern JM, Greer FM, Paxton T, Panico M, Morris HR, Feldman RG, and Santangelo JD. (2002) Identification of major outer surface proteins of Streptococcus agalactiae. Infect Immun. 70:1254-1259.

Kawano Y, Ito Y, Yamakawa Y, Yamashino T, Horii T, Hasegawa T, Ohta M. (2000) Rapid isolation and identification of staphylococcal exoproteins by reverse phase capillary high performance liquid chromatography-electrospray ionization mass spectrometry. FEMS Microbiol Lett. 189:103-8.

Kent J. Voorhees, Franco Basile, Michael B. Beverly, Christy Abbas-Hawks, Alan Hendricker, R. B. Cody and Ted L. Hadfield. (1997) The use of biomarker compounds for the identification of bacteria by pyrolysis--mass spectrometry. Journal of Analytical and Applied Pyrolysis. 40-41:111-134

Kirkpatrick C, Maurer LM, Oyelakin NE, Yoncheva YN, Maurer R, and Slonczewski JL.(2001) Acetate and formate stress: opposite responses in the proteome of *Escherichia coli*. J. Bacteriol. 183:6466-6477.

Lay JO Jr. (2001) MALDI-TOF mass spectrometry of bacteria. Mass Spectrom Rev. 20:172-94.

Lucarelli C, Radin L, Corio R, Eftimiadi C. (1990) Applications of high-performance liquid chromatography in bacteriology. I. Determination of metabolites. J Chromatogr. 515:415-34.

Masse R, Lalanne D, Messier F, Sylvestre M. (1989) Characterization of new bacterial transformation products of 1,1,1-trichloro-2,2-bis-(4-chlorophenyl) ethane (DDT) by gas chromatography/mass spectrometry. Biomed Environ Mass Spectrom. 18:741-52.

Phillips, I. (1990) New methods for identification of obligate anaerobes, Reviews of Infectious Diseases, Volume 12 Supplement 2, January - February 1990, Pages S127-S132

Rabilloud T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics. 2:3-10. Review.

Rosen R, Buttner K, Becher D, Nakahigashi K, Yura T, Hecker M, and Ron EZ. (2002) Heat shock proteome of *Agrobacterium tumefaciens*: evidence for new control systems. J. Bacteriol. 184:1772-1778.

Sabarth N, Lamer S., Zimny-Arndt U., Jungblut P., Meyer T., Bumann D. (2002) Identification of surface proteins of Helicobacter pylori by selective biotinylation, affinity purification and two-dimensional gel electrophoresis. Journal of Biochemistry (In Press).

Sinha P, Poland J, Schnolzer M, and Rabilloud T. (2001) A new silver staining apparatus and procedure for matrix-assisted laser desorption/ionization-time of flight analysis of proteins after two-dimensional electrophoresis. Proteomics. 1:835-840.

Smole SC, King LA, Leopold PE, Arbeit RD. (2002) Sample preparation of Gram-positive bacteria for identification by matrix assisted laser desorption/ionization time-of-flight.
J Microbiol Methods. 48:107-15.

Stein T, Entian KD. (2002) Maturation of the lantibiotic subtilin: matrix-assisted laser desorption/ionization time-of-flight mass spectrometry to monitor precursors and their proteolytic processing in crude bacterial cultures. Rapid Commun Mass Spectrom. 16:103-10.

Szponar B, Larsson L. (2001) Use of mass spectrometry for characterising microbial communities in bioaerosols. Ann Agric Environ Med. 8:111-7.

Vaidyanathan S, Kell DB, Goodacre R. (2002) Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. J Am Soc Mass Spectrom. 13:118-28.

Walker J, Fox AJ, Edwards-Jones V, Gordon DB. (2002) Intact cell mass spectrometry (ICMS) used to type methicillin-resistant Staphylococcus aureus: media effects and inter-laboratory reproducibility. J Microbiol Methods. 48:117-26.

Wattiez R, Hermans C, Bernard A, Lesur O, and Falmagne P. (1999) Human bronchoalveolar lavage fluid: two-dimensional gel electrophoresis, amino acid microsequencing and identification of major proteins. Electrophoresis. 20:1634-1645.

White DC, Lytle CA, Gan YD, Piceno YM, Wimpee MH, Peacock AD, Smith CA. Flash detection/identification of pathogens, bacterial spores and bioterrorism agent biomarkers from clinical and environmental matrices. J Microbiol Methods. 48:139-47.

Wittmann C, Heinzle E. (2001) Application of MALDI-TOF MS to lysine-producing Corynebacterium glutamicum: a novel approach for metabolic flux analysis. Eur J Biochem. 268:2441-55.

Wittmann C, Heinzle E. (2001) Modeling and experimental design for metabolic flux analysis of lysine-producing Corynebacteria by mass spectrometry. Metab Eng. 3:173-91.

8.2. Reference list for bacterial community analysis

Alexander R.J., and Raicht R.F. (1998) Purification of total RNA from human stool samples. Dig Dis Sci 43:2652-2658.

Amann R.I. (1995) In situ identification of micro-organisms by whole cell hybridization with rRNA-targeted nucleic acid probes Mol. Microbial Ecol. 3.3.6:1-15.

Amann R.I., Binder B.J., Olson R.J., Chisholm S.W., Devereux R., and Stahl D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analysing mixed microbial populations. Appl. Environ. Microbiol. 56: 1919-1925.

Amann R.I., Stromly J., Devereux R., Key R., and Stahl D.A. (1992) Molecular and microscopic identification of sulfate-reducing bacteria in multispecies biofilms. Appl. Environ. Microbiol. 58: 614-623.

Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiol. Rev. 24:1-7.

Attwood, T. and C. Miller. (2001). Which craft is best in bioinformatics? Computers and Chemistry 25:329–339.

Cooney RP, Pantos O, Tissier MD, Barer MR, ODonnell AG, and Bythell JC. (2002) Characterization of the bacterial consortium associated with black band disease in coral using molecular microbiological techniques. Environ Microbiol. 4:401-413.

Davis SA, and Gordon DM. (2002) The influence of host dynamics on the clonal composition of *Escherichia coli* populations. Environ Microbiol. 4:306-313.

Di Mattia E, Grego S, and Cacciari I. (2002) Eco-physiological characterization of soil bacterial populations in different states of growth. Microb Ecol. 2002 43:34-43.

Dollhopf SL, Hashsham SA, Tiedje JM. (2001) Interpreting 16S rDNA T-RFLP Data: Application of Self-Organizing Maps and Principal Component Analysis to Describe Community Dynamics and Convergence. Microb Ecol. 42:495-505.

Dominik K, and Hofle MG. (2002) Changes in bacterioplankton community structure and activity with depth in a eutrophic lake as revealed by 5S rRNA analysis. Appl Environ Microbiol. 68:3606-3613.

Inagaki F, Sakihama Y, Inoue A, Kato C, and Horikoshi K. (2002) Molecular phylogenetic analyses of reverse-transcribed bacterial rRNA obtained from deep-sea cold seep sediments. Environ Microbiol. 4:277-86.

Juretschko S, Loy A, Lehner A, and Wagner M. (2002) The microbial community composition of a nitrifying-denitrifying activated sludge from an industrial sewage treatment plant analyzed by the full-cycle rRNA approach. Syst Appl Microbiol. 25:84-99.

Hill J.A., Seipp R.P., Betts M., Hawkins L., Van Kessel A.G., Crosby W.L., and S.M. Hemmingsen. (2002) Extensive profiling of a complex microbial community by high-throughput sequencing. Appl. Environ. Microbiol. 68: 3055-3066.

Hopkins MJ, and Macfarlane GT. (2002) Changes in predominant bacterial populations in human faeces with age and with *Clostridium difficile* infection. J Med Microbiol. 51:448-454.

Laverman AM, Speksnijder AG, Braster M, Kowalchuk GA, Verhoef HA, and Van Verseveld HW. (2001) Spatiotemporal Stability of an Ammonia-Oxidizing Community in a Nitrogen-Saturated Forest Soil. Microb Ecol. 42:35-45.

Kaiser O, Puhler A, and Selbitschka W. (2001) Phylogenetic Analysis of Microbial Diversity in the Rhizoplane of Oilseed Rape (*Brassica napus* cv. Westar) Employing Cultivation-Dependent and Cultivation-Independent Approaches. Microb Ecol. 42:136-149.

Koizumi Y., Kelly J.J., Nakagawa T., Urakawa H., El-Fantroussi S., Al-Muzaini S., Fukui M., Urushigawa Y., and Stahl D.A. (2002) Parallel Characterization of anaerobic toluene- and ethylbenzene- degrading microbial consortia by PCR-denaturing gradient gel electrophoresis, RNA-DNA membrane hybridisation, and DNA microarray technology. Appl. Environ. Microbiol. 68: 3215-3225.

Muller AK, Westergaard K, Christensen S, and Sorensen SJ. (2002) The Diversity and Function of Soil Microbial Communities Exposed to Different Disturbances. Microb Ecol. 2002 Apr 30 [epub ahead of print]

Muyzer G., de Waal E.C., and Uitterlinden A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59: 695-700.

Ranjard L., Poly F., and Nazaret S. (2000) Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. Res. Microbiol. 151: 167-177.

Rimini R., Jansson B., Feger G., Roberts T.C., de Francesco M., Gozzi A., Faggioni F., Domenici E., Wallace D. M., Frandsen N., and Polissi A. (2000) Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. Mol. Microbiol. 36: 1279-1292.

Schabereiter-Gurtner C, Maca S, Kaminsky S, Rolleke S, Lubitz W, and Barisani-Asenbauer T. (2002) Investigation of an anaerobic microbial community associated with a corneal ulcer by denaturing gradient gel electrophoresis and 16S rDNA sequence analysis. Diagn Microbiol Infect Dis. 43:193-199.

Schabereiter-Gurtner C, Saiz-Jimenez C, Pinar G, Lubitz W, and Rolleke S. (2002) Phylogenetic 16S rRNA analysis reveals the presence of complex and partly unknown bacterial communities in Tito Bustillo cave, Spain, and on its Palaeolithic paintings. Environ Microbiol. 4:392-400.

Sessitsch A, Gyamfi S, Stralis-Pavese N, Weilharter A, and Pfeifer U. (2002) RNA isolation from soil for bacterial community and functional analysis: evaluation of different extraction and soil conservation protocols. J Microbiol Methods. 51:171-179.

Simpson JM, Martineau B, Jones WE, Ballam JM, and Mackie RI. (2002) Characterization of Fecal Bacterial Populations in Canines: Effects of Age, Breed and Dietary Fiber. Microb Ecol. (in press)

Song B, Kerkhof L, and Haggblom M. (2002) Characterization of bacterial consortia capable of degrading 4-chlorobenzoate and 4-bromobenzoate under denitrifying conditions. FEMS Microbiol Lett. 213:183.

Stach JE, and Burns RG. (2002) Enrichment versus biofilm culture: a functional and phylogenetic comparison of polycyclic aromatic hydrocarbon-degrading microbial communities. Environ Microbiol. 4:169-82.

Valinsky L., Vedova G.D., Scupham A.J., Alvey S., Figueroa A., Yin B., Hartin R.J., Chrobak M., Crowley D.E., Jiang T., and Borneman J. (2002) Analysis of bacterial community composition by oligonucleotide fingerprinting of rRNA genes. Appl. Environ. Microbiol. 68: 3243-3250.

Van Wintzingerode F., Göbel U.B., and E. Stackebrandt (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev. 21:213-229.

Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, and de Vos WM. (2002) Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol. 68:3401-3407.

8.3. Reference list for DNA chip technology

Alland D, Kramnik I, Weisbrod TR, Otsubo L, Cerny R, Miller LP, Jacobs WR Jr, Bloom BR. (1998) Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 95:13227-13232.

Collado-Vides J. (1991) A syntactic representation of units of genetic information--a syntax of units of genetic information. J Theor Biol. 148:401-429.

DeRisi JL, Iyer V, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680-686.

Dybvig K. (1993) DNA rearrangements and phenotypic switching in prokaryotes. Mol Microbiol. 10:465-471.

Gill RT, Katsoulakis E, Schmitt W, Taroncher-Oldenburg G, Misra J, Stephanopoulos G. (2002) Genome-wide dynamic transcriptional profiling of the light-to-dark transition in *Synechocystis* sp. strain PCC 6803. J Bacteriol. 184:3671-3681.

Nakabachi A, Ishikawa H. (2001) Expression of host S-adenosylmethionine decarboxylase gene and polyamine composition in aphid bacteriocytes.
Insect Biochem Mol Biol. 31:491-496.

Perini LT, Doherty EA, Werner E, Senear DF. (1996) Multiple specific CytR binding sites at the Escherichia coli deoP2 promoter mediate both cooperative and competitive

interactions between CytR and cAMP receptor protein. J Biol Chem. 271:33242-33255.

Rosenow C, Saxena RM, Durst M, Gingeras TR. (2001) Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. Nucleic Acids Res. 29:E112

Sabatti C, Rohlin L, Oh MK, Liao JC. (2002) Co-expression pattern from DNA microarray experiments as a tool for operon prediction. Nucleic Acids Res. 30:2886-2893.

Serrano M, Salas M, Hermoso JM. (1993) Multimeric complexes formed by DNA-binding proteins of low sequence specificity. Trends Biochem Sci. 18:202-206.

Stephens C. (2001) Bacterial cell cycle: seeing the big picture with microarrays. Curr Biol. 11:R222-225.

Wong KK, McClelland M. (1994) Stress-inducible gene of *Salmonella typhimurium* identified by arbitrarily primed PCR of RNA. Proc Natl Acad Sci U S A. 91:639-643.

Ye RW, Wang T, Bedzyk L, Croker KM. (2001) Applications of DNA microarrays in microbial systems. J Microbiol Methods. 47:257-272.

8.4. Reference list for genomic analysis

Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiol. Rev. 24:1-7.

Ashelford KE, Weightman AJ, Fry JC. (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. Nucleic Acids Res. 30:3481-3489.

Casjens S. (1998). The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. 32:339-377.

Chen F, Suttle CA. (1995) Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Appl Environ Microbiol. 61:1274-1278.

Geiss U, Vinnemeier J, Kunert A, Lindner I, Gemmer B, Lorenz M, Hagemann M, Schoor A. (2001) Detection of the isiA gene across cyanobacterial strains: potential for probing iron deficiency. Appl Environ Microbiol. 67:5247-5253.

Janssen, P., R. Coopman, G. Huys, J. Swings, Marjo Bleeker, P. Vos, M. Zabeau, and K. Kersters. (1996) Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. Microbiology 142:1881-1893.

Janssen, P. (2001). Genotypic analysis of microbial genomes by selective restriction fragment analysis. *In* New Approaches for the Generation and Analysis of Microbial Fingerprints. Ed. *L. Dijkshoorn, M. Struelens, & K. Towner*. Elsevier, Amsterdam.

Janssen, P, Enright A, Audit B, Cases I, Goldovsky L, Harte N, Kunin, V, and Ouzounis, C (2002). COmplete GENome Tracking (COGENT): an open-source environment for computational genomics. Bioinformatics (In Press).

Knablein J, Dobbek H, Ehlert S, Schneider F. (1997) Isolation, cloning, sequence analysis and X-ray structure of dimethyl sulfoxide/trimethylamine N-oxide reductase from Rhodobacter capsulatus. Biol Chem. 378:293-302.

Knablein J, Mann K, Ehlert S, Fonstein M, Huber R, Schneider F. (1996) Isolation, cloning, sequence analysis and localization of the operon encoding dimethyl sulfoxide/trimethylamine N-oxide reductase from *Rhodobacter capsulatus*. J Mol Biol. 263:40-52.

Kulaeva OI, Koonin EV, McDonald JP, Randall SK, Rabinovich N, Connaughton JF, Levine AS, Woodgate R. (1996) Identification of a DinB/UmuC homolog in the archeon *Sulfolobus solfataricus*. Mutat Res. 357:245-53.

Rimini R., Jansson B., Feger G., Roberts T.C., de Francesco M., Gozzi A., Faggioni F., Domenici E., Wallace D. M., Frandsen N., and Polissi A. (2000) Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. Mol. Microbiol. 36: 1279-1292

Roberts GA, Grogan G, Greter A, Flitsch SL, Turner NJ. (2002) Identification of a new class of cytochrome P450 from a *Rhodococcus* sp. J Bacteriol. 184:3898-908.

Schloter M, Lebuhn M, Heulin T, Hartmann A. (2000). Ecology and evolution of bacterial microdiversity. FEMS Microbiol Rev. 24:647-60

Tholl D, Ober D, Martin W, Kellermann J, Hartmann T. (1996) Purification, molecular cloning and expression in *Escherichia coli* of homospermidine synthase from *Rhodopseudomonas viridis*. Eur J Biochem. 240:373-9.

van Eldere J, Janssen P, Hoefnagels-Schuermans A, van Lierde S, Peetermans WE. (1999). Amplified-fragment length polymorphism analysis versus macro-restriction fragment analysis for molecular typing of Streptococcus pneumoniae isolates. J Clin Microbiol. 37:2053-7.

Vaneechoutte, M. (1996). DNA fingerprinting technques for microorganisms. A proposal for classification and nomenclature. Molecular Biotechnology 6:115-142.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, and Zabeau.M. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-14.

Vonshak A., (1997) *Spirulina platensis* (*Arthrospira*) : Physiology, Cell Biology and Biotechnology, Taylor & Francis.

8.5. Reference list for flow cytometry

Albrecht, J; Schmid, E W; Sssmuth, R. 1986. Some remarks about laser-induced mass spectrometry of bacteria Zeitschrift Fur Naturforschung 41:337-342

Bentley-Mowat, J.A. (1982) Application of fluorescence microscopy to pollution studies on marine phytoplankton. Bot Mar 25: 203-204.

Button, D.K. and Robertson, B.R. (1989) Kinetics of bacterial processes in natural aquatic systems based on biomass as determined by high-resolution flow cytometry. Cytometry 10:558-563.

Campbell, L. and Vaulot, D. (1993) Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). Deep-Sea Res. 40:2043-2060.

Campbell, L., Nolla, H.A., and Vaulot, D. (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. Lintnol. Oceanogr. 39:954-961.

Cavender-Bares, K.K., Frankel, S.L., and Chisholm, S.W. (1998) A dual sheath flow cytometer for shipboard analyses of phytoplankton communities from the oligotrophic oceans. Limnol. Oceanogr: 43:1383-1388.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Waterbury, J., Goericke, R., and Welschmeyer, N. (1988) A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone. Nature 334:340-343.

Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., West- Johnsrud, L., and Zettler, E.R. (1992) *Ptochlorococcus marinus* nov. gen. nov. sp.: An oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. Arch. Microbiol. 157:297- 300.

Courties, C., Vaquer, A., Trousselier, M., Lautier, J., Chrétiennot-Dinet, M-J., Neveux, J., Machado, C., and Claustre, H. (1994) Smallest eukaryotic organism. Nature 3701255.

Dorsey, J., Yentsch,,C.M., Mayo, S., McKenna, C. (1989) Rapid analytical technique for the assessment of cell metabolic activity in marine microalgae. Cytometry 10: 622-628.

Dubelaar, G.B.J., Groenewegen, A.C., Stokdijk, W., Van Den Engh, G.J., and Visser, J.W.M. (1989) Optical plankton analyser: A flow cytometer for plankton analysis, II: Specifications. Cytometry 10:529-539.

Dusenberry, J.A. and Frankel, S.L. (1994) Increasing the sensitivity of a FACScan flow cytometer to study oceanic picoplankton. Limnol. Oceanogr 39:206-210.

Frank J. Jochem F. J., Dark survival strategies in marine phytoplankton assessed by cytometric measurement of metabolic activity with fluorescein diacetate. Marine Biology, in press. http://www.jochemnet.de/fda3.html

Hobbie, J.E., Daley, R.J., and Jasper, S. (1977) Use of Nuclepore filters for counting bacteria by fluorescence microscopy. Appl. Environ. Microbiol. 33:1225-1228.

Kriss, A.E. and Rukina, E.A. (1947) Bacteriophages in the sea. Dokl. Akad. Nauk SSSR 57:833-836.

Lebaron, P., Catala, P, and Parthuisot, N. 1998. Effectiveness of SYTOX Green stain for bacterial viability assessment. Appl. Environ. Microbiol. 98:2697-2700.

Li, W.K.W. and Wood, A.M. (1988) Vertical distribution of North Atlantic ultraphytoplankton: Analysis by flow cytometry and epifluorescence microscopy. Deep Sea Res. 35:1615-1638.

Li, W.K.W., Jellett, J.F., and Dickie, PM. (1995) DNA distribution in planktonic bacteria stained with TOTO or TO-PRO. Limnol. Oceanogr. 40: 1485-1495.

Marie, D., Vaulot, D., and Partensky, F. (1996) Application of the novel nucleic acid dyes YOYO- 1, YO-PRO-1 and PicoGreen analysis of marine prokaryotes. Appl. Environ,. Microbiol. 62: 1649-1655.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green-I. Appl. Environ. Microbiol. 93:186-193.

Monger, B.C. and Landry, M.R. (1993) Flow cytometric analysis of marine bacteria with Hoechst 33342. Appl. Environ. Microbiol. 59:905-911.

Olson, R.J., Chisholm, S.W., Zettler, E.R., and Armbrust, E.V. (1988) Analysis of *Synechococcus* pigment types in the sea using single and dual beam flow cytometry. Deep Sea Res. 35:425- 440.

Partensky. F., Blanchot, J., Lantoine, F., Neveux, J., and Marie, D. (1996) Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. Deep Sea Res 43:1191-1213.

Partensky, F., Hess, W.R., and Vaulot, D. (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Mictvbiol. Mol. Biol. Rev 63:106-127.

Porter, J., Daniel, D., Pickup, R., Edwards, C. (1996) Fluorescent Probes and Flow Cytometry:New Insights Into Environmental Bacteriology. Cytometry. 23:91-98.

Porter, J., Pickup, R. and Edwards, C. (1997) Evaluation of Flow Cytometric Methods for the Detection and Viability Assessment of Bacteria from Soil. Soil Biol. Biochem. 29:91-100.

Porter, J., Deere, D., Hardman, M., Edwards, C., Pickup, R., (1997) Go with the flow-use of flow cytometry in environmental microbiology. FEMS Microbiol. Ecol. 24, 93-101.

Proctor, L.M. and Fuhrman, J.A. (1990) Viral mortality of marine bacteria and cyanobacteria. Nature 343:60-62.

Robertson, B.R., Button, D.K., and Kloch, A.L. (1998) Determination of the biomasses of small bacteria at low concentrations in a mixture of species with forward light scatter measurements by flow cytometry. Appl. Environ. Microbiol. 64:3900-3909.

Shapiro, H.M. (2000) Microbial analysis at single cell level: tasks and techniques. J. of Microbiol Meth 42:3-16.

Shapiro, L.P. and Haugen, E.M. (1988) Seasonal distribution and temperature tolerance of Synechococcus in Boothbay Harbor, Maine. Estuarine
Coastal Shelf Sci. 26:517-525.

Spencer, R. (1955) A marine bacteriophage. Nature 175:690. 11.11.14 Supplement 10 Current Protocols in Cytometry

Spencer, R. (1960) Indigenous marine bacteriophages. J. Bacteriol. 79:614.

Van Etten, J.L., Burbank, D.E., Schuster, A.M., Meints, R.H., (1985) Lytic viruses infecting a *Chlorella* -like alga. Virology 140, 135-143.

Vaulot, D. (1989) CYTOPC: Processing software for flow cytometric data. Signal Noise 2:8.

Vaulot, D., Courties, C., and Partensky, F (1989) A simple method to preserve oceanic phytoplankton for flow cytometric analyses. Cytometry 10:629-635.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979) Widespread occurrence of a unicellular, marine planktonic, cyanobacterium. Nature 277:293-294.

Wilson, W.H., Joint, I.R., Carr, N.G., Mann, N.H., (1993) Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. Strain WH7803. Appl. Environ. Microbiol. 59, 3736-3743.

Internet Resources

http://www.sb-roscoff.fr/Phyto/cyto.html
http://carl.im.uib.no/sup
http://www.flowcytometry.org
http://CCMP.bigelow.org
http://www.isac-net.org/techological/topoics/flowCytometry.htm
http://www4.umdnj.edu/cytoweb/intro.htm)