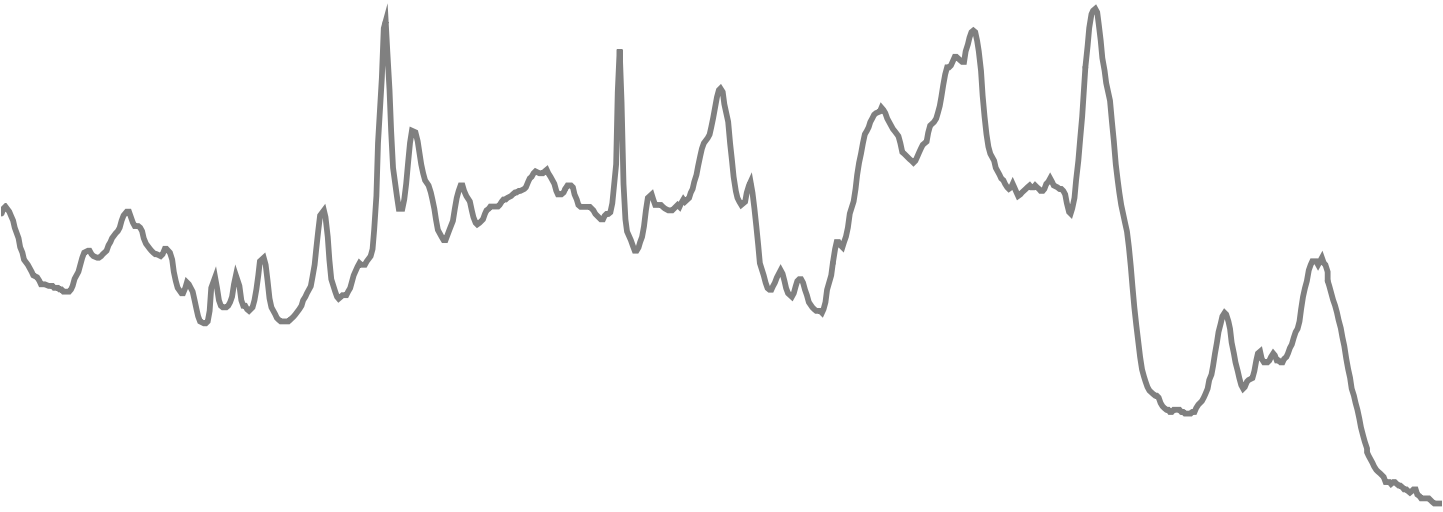
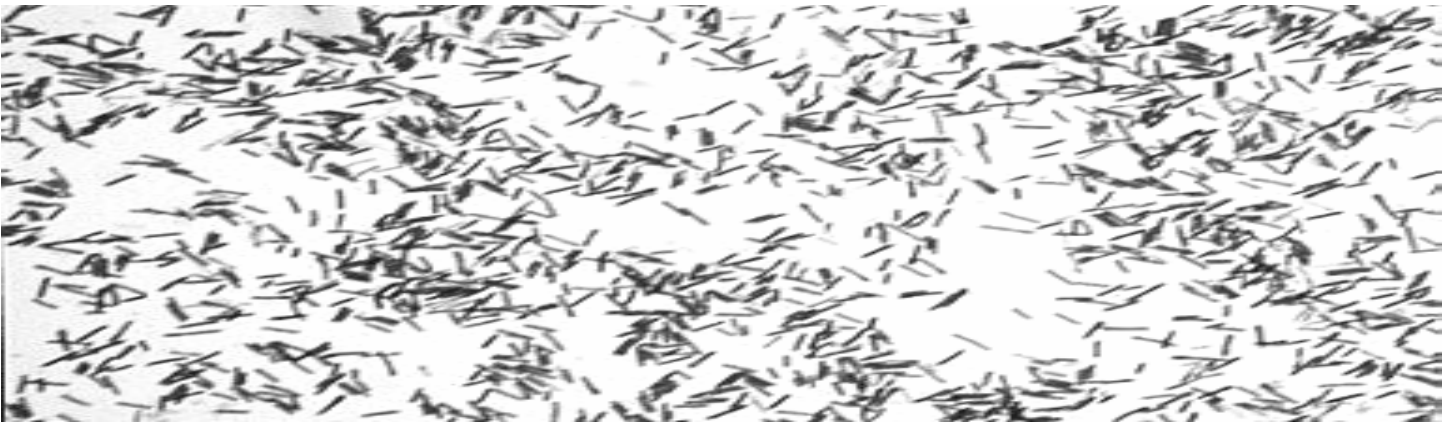


Evaluation of Raman spectroscopy as identification tool for bacteria



UNIVERSITEIT GENT

Proefschrift neergelegd tot het verkrijgen van de graad van Doctor in de wetenschappen, Scheikunde

Didier Hutsebaut

Evaluation of Raman spectroscopy as identification tool for bacteria

Evaluatie van Raman spectroscopie als een identificatiemethode voor bacteriën

Didier Hutsebaut

Faculteit Wetenschappen
Vakgroep Analytische chemie

&

Vakgroep Biochemie, Fysiologie en Microbiologie,
Laboratorium voor Microbiologie

Academiejaar 2004 - 2005

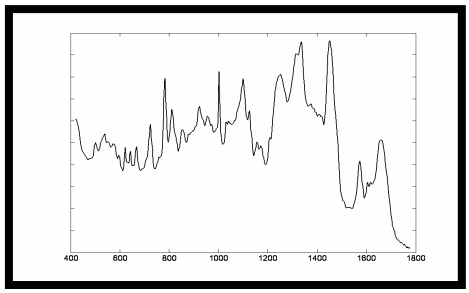
Proefschrift ingediend tot het verkrijgen van de graad van
Doctor in de wetenschappen, Scheikunde

Promotor:
Prof. Dr. Luc Moens

Co-promotoren:
Prof. Dr. Paul De Vos
&
Dr. Peter Vandenabeele

B. cereus viewed by a layman (actual size)

B. cereus viewed by a microbiologist (600 X)



B. cereus viewed by a spectroscopist

Content

Introduction 5

Chapter I
On the chemistry between the genus *Bacillus* and Raman spectroscopy (an introductory chapter) 9

Chapter II
Evaluation of an accurate calibration and spectral standardization procedure for Raman Spectroscopy 77

Chapter III
Growth conditions disclosed by three *Bacillus* species and how they affect the achievable taxonomic resolution of Raman spectroscopy 106

Chapter IV
Raman micro-spectroscopy as an identification tool within the phylogenetically homogeneous '*Bacillus subtilis*'-group 127

Chapter V
Exploration of micro-Raman spectroscopy as a taxonomic tool within the '*Bacillus cereus*' group: A comparison with FAFLP 152

Chapter VI
Discussion 177

Chapter VII
Summary & Future prospects 182

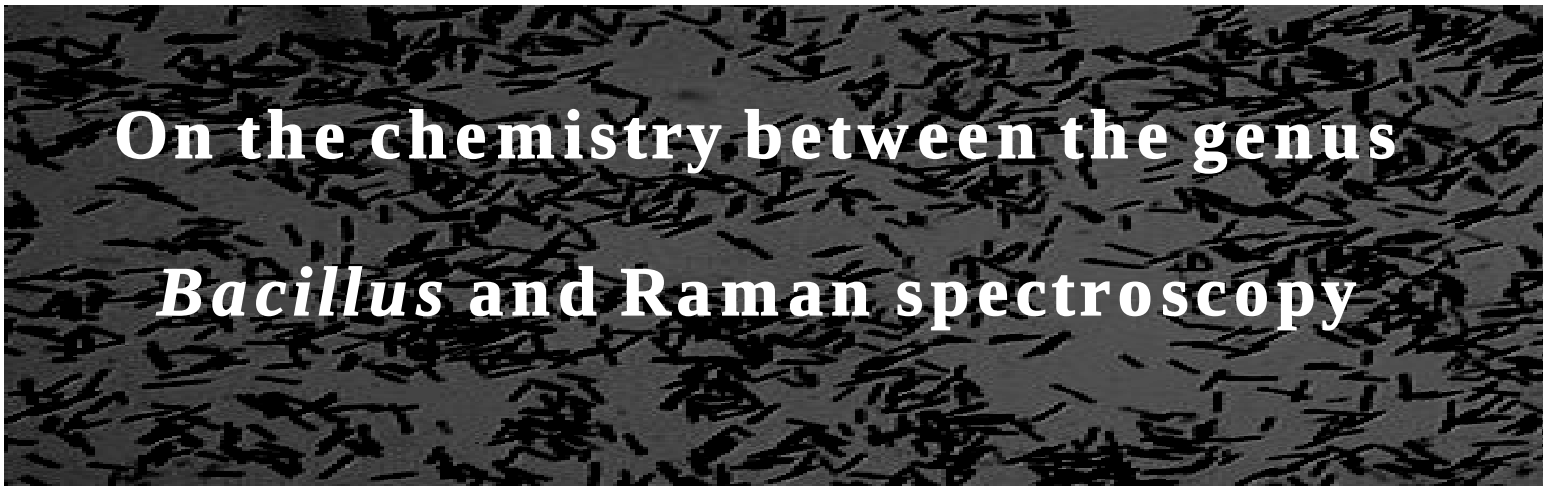
Samenvatting 189

Dankwoord 194

Appendix 199

Introduction

The aim of this thesis is to show, by means of well-chosen model systems, whether Raman spectroscopy can be considered as a potential new tool for identification of bacteria and/or in taxonomic studies of bacteria. Chapter I overviews the most relevant bio-molecules to characterize bacterial diversity and explains how these bio-molecules are analyzed for taxonomic purposes. Furthermore, the analytical value and theoretical background of Raman spectroscopy and how it can contribute to a more profound analysis of bacteria in the context of polyphasic studies are discussed. Chapter II reports on the instrumental protocol as used today in our laboratory and that was developed during this PhD work. A routine for spectral standardization is described, allowing correction for small spectral shifts. Hence, the long term stability of the spectral library is aimed and the instrumental stability is addressed. The methods described in Chapter II are subsequently used in the chapters IV and V. As a logical next step, the reproducibility of Raman analysis and its potential as a taxonomic tool for bacteria is studied in Chapter III. In particular, the effect of growth conditions on the achievable taxonomic resolution is investigated and the basis of the experimental protocol that is used during later experiments is established. The experiments that lay at the foundation of this chapter were performed at the Center for Optical Diagnostics and Therapy (CODT), Erasmus MC, Rotterdam. Reliable identification is one of the main issues of bacterial taxonomy. Especially identification of closely related and phenotypically very similar bacterial species is often problematic. Hence, the discrimination by Raman spectroscopy at the bacterial species level is evaluated by analyzing the phylogenetically homogenous '*Bacillus subtilis*'-group. The selection of this taxon as a first model system is discussed and the proposed protocols are validated by a blind study in order to obtain a realistic view on the potential of Raman spectroscopy as an identification tool for bacteria (Chapter IV). In Chapter V, the potential of this vibrational technique as a tool in polyphasic taxonomic studies is investigated within the '*Bacillus cereus*'-group. It is demonstrated that the groupings obtained by cluster analysis of Raman spectral data are in good agreement with those obtained based on fluorescent amplified fragment length polymorphism (FAFLP) analysis. In order to objectively determine the relation between Raman spectroscopy and this genomic fingerprinting technique, a correlation study is performed between the original data matrices revealed by both techniques. This thesis is concluded by a critical discussion (Chapter VI), a global summary, future prospects and suggestions for improvement (Chapter VII°). In the Appendix, the mathematical procedures (e.g. UPGMA, Canberra metric) used in this work are explained in short.



On the chemistry between the genus *Bacillus* and Raman spectroscopy

Chapter I

I. On the chemistry between the genus *Bacillus* and Raman spectroscopy

The linkage between molecular spectroscopy and its potential as an identification (or even a taxonomic) tool needs the understanding of which molecules contribute to the overall cell-composition of micro-organisms. This chapter overviews the most important groups of biochemical building blocks from a taxonomic point of view (DNA, proteins, lipids, polysaccharides, teichoic acids, etc.) together with their possible presence in certain genera and species. Some modern techniques (FAME, SDS-PAGE, FAFLP, ARDRA …) for the identification and characterization of strains via these taxonomic markers are discussed, together with their taxonomic resolution. Hence it is aimed to provide a better understanding on the potential of Raman spectroscopy in modern microbial laboratories. Next, to provide the reader an insight in this molecular spectroscopic technique the theoretical principles of Raman spectroscopy are summarized. Finally, the present state of vibrational methods as a microbial tool and some problems are sketched as an introduction to the experimental work.

1. A chemical view on cell constituents

1.1. Introduction

The main constituent of prokaryotic cells (e.g. Bacteria) is water, which represents about 70% of the total weight of an actively growing cell. The other 30% is represented by macromolecules and in a lesser amount monomers as well as a variety of inorganic ions. Table 1 gives an indicative overview of the chemical composition of a prokaryotic cell, expressed as the percentage of dry weight. Prokaryotic cells can produce thousands of proteins, which are polymers of different single amino acids, but not all are produced at the same moment or amount. Proteins are found throughout the cell in both structural as catalytic (enzymatic) roles. Nucleic acids are the polymer version of nucleotides and are represented in the cell by RNA and DNA. DNA makes up a relatively insignificant amount of the bacterial cell, but plays a critical role to the function of the cell as the repository of genetic information needed for reproduction[89].

Lipids are another group of macromolecules of key importance for the bacterial cell since they play a crucial role in membrane structure as well as in serving as storage depots for excess carbon. Polysaccharides are polymers of sugars and are primary present in the cell wall. However, polysaccharides like glycogen can be major storage forms of carbon and energy in the cell.

Table 1: Chemical composition of a prokaryotic cell[a]

Molecule	Percent of dry cell weight[b]
Total macromolecules	**96**
Protein	55
Polysaccharide	5
Lipid	9.0
Lipopolysaccharide	3.4
DNA	3.1
RNA	20.5
Total monomers	**3**
Amino acids and precursors	0.5
Sugars and precursors	2
Nucleotides and precursors	0.5
Inorganic Ions	**1**
Total	**100%**

[a] Data from Neidhardt, F.C. et al. (eds), 1996. *Escherichia coli* and *Salmonella typhimurium* – Cellular and molecular Microbiology, 2nd edition. American society for Microbiology, Washington, DC.

[b] Dry weight of an actively growing cell of *E. coli* 2.8×10^{-13}g, total weight (70 % water) = 9.5×10^{-13} g

1.2. Nucleic Acids (RNA and DNA)

Nucleotides are the elementary building-blocks of nucleic acids. In a nucleotide, a base (Figure 1) is attached to a pentose sugar by a glycosidic linkage. The exact sequence of nucleotides in a DNA or RNA molecule is referred to as its primary structure and represents the genetic information necessary to reproduce an identical copy of the organism[89].

While DNA carries the genetic information necessary for cell function and reproduction, RNA is an intermediary molecule to convert the genetic information contained by the DNA into defined amino acids sequences in proteins. RNA is, next to proteins, the most abundant group of macromolecules in an actively growing prokaryotic cell. This is easily understood since there are thousands of ribosomes in each cell and ribosomes are composed mainly by RNA (rRNA) and protein. Smaller amounts of RNA are found in the cell as messenger RNA (mRNA) and transfer RNA (tRNA), both playing a key-role in protein synthesis[89].

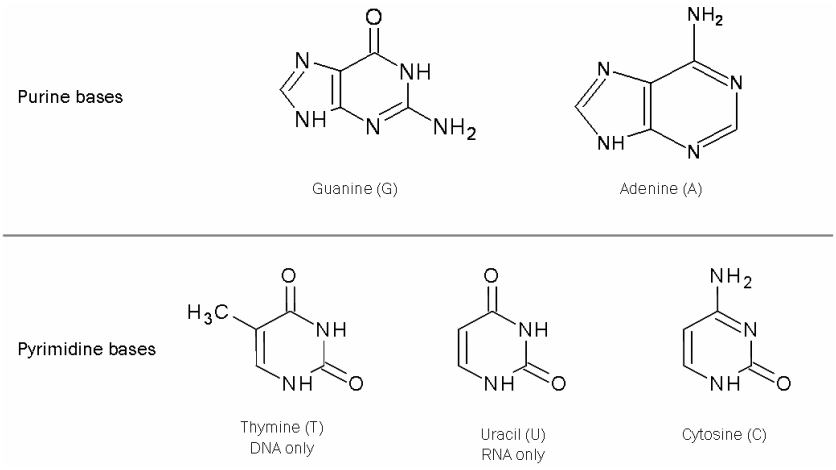


Figure 1: Overview of the purine and pyrimidine bases of DNA and RNA. Uracil replaces thymine in RNA.

The application of genomic techniques to determine nucleotide sequences (e.g. 16S rDNA / -rRNA sequencing) or similarities between the whole bacterial genome (DNA-DNA homology studies) are powerful taxonomic tools. To better understand the success of these genomic methods, it is important to pinpoint some drawbacks of classic phenotypic methods:

1. They all rely on cultivation and most of them are time consuming. Application of phenotypic methods (including Raman spectroscopy) is hence restricted to cultivable bacteria.

2. Only a minor fraction (0.1 – 10%) of the bacteria can be cultivated using standard techniques[2].

3. To form plate colonies many generations are necessary and the organism may deviate from its physiology from the population in nature.

In contrast, genotypic techniques are more 'stable' and sometimes less-time consuming. They are very useful to assign strains into specific groups and the sequence of RNA and chromosomal DNA is not affected by growth conditions. As a consequence of technological progress, modern taxonomy is dominated by these methods, primarily because the view on taxonomy is that it should reflect the natural relationships between strains/species as encoded in their DNA[163].

1.2.1. Mol % G + C composition

A classic parameter in the description of taxa is the determination of the guanine (G) and cytosine (C) content of the overall bacterial genome, because it is a stable biological parameter of a particular organism. In well defined bacterial species the variation in the G + C content should not increase more than 3 % and no more than 10 % within a well-defined genus[150].

The direct chemical composition of DNA can be determined by HPLC[98] after which the mol % G +C content is calculated from the corresponding peak areas of the nucleosides. The % G + C range for the different genera that emerged from *Bacillus sensu lato* and for other genera belonging to this phylogenetic group are shown in Table 2.

Table 2: mol % G +C distribution of the genus *Bacillus* and relatives[38]

Taxon	% G + C range
Alicyclobacillus	51.6 - 62.3
Amphibacillus	36
Bacillus sensu stricto (rRNA groups 1 and 2)*	33.2 - 64.6
Brevibacillus	42.8 - 57.4
Gracilibacillus	38 - 39
Halobacillus	40 - 43
Paenibacillus	39.3 - 54
Virgibacillus	37
Thermophiles (*B. stearothermophilus* and relatives; rRNA group 5)*	36 - 53.7

* see Ash *et al.* (1991, 1993)

1.2.2. DNA-DNA hybridization

DNA-DNA hybridization is considered as 'the gold standard' in the phylogenetic definition of new species and when a clear allocation of new strains to the species level is needed, DNA-DNA hybridization studies are recommended. Microbial taxonomists define a species quite arbitrarily as strains that exhibit DNA-DNA homology values equal or greater than 70%[167].

Due to the experimental limitations, DNA-DNA relatedness studies are rather impractical when analyzing large amounts of strains and they suffer from some important drawbacks. First, DNA-DNA hybridizations are time consuming and they are only performed by laboratories specialized in bacterial systematics[16]. Second, they are not sensitive enough to detect close relationships between strains and populations. Third, hybridization always occurs between the DNA of an (unknown) research strain and the type strain of a particular species. Hence, they provide pairwise similarities between strains, but no descriptive information of individual strains and thus – in contrast to fingerprints – no databases can be generated with the purpose of identification[164].

1.2.3. 16S rDNA / rRNA sequences

Sequence analyses of ribosomal nucleic acids were established in the late 1980's as a useful molecular chronometer to deduce phylogenetical relationships because changes in nucleotide sequences are deem to occur in a clocklike manner[175].

Two organisms that show 16S rDNA sequence similarities of 97% or lower have less than 70% DNA-DNA relatedness and should be considered as belonging to different species[167]. However, the comparison of 16S rRNA sequences show that, for example in the genus *Bacillus*, closely related species (e.g. *B. atrophaeus* and *B. mojavensis*, *B. lentus* and *B. insolitus*, *B. psychrophilus* and *B. psychrosaccharoliticus*) could not be distinguished from eachother by the conserved regions which are important for phylogenetic investigations[177].

Comparison of 16S rDNA-similarity and DNA-similarity data (*Bacillus psychrophilus* and *Bacillus globisporus*) indicate that a nearly complete 16S-rDNA sequence is not always suitable for identification at the species level (99.8 % 16S rDNA similarity – 23 % DNA similarity)[50]. However, closely related species can generally be differentiated using the highly variable 16S rDNA sequences (Hypervariant region (HV region))[16,58]. Still, some species (*B. atrophaeus*^{T} and *B. mojavensis*^{T}) remain indistinguishable based on their highly variable 16S rDNA sequence[58]. In this context it is important to mention that two species which share nearly identical 16S rDNA sequences can be clearly distinguished based on DNA-relatedness data.

1.2.4. Repetitive element sequence-based polymerase chain reaction (Rep-PCR) fingerprinting

Rep-PCR fingerprinting uses DNA-primers corresponding to naturally occurring repetitive elements in bacteria, such as BOX, ERIC (enterobacterial repetitive intergenic consensus), REP (repetitive extragenic palindromic) and $(GTG)_5$. These primers enable the amplification (via the polymerase chain reaction - PCR) of differently sized DNA fragments lying between these elements.

PCR is an enzyme-driven process for amplifying short regions of DNA *in vitro*. The method relies on knowing at least partial sequences of the target DNA *a priori* and using them to design oligonucleotides primers (e.g. $(GTG)_5$ primer) that hybridize specifically to the target sequences[179]. In PCR, the target DNA is copied by a thermostable DNA polymerase enzyme, in the presence of nucleotides and primers. Trough multiple cycles of heating and cooling in a thermocycler to produce rounds of target DNA denaturation, primer hybridization, and primer extension, the target DNA is amplified exponentially (Figure 2).

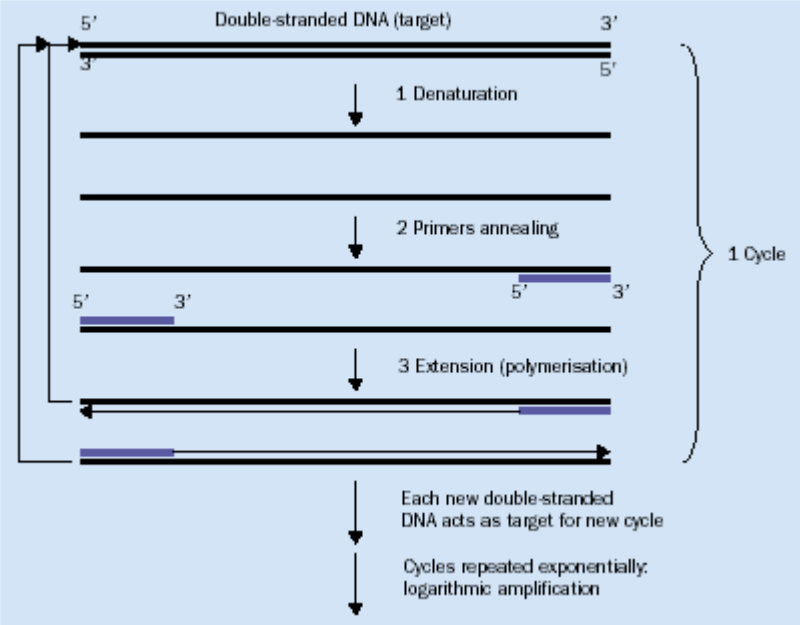


Figure 2: Schematic of PCR. Each PCR reaction consists of three major steps: (1) denaturation of template DNA into single-stranded DNA; (2) primers annealing to their complementary target sequences; (3) extension of primers via DNA polymerization to generate a new copy of the target DNA. At the end of each cycle the newly synthesised DNA acts as new targets for the next cycle. Subsequently, by repeating the cycle multiple times, logarithmic amplification of target DNA occurs. Reproduced from Yang and Rothman[179].

Rep-PCR was suggested as suitable for high throughput of strains and as a reliable tool for classifying a wide range of Gram negative and several Gram positive bacteria[112,165]. It is shown that rep-PCR is a reproducible and simple method to differentiate at the species, sub-species and potentially up to the strain level as shown within e.g. the genus *Bifidobacterium*[96], *Lactobacillus*[53], *Bacillus*[20,34], etc.

1.2.5. *Randomly amplified polymorphic DNA (RAPD)*

The aim of this technique is to reduce the visualized banding of the complex genetic pattern in order to obtain a workable blueprint of the variation of the bacterial genome. This can be achieved, for example, by the application of a hybridization procedure to reduce the number of visualized bands (ribotyping, below, § 1.2.8) or by the selective amplification of parts of the bacterial genome as in rep-PCR (see §1.2.4.), AFLP (see below, § 1.2.6) and RAPD.

In RAPD, oligonucleotides of about 10-20 bases long are used as arbitrary primers in PCR[172]. The primers anneal to the complementary or partially complementary sequences in the target DNA (the complete genome) and the DNA between the binding sites can be amplified[135]. The RAPD fingerprint patterns consist of an array of anonymous DNA amplicons.

Generally, RAPD fingerprints are commonly used for the typing of strains and, to some extent, to differentiate strains within the same species[62,131]. However, the banding pattern is not always very reproducible when no precautionary measures are taken for careful optimization and standardization. For example, differences in thermal cyclers, DNA-polymerases and concentrations, DNA preparation methods can cause variations in the RAPD patterns[38,135,160]. As result, the obtained banding patterns are not suited for comparative analysis in the long term, databases are of limited use and data exchanged between different laboratories have to be interpreted with great care[38].

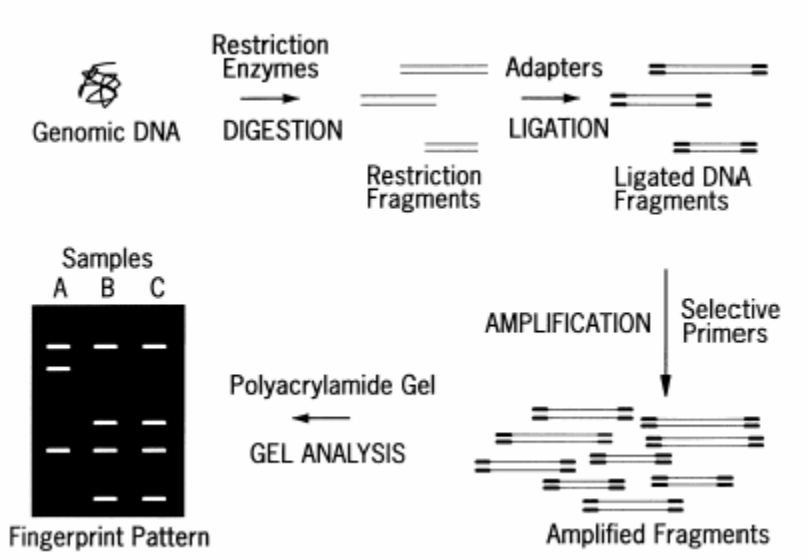
1.2.6. Fluorescent Amplified fragment length polymorphism (FAFLP)

FAFLP is a molecular technique for fingerprinting DNA's of any origin or complexity. The main advantages are its reproducibility and its capacity to inspect an entire genome for polymorphism. The technique is based on the selective amplification of restriction fragments from a total digest of DNA using the polymerase chain reaction (PCR)[166]. Compared to the rep-PCR method mentioned above, the FAFLP technique is much more complex, but also more reproducible. FAFLP can be subdivided in four steps (Figure 3)[12,38]:

1.Digestion of genomic DNA: a restriction of the bacterial genome of high purity is performed by two different restriction enzymes, randomly yielding three different kinds of DNA-fragments in relation to their sticky ends.

2.Ligation of adapters: short, double stranded oligonucleotides (adaptors) homologous to one 5'- or 3'-end generated during restriction digestion are ligated to the resulting sticky ends (templates).

3.<u>Selective amplification</u>: selective PCR amplification is performed by using primers complementary to the adapter and restriction site sequence, as well as an extra base. Only those fragments with complementary nucleotides extending beyond the restrictions site will be amplified by the selective primers under stringent annealing conditions.



<u>Figure 3</u>: A schematic displaying the four basic steps of AFLP: digestion, ligation, amplification and gel analysis. Figure reproduced from Blears *et al.*[12]

4.<u>Gel analysis</u>: due to selective amplification, a strong reduction of the complexity of the banding patterns is obtained, resulting in a workable and reproducible pattern of about 40-100 bands. This pattern can be digitized and handled by suitable computer software for numerical analysis[72].

The produced fingerprints are reliable, reproducible both within and between laboratories and are relatively easy and inexpensive to generate[12]. It has shown to differentiate between *Xanthomonas* species[72], *Aeromonas* species[68] and *Acinetobacter* species[73] and its results are generally consistent with those obtained by DNA-DNA hybridization[23,125]. Although it has been used to discriminate between *Bacillus larvae* ssp. *larvae* and *Bacillus larvae* ssp. *pulvifaciens*, in an epidemiological study of *Bacillus cereus* and for the genetic comparison of *Bacillus anthracis* and its closest relatives, the application of FAFLP on taxonomic studies within *Bacillus* is generally restricted[38].

1.2.7. Restriction fragment length polymorphism (RFLP) of the complete bacterial genome and pulsed field gel electrophoresis (PFGE)

During RFLP, purified DNA of the complete genome is restricted by selected endonucleases, yielding fragments that can be separated in an agarose gel. This generates a banding pattern based on different fragment lengths, determined by the restriction sites identified by the particular endonucleases used. Variations in the banding pattern among strains can occur as a result of e.g. mutations in restriction site sequences, deletion of restriction sites, etc.[54]. RFLP analysis results in very complex patterns of DNA fragments that are difficult to compare because they appear more as smear-like patterning[38].

PFGE is a DNA 'fingerprinting' technique that uses rare-cutting restriction enzymes (e.g. *Apa*I, *Asc*I, *Eag*I)[11] on the entire DNA genome[160,38] hence drastically reducing the number of DNA fragments. Figure 4 illustrates an example of a PFGE profile. Actually, due to the complexity of the RFLP patterns, rare cutting enzymes were used to simply the resulting fingerprints, which resulted in high molecular weight fragments. This required the development of a specific technique, known as PFGE, for the separation on agarose gels[156].

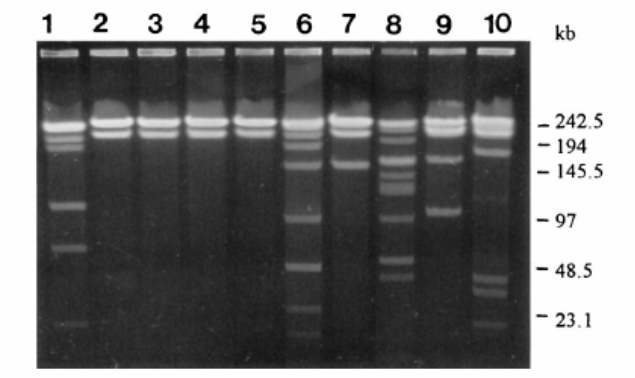


Figure 4: Example of PFGE profiles obtained from *Lactobacillus* strains as determined with restriction enzyme *Sfi*I. Figure reproduced from Tynkkynen[160].

PFGE has shown superior discriminatory power in comparison to other typing methods (e.g. ribotyping, RAPD) in strain differentiation[102,160], but due to the labour intensity of PFGE it is not a feasible technique for large scale typing of isolates.

1.2.8. Ribotyping

Ribotyping, also referred to as 'molecular fingerprinting', is a variation of RFLP analysis of the genomic DNA, where certain fragments are highlighted by probing, to obtain less complex patterns that are easier to interpret.

Genomic DNA is cut with restriction enzymes such as *Eco*RI, *Pvu*II etc. After separation on an agarose gel, the fragments are denatured and blotted (cf. Southern Blotting) on a filter A final hybridization step with a labelled fragment of the rDNA operon (= the probe), will reveal the distribution of the complement of this probe over the restriction pattern on the filter (Figure 5)[38].

These probes vary from partial sequences of the rDNA genes or their spacer regions[55,91,102] to the whole rDNA operon. The latter is used in the automated form using the Riboprinter™ (Qualicon, Wilmington, DE). If the probe contains conserved regions of rDNA, it can be used for the ribotyping of a wide range of bacteria. The procedure hence provides a DNA fingerprint of bacterial genomes coding for ribosomal ribonucleic acids (16S rDNA), which are highly conserved in microorganisms.

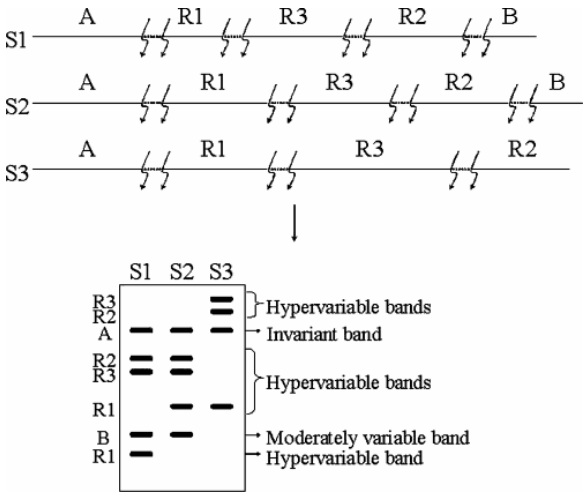


Figure 5: Restriction fragment length polymorphism with southern blot hybridization to a complex probe. S1 – S3, strains 1 – 3. Restriction sites of the enzyme chosen are represented by broken arrows. A, invariant sequence; B, moderately variable sequence; R1 – R3, hypervariable sequences, generally repetitive sequences. Strains S1 and S2 would e considered as highly related, since they only differ in one hypervariable band. In contrast, S1 and S3 can be considered as unrelated, since they share only the invariant band. Reproduced from Gil-Lamaignere *et al.*[54]

The discriminatory power of the technique depends on the size of the probe, but also on the restriction enzyme(s) used (e.g. *Eco*RI, *Pvu*II or *Mlu*I). Also, the number and the distribution of the rDNA operons over the bacterial genome, is of importance. Indeed, bacteria with only a few rDNA operons may show a very limited number of bands in their ribopattern[38].

Ribotyping has been shown to be useful for differentiating at the species level for example within *Sphingomonas*[17], at the sub-species level within the genera *Leuconostoc*[10] and *Lactobacillus*[83], and at the strain level within *Lactobacillus*[160].

1.3. Proteins

1.3.1. Proteins

Proteins are polymers of amino acids, linked to each other by peptide bounds. Most amino acids merely consist of carbon, hydrogen, oxygen and nitrogen, altough 2 of 21 common amino acids found in cells also contain sulphur (methionine and cysteine) and selenium (selenocysteine). Different amino acids only vary in the nature of their side chain attached to the α-carbon. These side chains vary considerably from as simple as a single hydrogen atom in glycine (Gly) to aromatic ring structures in phenylalanine and tryptophan (Figure 6) and determine the chemical properties of the different amino acids[89].

For example, side chains containing a carboxylic group (e.g. aspartate (Asp), glutamate (Glu)) render the amino acid acidic. Alternatively, cysteine (Cys), that is characterized by the presence of sulfhydryl group (- SH) is frequently important in connecting one chain of amino acids to another by disulfide linkage (R –S- S –R). Proline (Pro) is an exception on this rule, since it lacks a free amino group; hence it is often referred to as an imino acid, rather than an amino acid.

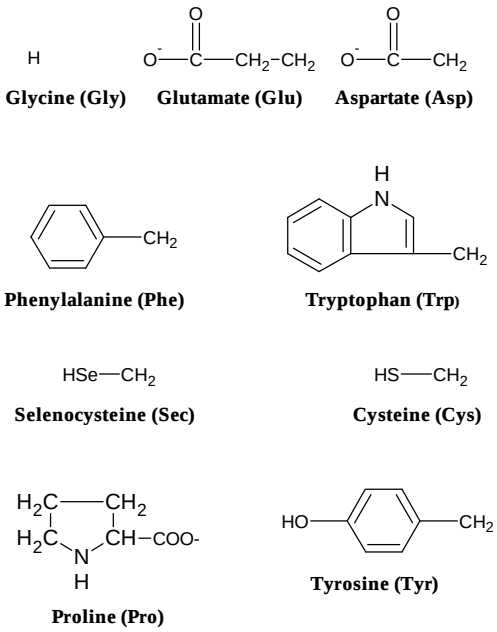


Figure 6: Chemical structure of the side chains of eight amino acids and proline. The three letter codes for the amino acids are given between brackets. For the imino acid proline the entire structure is shown, not just the side chain.

1.3.2. Polyacrylamide gel electrophoresis (SDS-PAGE)

The genetic code for cellular proteins represents a large part of the bacterial genome and under strictly standardized conditions closely related bacteria will have similar protein contents. Analysis of the whole-cell protein content hence reflects the genetic organization of an organism[113].

Although different approaches exist for protein analysis based on e.g. water soluble protein profiles[101], multi locus enzyme electrophoretic patterns[101,63] or differences between the outer membrane protein composition[44,114], profiles from the whole bacterial cell proteins are the most widely used in bacterial identification, because they probably yield the largest amount of information.

Polyacrylamide gel electrophoresis (PAGE) of whole-organism proteins is a powerful, cheap and relatively simple technique. Prior to PAGE-analysis, proteins are usually released and solubilized by detergents such as sodium dodecyl sulphate (SDS). Gels are stained to reveal the protein bands and the obtained patterns can be evaluated visually or computer assisted[121]. The method is considerably faster than DNA-DNA reassociation for species delineation and most often a good correlation between high similarity in whole-cell protein content and DNA-DNA hybridization studies is reported (e.g. by Owen *et al.*)[117].

Another advantage of PAGE is that a set of data is obtained for each strain (a protein 'fingerprint'), whereas DNA-DNA hybridization is based on comparisons with reference strains. The results of sodium dodecyl sulphate polyacrylamide gel electrophoresis, shortly SDS-PAGE, are therefore amenable to numerical analysis and can be used to establish databases for identification. Moreover, methodological differences between laboratories appear to have little effect on the identification[28]. The comparison of whole cell protein patterns (SDS-PAGE) has proven to be stable and extremely reliable for comparing and grouping of closely related strains[22].

The comparison of whole-cell protein analysis is relevant only when highly standardized conditions of growth are combined with a rigorously standardized procedure for analysis. The latter can be achieved by the inclusion of a carefully chosen bacterial pattern (e.g. *Psychrobacter immobilis* LMG 1121) on the outer lanes and the central lane of the gel. If a molecular marker is loaded as well, the molecular weights can easily be estimated[38]. In general an 80% protein pattern similarity in an unweigthed pair group method analysis (UPGMA) can be considered as an acceptable basis for species delineation and generally corresponds to 70% or more DNA relatedness between the groups[38].

1.4. Lipids

1.4.1. Fatty acids

Fatty acids (FAs) are the main constituents of bacterial lipids and their composition in bacterial cells may significantly differ between taxa. Non-hydroxylated fatty acids with chain lengths of up to 20 carbon atoms that are found in plasma membranes and lipopolysaccharides, are taxonomically very useful and can be analyzed by gas/liquid chromatography of their methylesters (FAME). As an example, the mean and standard deviation values of fatty acids for five *Bacillus* species are given in Table 3.

Table 3: Mean values for fatty acids for five *Bacillus* species expressed as (area %). Standard deviations are given between brackets[126].

Species	i12:0	i13:0	a13:0	i14:0	14:0	i15:0	a15:0	15:0	U	i16:1	i16:0	16:1B	a16:1	16:0	a17:1	17:1B	i17:0	a17:0	18:0
B. amyloliquefaciens (n = 4)	--	0.21(0.73)	--	1.19 (0.41)	0.72 (1.5)	30.50 (6.38)	35.49 (4.60)	0.42 (0.86)	--	0.41 (0.83)	2.56 (1.13)	1.92 (1.83)	--	6.21 (7.86)	1.77 (0.57)	--	10.85 (2.01)	7.59 (4.19)	0.15 (0.52)
B. licheniformis (n = 21)	--	--	--	0.46 (1.73)	0.02 (0.48)	38.04 (17.45)	30.40 (13.99)	--	--	0.64 (2.91)	2.04 (7.52)	1.73 (6.23)	0.13(1.21)	2.03 (2.00)	2.10 (5.27)	0.62 (2.65)	10.05 (7.67)	10.17 (6.02)	--
B. megaterium (n = 26)	--	0.02 (0.55)	0.72 (5.42)	6.01 (7.58)	1.62(3.05)	33.59 (22.28)	43.37 (18.57)	0.11 (2.13)	0.55 (4.35)	0.86 (5.32)	0.88 (3.68)	2.85 (5.19)	0.40 (5.62)	2.56 (5.29)	0.45 (3.12)	0.42 (4.09)	1.32 (3.96)	3.00 (5.54)	
B. coagulans (n = 7)	--	--	--	0.41 (1.29)	1.36 (1.17)	8.78 (6.37)	54.71 (18.06)	--	--	--	3.13 (5.59)	1.08 (4.62)	--	1.80 (1.27)	--	--	0.79 (1.94)	27.92 (12.60)	--
B. pumilus (n= 26)	0.03(0.76)	0.14 (1.12)	0.46(2.35)	0.73 (1.27)	0.46 (1.04)	56.89 (19.98)	23.38 (11.54)	--	0.57 (3.51)	0.40 (2.03)	1.37 (2.13)	0.99 (1.85)	0.03 (0.64)	1.09 (2.70)	3.45 (5.05)	0.80 (2.55)	5.21 (6.24)	3.46 (3.78)	0.09 (2.17)

i12:0, 10-methyl undecanoic acid; i13:0, 11-methyl dodecanoic acid; a13:0, 10-methyl dodecanoic acid; i14:0, 12-methyl tridecanoic acid;

14:0, tetradecanoic acid; i15:0, 13-methyl tetradecanoic acid, a15:0, 12-methyl tetradecanoic acid, 15:0, pentadecanoic acid; U = unknown;

i16:1, 14-methyl pentadecenoic acid, i16:0, 14-methyl pentadecanoic acid; 16:1B, hexadecenoic acid (Branched); a16:1, 13-methyl pentadecenoic acid

16:0, hexadecanoic acid; a17:1, 14-methyl hexadecenoic acid; 17:1B, heptadecenoic acid (Branched); i17:0, 15-methyl hexadecanoic acid;

a17:0, 14-methyl hexadecanoic acid; 18:0, octadecanoic acid

n = number of strains tested

1.4.2. Fatty acid analysis (FAME)

FAME is the most commonly used method for the identification of micro-organisms based on FA-profiles and involves:

1. saponification of the whole cell fatty acids,
2. esterification with an alcohol (e.g. methanol),
3. extraction of fatty acid methyl esters (FAMEs) with an organic solvent (e.g. hexane – methyl-tert-butyl ether (50/ 50 % v/v)),
4. separation by gas chromatography (GC),
5. identification (qualitatively as well as quantitatively)[113].

The variability in chain length, double bound position, hydroxylation and substituent groups (e.g. cyclopropane) has proven to be very useful for the characterization of bacterial taxa[153]. The cellular fatty acid methyl ester constituent is a stable parameter provided that highly standardized culture conditions are used. When different growth conditions and/or test methodologies are used considerable variances in the FA-profiles appear[99]. As a result, databases for psychrofiles, mesophiles and thermophiles become incompatible due to the need for a standardized incubation temperature. However, when the growth conditions deviate from those of the library, FA-profiles are still comparable to each other provided they are obtained under identical growth conditions[34]. The method has reached a high degree of automation[163] making it cheap and relatively rapid. The distribution of these complex lipids, which show wide structural variation, is valuable for a first allocation at the genus level and in some groups at the species level.

1.4.3. Lipids

Lipids differ mainly from other biological macromolecules (proteins, DNA) since they can not be considered as a polymer of their monomers (fatty acids). Instead, simple lipids, also referred to as triglycerides, consist of three fatty acids linked to glycerol, a C_3 alcohol (Figure 7). Complex lipids are triglycerides that contain additional elements such as phosphorous, nitrogen or sulphur or small hydrophilic carbon compounds (e.g. sugars, ethanolamine, serine, choline)[89]. An important class of complex lipids are the phospholipids (e.g. phosphatidylethanolamine, phosfatidylserine, phosphatidylglycerol, phosphatidylcholine, phosphatidylinositol), as they play an important role in bacterial membranes (Figure 10)[74,128,170,171].

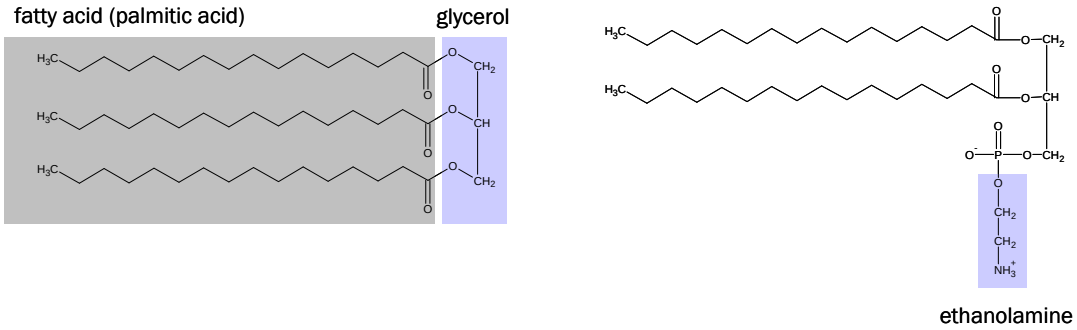


Figure 7: Left: triglyceride based of palmitic acid and glycerol. Right: phosphatidyl ethanolamine, a phospholipid, as an example of a complex lipid.

A variety of lipids is present in bacterial cells. Polar lipids are the major constituents of the lipid bilayer of bacterial membranes and have been studied frequently for classification and identification purposes[8,97].

Other types of lipids, such as sphingolipids occur in only a restricted number of taxa (e.g. the genus *Sphingomonas*) and were shown to be valuable taxanomic markers in these groups[6,84,109]. For the description and differentiation of Gram-positive taxa five phospholipids have been recognized (Table 4

Table 4) and used for classification, including *Actinomycetes*[16].

Table 4: Five recognized phospholipid types. Reproduced from Busse *et al.*[16].

Phospholipid types	Characteristic phospholipids
P I	No nitrogenous phospholipids
P II	Phosfatidylethanolamine
P III	Phosphatidylcholine
P IV	Glucosamine containing phospholipids
P V	Glucosamine containing phospholipids, phosphatidylglycerol

1.5. Polysaccharides

1.5.1. Sugars

Sugars are organic compounds containing carbon, hydrogen and oxygen in the ratio 1:2:1. C_5 sugars (pentoses) are of special significance because of their role as structural backbones of nucleic acids. C_6 (hexoses) are the monomeric constituents of cell wall polymers and energy reserves. Derivatives of simple carbohydrates can be synthesized by replacing one or more hydroxyl groups by other chemical species[89].

Sugar compositions (e.g. arabiose, galactose) from whole cells or cell-wall preparations are used for the description of Gram-positive taxa[154,155,181]. However, the resulting sugar pattern may be contaminated by non-peptidoglycan-linked saccharides from the capsules (Figure 8), cytoplasm or slimes and problems may arise due to the unavailability of some unusual sugars. The latter can be solved by co-chromatography with a sample from a strain known to posses this sugar[16].

Figure 8: Mucoid-type colonies of an encapsulated *Bacillus* species[79].

1.5.2. Polysaccharides

Bacterial cellular polysaccharides are composed of a variety of sugar monomers and are essential components of most bacteria (Gram positive and Gram negative)[49]. Table 5 gives an overview of some sugar monomers as they occur as characteristic components in isolated macromolecules (polysaccharides and glycoproteins) in some bacterial species and/or genera.

Gram positive bacteria (e.g. *Bacillus*, *Actinomycetes*) have polysaccharides that are absent in Gram-negative bacteria and that are covalently bound to the cell wall skeleton (peptidoglycan (PG) layer). The peptidoglycan layer (Figure 10) is a rigid, high molecular mass network outside the cytoplasmatic membrane (CM) maintaining the shape of the cells and protecting against osmotic lysis[82]. Peptidoglycan (or mureïn) contains the glucose derivatives N-acetylglucosamine and N-acetyl muramic acid and is an important macromolecule present in the bacterial cell wall (Figure 9). Cell walls of Gram-positive bacteria contain various peptidoglycan types that may be genus or species specific[139,149].

Table 5: Overview of some carbohydrate markers identified in bacteria[49].

Compound	Source	Organism
Muramic acid (3-O-lactylglucosamine)	Peptidoglycan	Bacteria but not elsewhere in nature
L-glycero-D-mannoheptose [L,D -heptose] D-glycero-D-mannoheptose [D,D - heptose]	Lipopolysaccharide	Gram negative but not Gram positive bacteria
L,D-Heptose but not D,D-heptose	Lipopolysaccharide	*Salmonella typhimurium* other *Enterobacteriaceae*
D,D -Heptose and L,D -heptose	Lipopolysaccharide	*Pasteurella multocida*
Galactose	Vegetative cell wall polysaccharide	*Bacillus anthracis* but not *Bacillus cereus*
Fucosamine (2-amino-2,6-dideoxygalactose)	Lipopolysaccharide	*Tatlockia* but not *Legionella*
Quinovosamine (2-amino-2,6-dideoxyglucose)	Lipopolysaccharide	*Legionella* but not *Tatlockia*
Quinovosamine	Lipopolysaccharide	*Brucella abortis, Brucella suis* and *Brucella melitiensis* but not *Brucella canis*
Yersiniose A (3,6-dideoxy-4-hydroxyethyl-D-xylohexose)	Lipopolysaccharide	*Tatlockia*

Gram negative bacteria (e.g. *Escherichia*, *Salmonella*, *Pseudomonas*), in contrast, have an additional outer membrane (OM) that contains lipopolysaccharides (LPS) anchored by their lipid-end in the membrane (Figure 10). Chemically, LPS are composed of a poly- or oligosaccharide and a lipid component and are characterized by the presence of L,D-heptose and D,D-heptose, which are generally absent in Gram positive bacteria.

LPS are often on the basis of serological identification by diagnostic antisera of many bacterial species (especially pathogenic bacteria) and is of importance as a taxonomic marker[13]. The structure of certain polysaccharides can be specific for groups of closely related bacteria and type–specific for strains from the same species[9].

Hence their compositional sugars (e.g. galactofuranose, galactopyranose) may serve as chemical markers to identify specific species or genera[49]. A distinction can be made between[29,39,168]:

1. Capsular polysaccharides (CPS) that form a thick outermost shell intimately associated with, and often covalently bound to, the cell wall (Figure 8).

2. Cell wall polysaccharides (CWS) *sensu stricto* that decorate the cell envelope and may be covalently bound to the peptidoglycan layer or loosely associated with it.

3. Extracellular polysaccharides (EPS) that, in contrast to capsules, do not remain attached to the cells but are released into the medium.

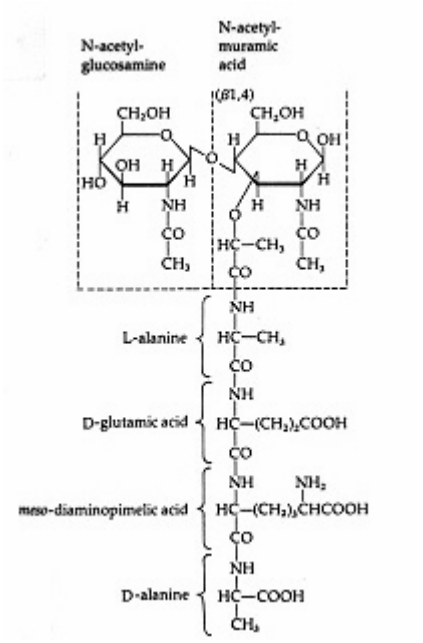


Figure 9: Structure of the muropeptide subunit of the peptidoglycan of *Bacillus megaterium*[79].

The procedure for LPS-analysis by gel-electrophoresis is time consuming, although a rapid screening method has been proposed[71,100,139].

1.6. Teichoic acids

Teichoic acids are water soluble phosphodiester-linked polymers of glycerol phosphate, ribitol phosphate, or other polyol phosphates, that may be substituted by sugars, amino-sugars or D-alanine. They occur as cell-wall teichoic acids, covalently bound to peptidoglycan and lipoteichoic acids, associated with the plasma membrane (Figure 10). Teichoic acids are diverse in structure and abundance, depending on species or strain, stage or rate of growth, pH of the medium, carbon source and availability of phosphate. By virtue of their lipid anchor, lipoteichoic acids and lipoglycans may remain attached to the cytoplasmatic membrane, but a fraction of them are found free in the cell wall or even released into the medium; teichoic acids and teichuronic acids, on the contrary, are covalently bound to the peptidoglycan layer[39].

Membrane-bound teichoic acids are present in all gram-positive species[3], whereas cell-wall bound teichoic acids are only present in some gram-positive species[81]. Teichoic acids can easily be extracted and purified[46] and analyzed by gas-liquid chromatography[47,48].

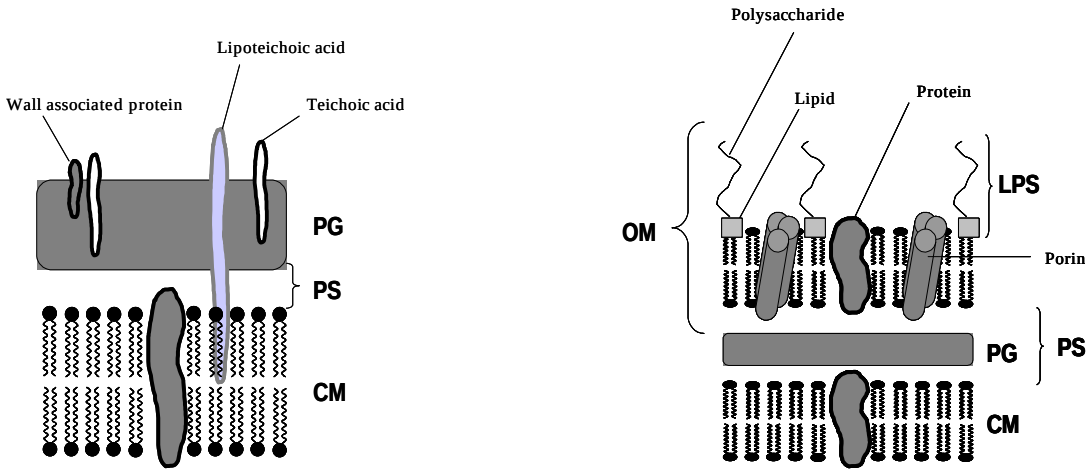


Figure 10: overview of the cell-wall from gram-positive (left) and gram- negative (right) bacteria. The cytoplasmatic membrane (CM) surrounds the cytoplasm. PS perispasmatic space, PG peptidoglycan, OM outer membrane and LPS lipopolysaccharide. Adapted from Maquelin[92] .

1.7. Polyamines

The polyamines concern a group of polycationic compounds that are primarily synthesized from amino acids and contain two or more amino groups. The number of naturally occurring aliphatic polyamines is rather low (Table 6).

Polyamines are extremely important for the cell. They affect chromosome structure and gene expression[180] and play a significant role in structural and functional organization by compacting DNA and neutralizing negative charges[24]. Furthermore they are implicated in RNA and protein biosynthesis, stabilize functional ribosome structures and control or modulate the fidelity of translation[24].

Putrescine (PUT) and spermidine (SPD) are wide spread examples of polyamines. Others, like cadaverine (CAD) and spermine (SPM) are less common. They form chemotaxonomic markers and have been investigated in a number of groups including *Bacillus*[19,61], *Vibrio*[40,178] and *Pseudomonas*[141].

However, polyamine patterns do not always allow a definite identification of bacterial strains. Some patterns are characteristic for different phylogenetic units and additional information is needed for differentiation (e.g. FA-profile, quinone patterns, etc.)[16]

Table 6: naturally occurring polyamines[14,32].

Structure	Name	Abbreviation
H_2N - $(CH_2)_3$-NH_2	1,3-Diaminopropane	DAP
H_2N - $(CH_2)_4$-NH_2	Putrescine	PUT
H_2N - CH_2-CHOH-$(CH_2)_2$-NH_2	2-Hydroxyputrescine	HPUT
H_2N - $(CH_2)_5$-NH_2	Cadaverine	CAD
H_2N - $(CH_2)_3$-NH-$(CH_2)_3$-NH_2	sym-Norspermidine	NSPD
H_2N - $(CH_2)_3$-NH-$(CH_2)_4$-NH_2	Spermidine	SPD
H_2N - $(CH_2)_4$-NH-$(CH_2)_4$-NH_2	sym-Homospermidine	HSPD
H_2N - $(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_3$-NH_2	sym-Norspermine	NSPM
H_2N - $(CH_2)_3$-NH-$(CH_2)_4$-NH-$(CH_2)_3$-NH_2	Spermine	SPM
H_2N - $(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_4$-NH_2	Thermospermine	TSPM
H_2N - $(CH_2)_4$-NH-$(CH_2)_3$-NH-$(CH_2)_4$-NH_2	Canavalmine	CANV
H_2N - $(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_3$-NH_2	Caldopentamine	CLPA
H_2N - $(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_3$-NH-$(CH_2)_4$-NH_2	Homocaldopentamine	HCPA

Polyamines are analyzed by HPLC of their corresponding dansylated polyamines as described by Scherer and Kneifel (1983)[136]. The method is reported to be suitable for investigation of large number of samples, because a low amount of individual biomass (40 mg of lyophilized cells) is needed. However, sample preparation is rather complex and time consuming[16].

1.8. Quinones

Respiratory isoprenoid quinones are present in the cytoplasmatic membranes of aerobic, facultative anaerobic and some strictly anaerobic bacteria, where they play an important role in electron transport and oxidative phosphorylation. The potential of analyzing the quinone system for the characterization of bacteria is based on the different types of quinones, the length of isoprenoid side chain and on the number of saturated isoprenoid units[16]. Some quinone types are preseneted in Figure 11.

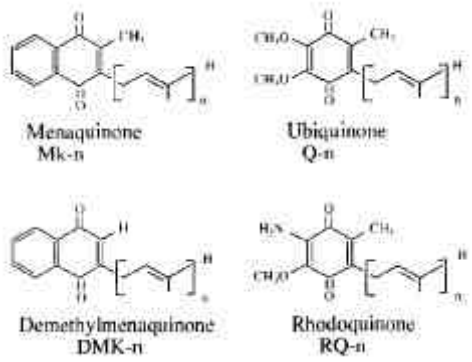


Figure 11: structure of some different quinone types (e.g. ubiquinones, menaquinones and their derivatives demethyl quinonone and rhodoquinone). Menaquinones, ubiquinones, etc. with n isoprene are abbreviated MK-n, Q-n, etc. respectively[25]. Reproduced from Collins and Jones[25].

The main types, menaquinones (MK-n) and ubiquinones (Q-n), exhibit ranges of isoprenologues, as their side chains have lengths (n) varying between 1 and 15 isoprene. Ubiquinones have only been detected in some members of the *Proteobacteria* (e.g. *Sphingomonas*, *Alcaligenes* and *Pseudomonas*), the so called α–, β– and γ-subclasses and hence their detection is a strong indication that a bacterial isolate belongs to one of these taxa.

In menaquinones, ring demethylation and the degree of side chain hydrogenation are also susceptible to variation and hence this kind of quinones is of more discriminatory value. Menaquinones and/or their derivatives are found in several members of the α–, β– and γ-subclasses of the *Proteobacteria*, but mainly in the δ– and ε–subclasses of *Proteobacteria*, the remaining Gram-negatives and in aerobic and facultative anaerobic Gram-positive bacteria (e.g. *Bacillus*).

Gram-positive bacteria are known to contain exclusively menaquinones. Here, not only the length of the isoprenoid side chains might vary, but also several units of the side chain can be saturated[16]. Depending on the taxon, the dominating presence of MK-6 to MK-14 with various degrees of saturated isoprenoid chains or combinations of these, have been reported for Gram-positives, including *Bacillus* (MK-7)[25,26,108,183].

Analysis is realized by high performance chromatography (HPLC) and the identification of the extracted quinones is determined by comparison with authentic standards, analyzed in parallel[138]. However, these references are not commercially available and have to be extracted from reference organisms with known quinone composition[16].

1.9. Diaminoacids

Diaminoacids are essential constituents of the peptidoglycan layer of Gram-positive and Gram-negative bacteria. While the amino acid composition of the peptidoglycan of Gram-negative bacteria is very uniform (meso-diaminopimelic acid is the sole diaminoacid), great diversity is revealed in Gram-positives. Indeed, differences in amino acid sequence of the peptide stems, the mode of cross-linkage between the stems and the diaminoacid present, give important information for the classification of Gram-positive bacteria[139]. Hence, diaminoacids are only useful as a chemotaxonomic marker for Gram-positive bacteria, where it is a feature for their identification. Diaminoacids detected are meso- and LL-diaminopimelic acid, L-ornithine, L-lysine, L-diaminobutiric acid and D-ornithine[139,181].

The use of cell-wall diaminoacids has shown to be extremely useful for the classification at the generic level in the coryneforms[118,148]. Diaminoacids from whole cells can be analyzed using thin layer chromatography (TLC) and co-migration of references as described by Schleifer[103]. The technique is inexpensive but rather time consuming.

2. The genus *Bacillus*

2.1. Description

The genus *Bacillus* is characterized by aerobic or facultative anaerobic, Gram-positive, rod-shaped, endospore-forming bacteria (Figure 12) that have quite different phenotypes and that are widely distributed in the environment[21]. Claus' definition has been used for many years and still remains of practical value, although it should be mentioned that some anaerobic species (*B. infernus, B. arsenicoselenatis* and *B. selenitireducens* (= microaerophile) were proposed to the genus *Bacillus*[137,140].

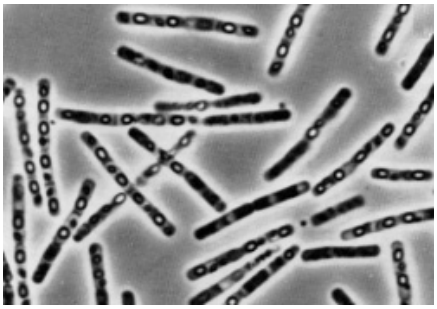


Figure 12: *Bacillus thuringiensis.* Phase photomicrograph of vegetative cells, intracellular spores (light) and parasporal crystals (dark) 1000X [78].

The vegetative cells range from 0.5 by 1.2 to 2.5 by 10 µm in diameter[159] and most of them can grow at optimal temperatures ranging from 25°C to 37°C. The majority of the strains are catalase positive, possess peretrichous flagella and sporulate under aerobic conditions, differentiating members of this genus from those of the genus *Clostridium* (exceptions given above). The genus *Bacillus* encompasses a variety of phenotypically heterogeneous species exhibiting a wide range of nutritional requirements, physiological and metabolic diversity and DNA base composition.

Many species of this genus can grow under extreme conditions that can be deducted from several of the species names given: *B. alcalophilus* (alkalophilic); *B. halodurans*[106], *B. saliphilus*[130] (halophilic); *B. psychrodurans, B. psychrotolerans*[1] (psychrophilic); *B. fumarioli*[87], *B. thermoamylovorans*[27] (thermophilic).

2.2. Phylogenetic heterogeneity within *Bacillus*

The extreme metabolic and phylogenetic heterogeneity of the representatives of this genus nowadays is reflected in the wide variety of ecological niches in which they are found and in the wide range of mol% G + C composition of the DNA of species that varies from 33.2 to 64.6%[38]. This wide range of DNA base composition and hence phylogenetic heterogeneity supports the splitting of the genus *Bacillus* into better defined genera. Comparison of 16S rDNA sequences of the type strains of the *Bacillus* species show clear phylogenetic relationships within the genus and provide a rational basis for the splitting of the genus[5].

***Bacillus* RNA group 1** constitutes the core of *Bacillus* (*Bacillus sensu stricto;* Figure 13) and contains the type species *B. subtilis.* The phylogenetic diversity of this group is still important and encompasses several well separated species clusters and single species lineages[5]. The groups consist of multiple subgroups of which a first one comprises *B. vallismortis*, *B. mojavensis*, *B. subtilis*, *B. amyloliquefaciens*, *B. atrophaeus* and *B. licheniformis*; a second group contains *B. cohnii*, *B. horikoshii* and *B. halmapalus*; a third group harbours *B. cereus*, *B. pseudomycoides*, *B. anthracis*, *B. thuringiensis*, *B. weihenstephanensis* and *B. mycoides;* a fourth one embraces *B. simplex, B. psychrosaccharalyticus,* the invalid species *B. maroccanus*[66] and two misclassified *Brevibacterium* and *Arthrobacter* species. Other species that formerly were classified as members to this subgroup were transferred to *Virgibacillus, Halobacillus,* etc.

***Bacillus* RNA group 2** contains *Bacillus* type organisms (*B. sphaericus*, *B. fusiformis*, *B. insolitus*, *B. pasteurii* and *B. psychrophilus*) that are intermixed with spherical sporeformers (*Sporosarcina*), nonsporeformingrods (*Kurthia*, *Caryophanon*) and cocci (*Planococcus*)[151]. *B. globisporus*, *B. pasteurii* and *B. psychrophilus* were recently reclassified into the genus *Sporosarcina*[182].

Figure 13: Neighbour joining tree of the genus *Bacillus, sensu stricto*. EMBL/Genbank/DDBJ accession numbers are given between brackets. Bar, 1% estimated sequence divergence

***Bacillus* RNA group 3 and 4**. Members of these groups were all reclassified as *Paenibacillus*[4] and *Brevibacillus*[143], respectively (Figure 14).

***Bacillus* RNA group 5** has recently been reclassified as the new genus *Geobacillus*[105] (Figure 14). This genus contains the species *G. stearothermophilus, G. thermooleovorans,* G. *thermocatenulatus, G. kaustophilus, G. thermoglucosidasius, G. thermodenitrificans, G. subterraneus* and *G. uzenensis.*

***Bacillus* RNA group 6** was defined by Nielsen *et al.* (1994) and mainly constitutes alkaliphilic or alkalitolerant species[107]. No formal nomenclature changes were proposed since then, but based on the distance-matrix-analyses of 16S rRNA sequences, the question raises whether this group should be reclassified to constitute more genera[76].

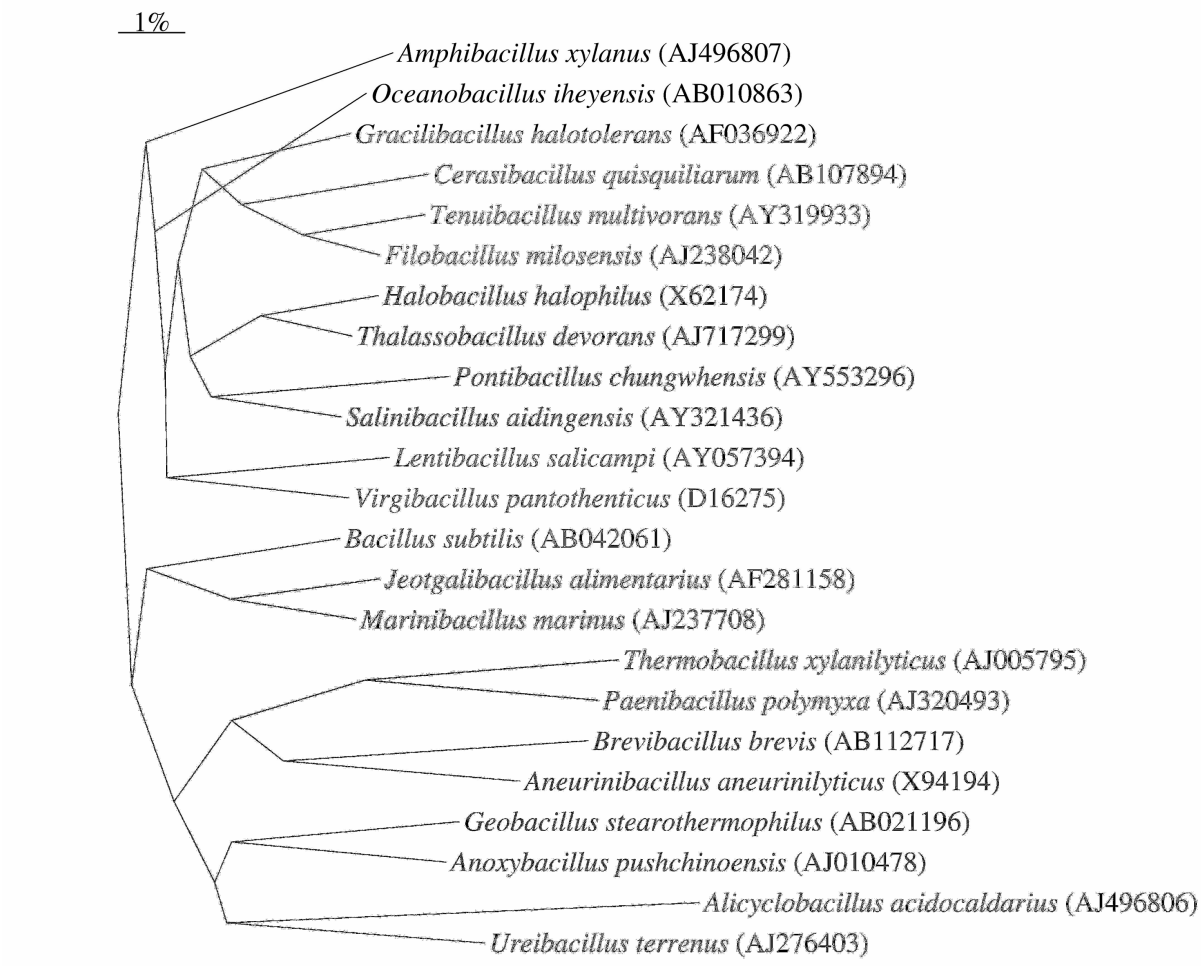


Figure 14: Neighbour joining tree including all type strains of genera that either include previous members of the genus or were newly described after the splitting up of the genus (*Bacillus sensu lato)*. EMBL/Genbank/DDBJ accession numbers are given between brackets. Bar, 1% estimated sequence divergence.

As mentioned, this only provided a basis in 1991 for the splitting of the genus which has been extensively performed since then. Due to new species descriptions the authentic genus *Bacillus* encompasses currently more than 100 species. The genus accommodates some important species as *B. subtilis* (type species), *B. anthracis*, *B. cereus*, *B. licheniformis*, *B. megaterium*, *B. pumilus*, *B. sphaericus* and *B. thuringiensis;* most of them have a large impact on human activity. The phylogenetic relationships of *Bacillus* species, based on their complete 16S rDNA sequence have been reported by Goto *et al.*[58]. While most strains enclosed in Goto's study could be discriminated, problems occurred in two groups: *B. subtilis* and *B. cereus* group.

2.3. Identification within *Bacillus*

2.3.1. Gram reaction

The majority of all aerobic spore-forming organisms (e.g. *Bacillus*) stains Gram-positive, although not necessarily in all phases of their cell cycle. Electron microscopic analysis showed that the structure of the cell wall, the so-called "Gram type", not always corresponds to the Gram-staining reaction[169].

For example, *B. hortii* has a Gram-positive cell wall structure but stains Gram negatively because the loss of the blue color occurs much faster than for more easily than the standard strain used for comparison. This is the result of a very thin peptidoglycan layer that does not retain the colour[51].

2.3.2. Morphology

Morphologic information on the colonies (colour, shapes, margins, etc.) grown on given substrates and/or microscopic information on cell morphology (flagellation, spores, etc.) provides a first indication on the identity of bacteria. It is an important and easy approach to check purity of the cultures and as first orientation towards final identification. A detailed overview on the morphology of different *Bacillus* groups is given by Fritze[51].

2.3.3. Physiology

The widely used diagnostic scheme for *Bacillus as* developed by Smith (1952) and Gordon (1973), also included biochemical and physiological tests and was effectively used during several years. *Bacillus* identification was generally perceived as complicated because of the need for special media and strain variation. Moreover, some species (e.g. *Brevibacillus* sp.) are usually inactive in routine biochemical tests so there are too few characters available to distinguish between these species, that are often phylogenetically closely related[85].

The opposite problem may also arise as is the case for *Paenibacillus* sp. that are highly active in routine tests so that distinctions between them are confined to rather variable characters. Today, the most widely used, commercially available methods for identifying members of the genus *Bacillus* are still based upon miniaturized developments of routine biochemical tests (API System, Biotype 100, Vitek, Biolog).

API (bioMérieux): This system currently allows identification of only 19 *Bacillus* species[85].

Biotype 100 (bioMérieux): Species of the genera *Brevibacillus* and *Aneurinibacillus* and the species from the *B. sphaericus* group are largely unreactive in the carbohydrate utilization tests of the API 50CHB gallery and insufficiently variable in the API 20E and supplementary tests, so that some 20 species are largely inseparable by this means.
The Biotype 100 system contains 99 tests for assimilation of carbohydrates, organic acids and amino acids. The kit has proven to be of great value in differentiating species in *Brevibacillus* and *Aneurinibacillus* and is promising for other unreactive species[85].

Vitek (bioMérieux): The Vitek system is now routinely used for the identification and determination of anti microbial susceptibility in thousands of laboratories worldwide. It identifies over 300 species of bacteria and yeasts, in both clinical and industrial environments. However, the present Vitek *Bacillus* card only identifies 17 species with confidence and is now expanded to recognize 34 well-defined species[85].

Biolog: The database currently contains 23 *Bacillus* species, *Brevibacillus brevis* and six *Paenibacillus* species[85]

2.3.4. Chemotaxonomic markers

Chemotaxonomic fingerprinting techniques applied to aerobic endospore formers include FAME profiling[76], PAGE-analysis[38] and a series of more 'exotic' fingerprinting techniques, as pyrolysis mass spectrometry (Py-MS)[145,146], MALDI-TOF-MS[41,42] and FT-IR spectroscopy[64]. The MS-based fingerprinting methods offer a variety of rapid, automated, high throughput techniques for the investigation of whole cell composition of bacteria. They have been well-evaluated and show to be valuable in bacterial systematics as demonstrated for e.g. *Bacillus*[144,145], *Actinomycetes*[134] and *Clostridium*[18].

Although, these spectrometric techniques (mass spectrometry, FT-IR spectrometry, etc.) are in full development, their increasing importance can not be denied in a polyphasic approach. They provide data on aspects of cell composition that otherwise can only be determined by slow, costly, low-throughput, conventional chemistry[90]. Their main advantages are the simplicity of sample preparation, lack of labile reagents, speed of analysis, low specimen processing cost and automation. These techniques can be applied over a broad range of species and combine high throughput with the possibility to identify strains at the genus, species and even strain level.

Today, however, only FAME analysis is supported by a commercially available database for routine identification (Microbial ID, Inc, Newark, DE). Although it does not provide a reliable stand-alone and accurate identification for *Bacillus*, it plays an important role both in polyphasic taxonomic studies of this genus (and relatives) and as a valuable screening tool of large groups of strains[75,76].

2.3.5. Serology

Some serological tests are available for *Bacillus* species. However, they are commonly plagued by problems of cross-reacting antigens and autoagglutination (in case of sporulation). A strain differentiation system for *B. cereus* based on flagellar antigens is available for investigations of food poisoning outbreaks and other *B. cereus* associated problems. The classification of *B. thuriengiensis* strains on the basis of H – antigens and antigenic subfactors is also available and superior to traditional phenotypic test schemes[85].

2.3.6. Genotypic methods

Studies of 16S rDNA and DNA have very valuable applications in the classification of aerobic, endosporeforming bacteria and have become suitable even for routine identification at the species level[45,58,86,174].

FAFLP has revealed sufficient variation for differentiating subspecies of *B. subtilis*[15]. It was also applied for the genetic comparison of *B. anthracis* and its closest relatives[67,70,77,157] and for the epidemiological typing of *B. cereus* isolates[129,162].
Rep-PCR has been used to unravel the genetic diversity of *B. sphaericus*[31], to evaluate the genetic relationship whithin the '*B. cereus*'-group[20], and for the identification and discrimination of clonally related *B. sporothermodurans* isolates in UHT treated milk[65]. More recently, rep-PCR has demonstrated the intra-species heterogeneity of *B. coagulans*[35] and the technique is generally used for screening purposes[37,36].

In a recent polyphasic approach, RAPD was applied to study the population structure of the human pathogen *B. cereus* and to asses the intraspecific biodiversity of this species[43]. Furthermore, it has been used for subtype characterization of clinical specific isolates from biofilms[33], for typing of *B. thuringiensis* serovars[52] and as one of the approaches in polyfasic, taxonomic studies[132].

The existing reports on ribotyping concern the intraspecific distribution of ribopatterning in correlation with, for example, food poisoning with *B. licheniformis*[133] or different contamination sources with *B. sporothermodurans*[59]. Since the method is automated (via the 'Riboprinter') it is easily applied as identification tool for bacteria originating from the pulp and paper industry[152].

The techniques mentioned, give only a restricted overview of some of the most commonly applied genotypic methods in the genus *Bacillus*.

3. Raman spectroscopy

3.1.1. Introduction

When monochromatic light falls on a sample, some of the light is transmitted, some is absorbed and some is scattered. The ratio of incident and transmitted light is a measure for the transmittance of the matter, which is classically determined in FT-IR spectroscopy. However, from the scattered light a small fraction (typically 10^{-6} – 10^{-8}) is shifted in wavelength, relative to the incident beam, due to the interaction with the molecular vibrations of all molecules illuminated by the monochromatic beam. If this scattered light is detected, a Raman spectrum can be obtained by measuring the intensity of the scattered radiation as a function of scattered wavelength, expressed as Raman wavenumber (cm^{-1}). When the wavelength of the incident light (λ_0) and scattered light (λ) are expressed in nanometres (nm), the Raman wavenumber (ν) is defined as:

$$Raman\ wavenumber\ (\nu) = \left(\frac{1}{l_0} - \frac{1}{l}\right) \times 10^{-7} \quad \text{Eq. 1}$$

Although depending on the nature of the sample, a Raman spectrum generally consists of sharp bands that are characteristic for the specific molecules in the sample (Figure 15).This gives Raman spectra a good specificity for qualitative analysis and discrimination among similar materials. Raman spectroscopy also provides quantitative information, since the intensity is proportional to the concentration (see below)[119].

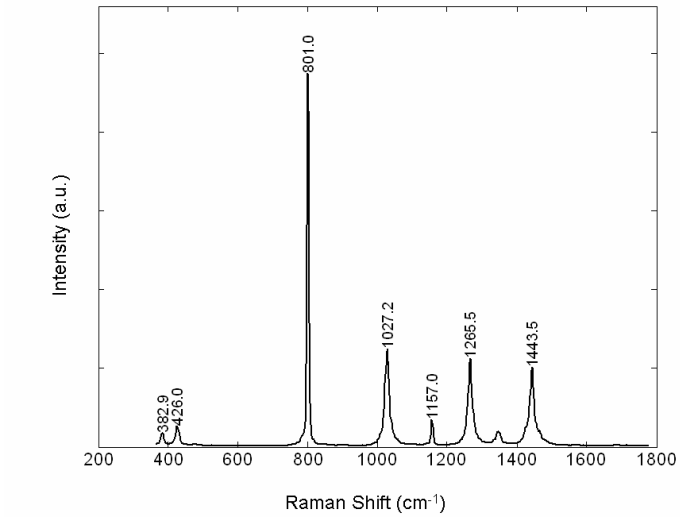


Figure 15: Representative Raman spectrum of cyclohexane as collected in our laboratory. Spectrometer calibrated as described in Chapter II. Accumulation time 60 s. a.u.: arbitrary units.

From a practical point of view, Raman spectroscopy has evolved into a fast and powerful analytical technique that is easily used by non-specialists, provided that the instrumentation is fully optimized for a specific application. Furthermore, Raman spectroscopy allows collection of spectra in less than a minute, in a non-invasive and non-destructive way.

3.1.2. Theoretical background: A quantum theoretical approach

Although the existence of Raman scattering was theoretically predicted earlier[147], it were Raman and Krishnan who first reported the observation of the Raman effect in 1928[127]. The Raman effect can best be described by means of an energy diagram as depicted in Figure 16.

For simplicity, the Raman effect is described as the inelastic scattering of light by matter. When an incident photon with energy $E = h\nu_0$ interacts with a molecule having different vibrational energy levels (v = 1, v = 2, etc.) most of the photons are scattered with the same energy as the incident beam.

However, a small fraction of photons with the same energy ($E = h\nu_0$) raises the energy state of he molecule from the electronic and vibrational ground state to a 'virtual energy level'. This virtual energy level is not a stationary, stable energy state of the molecule, but rather a distortion of the electron distribution of a covalent bond[119]. Since this is not a stable state, the molecule relaxes to the original electronic state by emitting a photon.

So, when a photon of energy, too low to excite an electronic transition, interacts with a molecule, it can be scattered in one of three ways. It can be elastically scattered (i.e. Rayleigh scattering) retaining its original energy or it can be inelastically scattered by either increasing (i.e. Stokes) or decreasing (i.e. Anti-Stokes) the energy level of the molecule. The three phenomena are illustrated in Figure 16 and summarized below[176]:

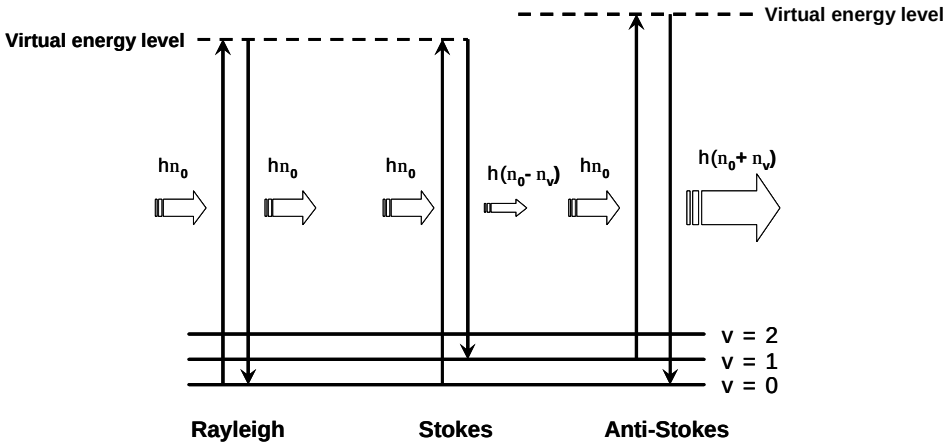


Figure 16: Overview of the different kinds of scattering that are observed during a Raman experiment. Rayleigh, Stokes and Anti-Stokes scattering occur via a virtual energy level. The ground electronic state consists of several vibrational energy levels (v =1, v = 2, etc.). The size of the arrow is proportional to the energy of the photon.

1. If the molecule returns to the vibrational energy level (v=0) from which it is started, the difference in energy between the virtual energy level and the final state is emitted by a photon with the same energy, and therefore the same wavelength as the initial photon. This event is called Rayleigh scattering.

2. If the molecule, after excitation to the virtual energy state, returns to a vibrational energy level higher than the one from which it is started, the energy difference between the virtual energy level and the final state (e.g. v = 1) is emitted by a photon with less energy than the incident photon. In this case the emitted photon has a longer wavelength than the initial photon and the vibrational energy of the molecule has increased. This event is called Stokes Raman scattering.

3. If the molecule returns to a lower vibrational energy (e.g. v = 0) level than the one from which it is started (e.g. v = 1), the energy difference between the virtual energy level and the final state is emitted by a photon with more energy than the incident photon and a shorter wavelength than the initial photon. The vibrational energy of the molecule has decreased. This event is called anti-Stokes Raman scattering.

In general discussions, 'Raman scattering' is assumed to be 'Stokes Raman scattering', because Stokes scattering is more intense than Anti-Stokes scattering. This is a consequence of the population of the vibrational ground-state which is higher than the excited vibrational levels at room temperature. The relative population of two energy levels at a certain temperature is given by the Maxwell-Boltzmann distribution.

3.1.3. The Maxwell-Boltzmann distribution

At thermal equilibrium the fraction of molecules in one vibrational energy level relative to another is given by the well-known Boltzmann distribution[119]:

$$\frac{N_1}{N_0} = \left(\frac{g_1}{g_0}\right) \times e^{[-(\Delta E)/kT]} \qquad \text{(Eq. 2)}$$

with:

N_1 = number of molecules at a higher vibrational energy level;

N_0 = number of molecules at a lower vibrational energy level;

g_1 = degeneracy of the higher vibrational level;

g_0 = degeneracy of the lower vibrational level;

ΔE = energy difference between the higher and the lower energy levels;

k = Boltzmann's constant (= 1.38066×10^{-23} J K^{-1});

T = temperature (in Kelvin, K);

At thermal equilibrium the number of molecules at a lower vibrational energy level is always larger than the number of molecules at a higher vibrational energy level and thus Stokes Raman scattering is more intense than Anti-Stokes Raman scattering.

3.1.4. Raman intensity

It was Placzek who derived for the first time an expression for Raman scattering intensity[120]. A detailed derivation is given elsewhere[88]. The intensity of Raman scattering can be expressed as:

$$I_R = \frac{2^4 p^3}{45 \times 3^2 \times c^4} \times \frac{h I_L N (u_0 - u)^4}{mn(1 - e^{-hu/kT})} [45(a'_a)^2 + 7(g'_a)^2] \qquad \text{(Eq. 3)}$$

With:

c = speed of light (= 2.99792458 X 10^8 ms^{-1});

h = Planck's constant (= 6.62608 X 10^{-34} J s);

I_L = excitation intensity;

N = number of scattering molecules in the irradiated volume;

ν = molecular vibrational frequency [Hz];

ν_0 = laser excitation frequency, [Hz];

μ = reduced mass of the vibrating atoms;

k = Boltzmann's constant (= 1.38066 X 10^{-23} J K^{-1})

T = Absolute temperature [K];

$\alpha_{a'}$ = mean value invariant of the polarisability tensor;

$\gamma_{a'}$ = anisotropy invariant of the polarisability tensor.

Equation 3 learns that the Raman scattering intensity is proportional to the number of illuminated molecules in the irradiated volume. This relationship allows quantitative analysis, on a theoretical basis. In addition, the scattered intensity depends on laser frequency - via the factor $(\nu_0 - \nu)^4$ - and the intensity of the incident light intensity, which is proportional to the laser power. Hence, when optimizing Raman scattering intensity, one can, on a theoretical basis, increase both laser power and frequency. One should keep in mind that the *measured* intensity is strongly affected by the collection optics (transmission, numerical aperture (N.A.)), reflection and transmission efficiency of the optical components (mirrors, lenses), the type of diffraction grating (blazed frequency, lines/mm) and quantum efficiency of the detector. Sample fluorescence and thermal degradation are limiting factors when laser frequency and power are selected. Often a trade-off exists between sensitivity and reduction of fluorescence and thermal degradation.

3.1.5. Raman wavenumber

The traditional units for the abscissa of a Raman spectrum are 'wavenumbers of shift from the exciting wavelength' or simply 'Raman wavenumber. Wavenumber is simply the reciprocal of wavelength (expressed in centimetres). Thus, a diode laser with an emission wavelength of 785 nm (785 x 10^{-7} cm) has a wavenumber of 12738.85 cm^{-1}.

Wavenumber is a unit of energy (E) and corresponds to[119]:

$$E = hu = \frac{hc}{l} = hcv \qquad \text{(Eq. 4)}$$

with:

h = Planck's constant (= 6.62608 X 10^{-34} J s);

ν = frequency of light;

c = speed of light (= 2.99792458 X 10^{8} ms^{-1});

λ = wavelength of light;

ω = wavenumber of light;

The values on the abscissa of a Raman spectrum are the difference between the excitation wavelength (laser wavelength) and the Raman wavelength, expressed as wavenumbers. Hence, when Raman scattering is detected with an (absolute) wavelength of 849.70 nm, using an excitation wavelength of 785 nm, the Raman wavenumber for this Raman band can easily be calculated:

$$Raman\ wavenumber = \frac{1}{785\,nm \times \left(\frac{1.0\,cm}{10^{7}\,nm}\right)} - \frac{1}{849.70\,nm \times \left(\frac{1.0\,cm}{10^{7}\,nm}\right)} = 970.00\,cm^{-1} \text{ (Eq.5)}$$

3.1.6. The application of Raman spectroscopy in microbiological laboratories

In times of PCR-based molecular techniques as 16S rDNA, rep-PCR, FAFLP and well established phenotypic methods (SDS-PAGE, FAME, API, Vitek) the question rises how 'just another' phenotypic method can contribute to routine identification / characterization in modern microbial laboratories. The answer hereto depends on the characteristics inherent to Raman spectroscopy, on the advantages towards other methods (genotypic as well as phenotypic) and the taxonomic resolution of Raman spectroscopy.

3.1.6.1. Characteristics & advantages

Raman spectroscopy is intrinsically an easy-to-use vibrational technique, is non-destructive, rapid, cost-effective, requires minimal sample preparation and is suitable for automation. The sensitivity and speed of Raman spectroscopy has increased to a level where spectra can be recorded of even one single-living cell in less than 300 seconds[115,122-124,161] and due to the confocal measurement principle a high spatial resolution is obtained. As a result, only small volumes are needed to obtain high quality spectra.

Together, high quality Raman spectra can be collected from micro-colonies (< 60 µm) which results in a significant reduction in culture time (< 7 h)[93], a considerable increase in the overall speed of analysis and thus in a high sample throughput. This is especially of interest when Raman spectroscopy is weighted against well-established phenotypic methods (e.g. FAME, SDS-PAGE) which all need at least 24 hours for standardized cell growth and additional time for sample preparation (esterification of fatty acids, preparation of gels,…).

Especially the ease of use and minimal sample handling are major advantages when compared to PCR-based genotypic techniques, for which the culture step is theoretically superfluous. These molecular techniques are known to be expensive and require highly skilled personnel, limiting their use as routine application[179].

Furthermore, Raman spectroscopy provides molecular and (optional) structural information on the complex collection of biochemical molecules within a micro-organism and consequently can be regarded as molecular fingerprints of the overall cell-composition[64,104,142,158]

In contrast to other taxonomic tools (SDS-PAGE, FAME, rep-PCR), which are focussed on one particular group of taxonomic discriminative molecules, Raman spectroscopy gives an enlarged view on micro-organisms by revealing a complete phenotypic picture with a minimum of work. Hypothetically, spectral analysis might thus include phenotypes that are not used hitherto for taxonomic studies or identification.

3.1.6.2. Possibilities of vibrational methods (Raman and FT-IR spectroscopy)

The possibility of Raman spectroscopy to identify a restricted number of species was recently illustrated. Clinical relevant bacterial strains were successfully discriminated at the species level[69,80,95] indicating the advantages of this technique in a specific niche. Raman spectroscopy can be regarded as superior to more conventional phenotypic methods (API, BIOLOG) and more broadly applicable than genomic methods[80]. The taxonomic resolution of vibrational spectroscopies (Raman and FT-IR spectroscopy) is suggested to be at the species level and perhaps even the subspecies level[94].

These methods were claimed to be highly reproducible and hence suggestions that they are able to discriminate at the strain level could be expected[64,80,116]. However, it should be underlined that these studies were often performed on a limited set of strains and hence the naturally occurring biological and ecological diversity was insufficiently included. Moreover, since the exact details on the planning of experiments are often not provided, interference of variable experimental parameters (e.g. composition of medium, sampling) with the groupings of closely related species or strains could be expected.

Oust *et al.* used in total fifty-six *Lactobacillus* strains assigned to *L. sakei*, *L. plantarum* and *L. curvatus* and performed three independent parallel experiments. Her study indicated that strain groupings (by HCA) became unclear when the library was expanded from 29 strains to 56 strains[116]. Indeed, HCA is not suited for analysis of complete vibrational spectra with respect to differentiation of species, when the main variation in the spectra is not related to variation between the species[80,116].

In addition, an FT-IR database (730 reference strains covering 220 species out of 46 genera of coryneform bacteria and related taxa) was validated internally and resulted in 93.9 % correct identification at the strain level, 98.8 % at the species level and 99.5 % at the genus level. However, when validated by simulated external validation (SEV), a correct identification of 87.3% at the species level and 95.4 % at the genus level was obtained[111]. Using SEV, the identification accuracy is considerably lower at the species level. SEV uses the average of one stain and builds the database without the average spectrum of this particular strain. The strain left-out is consequently identified at the species and/or genus level. This validation approach is comparable to a 'leave-one-strain-out' approach[95]. It is clear that SEV does not allow validation at the strain level and still gives a too optimistic view on the identification abilities since not all (possible) variations or 'uncertainties' in experimental factors are included (e.g. medium, temperature, sampling, human factor…). Some of these factors are expected to affect identification in an important way.

Other studies have reported on the application of FT-IR spectroscopy for some species of the genera *Lactobacillus*[30], *Actinomycetes*[60], *Bacillus*[7], *Streptococcus*[57] and *Acinetobacter*[173] but were based on a very limited number of species and were validated by only a few strains that were often part of the reference library as well. Hence it is still unclear of vibrational spectroscopies are indeed a competitive identification method.

Thus, based on its inherent characteristics and advantages towards classic microbial tools, Raman spectroscopy can be considered as a valuable tool in modern microbial laboratories. Although the potential of vibrational methods as microbial identification techniques, their taxonomic applications today are insufficiently supported because:

1. Often, too few species, which are often phylogenetically not sufficiently related, are included in the performed studies to adequately validate the discriminative power of vibrational methods.

2. The experiments might be performed under 'ideal' circumstances excluding certain potential experimental variations and hence lead to a less realistic view on the possibilities of these techniques.

3. Correlation of vibrational methods with general accepted phenotypic or genotypic methods are reported exceptionally[110]. As such, the taxonomic resolution of vibrational methods is insufficiently understood.

This thesis intends to contribute to the further development of Raman spectroscopy as a reliable microbial identification tool by tackling these three topics in the following chapters.

Reference List

1. **Abd El-Rahman, H. A., D. Fritze, C. Sproer, and D. Claus.** 2002. Two novel psychrotolerant species, *Bacillus psychrotolerans* sp. nov. and *Bacillus psychrodurans* sp. nov., which contain ornithine in their cell walls. Int J Syst Evol Microbiol **52**:2127-2133.

2. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev **59**:143-169.

3. **Archibald, A. R. and J. Baddiley.** 1966. The teichoic acids. Adv Carbohydr Chem Biochem **21**:323-375.

4. **Ash, C., F. G. Priest, and M. D. Collins.** 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus.* Antonie Van Leeuwenhoek **64**:253-260.

5**. Ash, C., J.A.E. Farrow, S. Wallbanks, and M.D. Collins**. 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Letters in Applied Microbiology **13**:202-206.

6. **Balkwill, D. L., G. R. Drake, R. H. Reeves, J. K. Fredrickson, D. C. White, D. B. Ringelberg, D. P. Chandler, M. F. Romine, D. W. Kennedy, and C. M. Spadoni.** 1997. Taxonomic study of aromatic-degrading bacteria from deep-terrestrial-subsurface sediments and description of *Sphingomonas aromaticivorans* sp. nov., *Sphingomonas subterranea* sp. nov., and *Sphingomonas stygia* sp. nov. Int J Syst Bacteriol **47**:191-201.

7. **Beattie, S. H., C. Holt, D. Hirst, and A. G. Williams.** 1998. Discrimination among *Bacillus cereus*, *B. mycoides* and *B. thuringiensis* and some other species of the genus *Bacillus* by Fourier transform infrared spectroscopy. FEMS Microbiol Lett **164**:201-206.

8. **Bendinger, B., R. M. Kroppenstedt, S. Klatte, and K. Altendorf.** 1992. Chemotaxonomic differentiation of coryneform bacteria isolated from biofilters. Int J Syst Bacteriol **42**:474-486.

9. **Bergstrom, N., P. E. Jansson, M. Kilian, and U. B. Skov Sorensen.** 2000. Structures of two cell wall-associated polysaccharides of a *Streptococcus mitis* biovar 1 strain. A unique teichoic acid-like polysaccharide and the group O antigen which is a C-polysaccharide in common with pneumococci. Eur J Biochem **267**:7147-7157.

10. **Bjorkroth, K. J., R. Geisen, U. Schillinger, N. Weiss, P. De Vos, W. H. Holzapfel, H. J. Korkeala, and P. Vandamme.** 2000. Characterization of *Leuconostoc gasicomitatum* sp. nov., associated with spoiled raw tomato-marinated broiler meat strips packaged under modified-atmosphere conditions. Appl Environ Microbiol **66**:3764-3772.

11. **Bjorkroth, K. J., P. Vandamme, and H. J. Korkeala.** 1998. Identification and characterization of *Leuconostoc carnosum*, associated with production and spoilage of vacuum-packaged, sliced, cooked ham. Appl Environ Microbiol **64**:3313-3319.

12. **Blears, M. J., S.A. De Grandis, H. Lee, and J.T. Trevors**. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology and Biotechnolgy**21**:99-114.

13. **Brade, H., L. Brade, and E. T. Rietschel.** 1988. Structure-activity relationships of bacterial lipopolysaccharides (endotoxins). Current and future aspects. Zentralbl Bakteriol Mikrobiol Hyg [A] **268**:151-179.

14. **Kneifel, H., K.O. Stetter, J.R. Andreesen, J. Wiegel, H. Konig, and S.M. Schoberth.** 1986. Distribution of polyamines in representative species of archaebacteria. Sys Appl Microbiol **7**:241-245

15. **Burke, S. A., J. D. Wright, M. K. Robinson, B. V. Bronk, and R. L. Warren.** 2004. Detection of molecular diversity in *Bacillus atrophaeus* by amplified fragment length polymorphism analysis. Appl Environ Microbiol **70**:2786-2790.

16. **Busse, H. J., E. B. Denner, and W. Lubitz.** 1996. Classification and identification of bacteria: current approaches to an old problem. Overview of methods used in bacterial systematics. J Biotechnol **47**:3-38.

17. **Busse, H. J., A. Kainz, I. V. Tsitko, and M. Salkinoja-Salonen.** 2000. Riboprints as a tool for rapid preliminary identification of sphingomonads. Syst Appl Microbiol **23**:115-123.

18. **Cartmill, T. D., K. Orr, R. Freeman, P. R. Sisson, and N. F. Lightfoot.** 1992. Nosocomial infection with Clostridium difficile investigated by pyrolysis mass spectrometry. J Med Microbiol **37**:352-356.

19. **Chen, K. Y. and S. Cheng.** 1988. Polyamine metabolism in an obligately alkalophilic *Bacillus alcalophilus* that grows at pH 11.0. Biochem Biophys Res Commun **150**:185-191.

20. **Cherif, A., L. Brusetti, S. Borin, A. Rizzi, A. Boudabous, H. Khyami-Horani, and D. Daffonchio.** 2003. Genetic relationship in the *'Bacillus cereus* group' by rep-PCR fingerprinting and sequencing of a *Bacillus anthracis*-specific rep-PCR fragment. J Appl Microbiol **94**:1108-1119.

21. **Claus, D., and R. C. W. Berkeley.** *Genus Bacillus* Cohn 1872, 174 AL, p. 1105-1139. Baltimore: Wiliams & Wilkins, Bergey's Manual of systematic bacteriology.

22. **Coenye, T., E. Falsen, M. Vancanneyt, B. Hoste, J. R. Govan, K. Kersters, and P. Vandamme.** 1999. Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. Int J Syst Bacteriol **49**:405-413.

23. **Coenye, T., L. M. Schouls, J. R. Govan, K. Kersters, and P. Vandamme.** 1999. Identification of *Burkholderia* species and genomovars from cystic fibrosis patients by AFLP fingerprinting. Int J Syst Bacteriol **49**:1657-1666.

24. **Cohen, S.** 1998. A Guide to the Polyamines. Oxford University Press, Oxford.

25. **Collins, M. D. and D. Jones.** 1981. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. Microbiol Rev **45**:316-354.

26. **Collins, M. D., D. Jones, M. Goodfellow, and D. E. Minnikin.** 1979. Isoprenoid quinone composition as a guide to the classification of *Listeria*, *Brochothrix*, *Erysipelothrix* and *Caryophanon*. J Gen Microbiol **111**:453-457.

27. **Combet-Blanc, Y., B. Ollivier, C. Streicher, B. K. Patel, P. P. Dwivedi, B. Pot, G. Prensier, and J. L. Garcia.** 1995. *Bacillus thermoamylovorans* sp. nov., a moderately thermophilic and amylolytic bacterium. Int J Syst Bacteriol **45**:9-16.

28. **Costas, M., B. Pot, P. Vandamme, K. Kersters, R. J. Owen, and L. R. Hill.** 1990. Interlaboratory comparative study of the numerical analysis of one-dimensional sodium dodecyl sulphate-polyacrylamide gel electrophoretic protein patterns of *Campylobacter* strains. Electrophoresis **11**:467-474.

29. **Coyette, J. and J. M. Ghuysen.** 1970. Structure of the walls of *Lactobacillus acidophilus* strain 63 AM Gasser. Biochemistry **9**:2935-2943.

30. **Curk, M.C., F. Peladan, and J.C. Hubert.** 1994. Fourier-Transform Infrared (FTIR) Spectroscopy for identifying *Lactobacillus* species. FEMS Microiology Letters **123**:241-248.

31. **da Silva, K. R., L. Rabinovitch, and L. Seldin.** 1999. Phenotypic and genetic diversity among *Bacillus sphaericus* strains isolated in Brazil, potentially useful as biological control agents against mosquito larvae. Res Microbiol **150**:153-160.

32. **Busse, J., and G. Auling.** 1988. Polyamine pattern as a chemotaxonomic marker within the *Proteobacteria*. Sys Appl Microbiol **11**:1-8

33. **Dautle, M. P., R. L. Ulrich, and T. A. Hughes.** 2002. Typing and subtyping of 83 clinical isolates purified from surgically implanted silicone feeding tubes by random amplified polymorphic DNA amplification. J Clin Microbiol **40**:414-421.

34. **De Clerck, E. and P. De Vos.** 2002. Study of the bacterial load in a gelatine production process focussed on *Bacillus* and related endosporeforming genera. Syst Appl Microbiol **25**:611-617.

35. **De Clerck, E., M. Rodriguez-Diaz, G. Forsyth, L. Lebbe, N. A. Logan, and P. DeVos.** 2004. Polyphasic characterization of *Bacillus coagulans* strains, illustrating heterogeneity within this species, and emended description of the species. Syst Appl Microbiol **27**:50-60.

36. **De Clerck, E., M. Rodriguez-Diaz, T. Vanhoutte, J. Heyrman, N. A. Logan, and P. De Vos.** 2004. *Anoxybacillus contaminans* sp. nov. and *Bacillus gelatini* sp. nov., isolated from contaminated gelatin batches. Int J Syst Evol Microbiol **54**:941-946.

37. **De Clerck, E., T. Vanhoutte, T. Hebb, J. Geerinck, J. Devos, and P. De Vos.** 2004. Isolation, characterization, and identification of bacterial contaminants in semifinal gelatin extracts. Appl Environ Microbiol **70**:3664-3672.

38. **De Vos, P.** 2002. Nucleic Acid Analysis and SDS-Page of Whole-cell Proteins in *Bacillus* Taxonomy, p. 141-159. *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

39. **Delcour, J., T. Ferain, M. Deghorain, E. Palumbo, and P. Hols.** 1999. The biosynthesis and functionality of the cell-wall of lactic acid bacteria. Antonie Van Leeuwenhoek **76**:159-184.

40. **Denner, E. B., D. Vybiral, U. R. Fischer, B. Velimirov, and H. J. Busse.** 2002. *Vibrio calviensis* sp. nov., a halophilic, facultatively oligotrophic 0.2 microm-fiIterabIe marine bacterium. Int J Syst Evol Microbiol **52**:549-553.

41. **Dickinson, D. N., M. T. La Duc, M. Satomi, J. D. Winefordner, D. H. Powell, and K. Venkateswaran.** 2004. MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of *Bacillus pumilus* spores. J Microbiol Methods **58**:1-12.

42. **Dieckmann, R., I. Graeber, I. Kaesler, U. Szewzyk, and H. von Dohren.** 2005. Rapid screening and dereplication of bacterial isolates from marine sponges of the Sula Ridge by Intact-Cell-MALDI-TOF mass spectrometry (ICM-MS). Appl Microbiol Biotechnol **67**:539-548.

43. **Ehling-Schulz, M., B. Svensson, M. H. Guinebretiere, T. Lindback, M. Andersson, A. Schulz, M. Fricker, A. Christiansson, P. E. Granum, E. Martlbauer, C. Nguyen-The, M. Salkinoja-Salonen, and S. Scherer.** 2005. Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. Microbiology **151**:183-197.

44. **Ehret, W. and G. Ruckdeschel.** 1985. Membrane proteins of legionellaceae. I. Membrane proteins of different strains and serogroups of *Legionella pneumophila*. Zentralbl Bakteriol Mikrobiol Hyg [A] **259**:433-445.

45. **Felske, A., A. Wolterink, R. van Lis, W. M. de Vos, and A. D. Akkermans.** 1999. Searching for predominant soil bacteria: 16S rDNA cloning versus strain cultivation. FEMS Microbiol Ecol **30**:137-145.

46. **Fischer, W., H. U. Koch, and R. Haas.** 1983. Improved preparation of lipoteichoic acids. Eur J Biochem **133**:523-530.

47. **Fischer, W., H. U. Koch, P. Rosel, and F. Fiedler.** 1980. Alanine ester-containing native lipoteichoic acids do not act as lipoteichoic acid carrier. Isolation, structural and functional characterization. J Biol Chem **255**:4557-4562.

48. **Fischer, W., P. Rosel, and H. U. Koch.** 1981. Effect of alanine ester substitution and other structural features of lipoteichoic acids on their inhibitory activity against autolysins of *Staphylococcus aureus*. J Bacteriol **146**:467-475.

49. **Fox, A.** 1999. Carbohydrate profiling of bacteria by gas chromatography-mass spectrometry and their trace detection in complex matrices by gas chromatography-tandem mass spectrometry. J Chromatogr A **843**:287-300.

50. **Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk Jr.** 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol **42**:166-170.

51. **Fritze, D.** 2002. *Bacillus* identification - Traditional approaches, p. 100-122. Blackwell Publishing, *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

52. **Gaviria Rivera, A. M. and F. G. Priest.** 2003. Molecular typing of *Bacillus thuringiensis* serovars by RAPD-PCR. Syst Appl Microbiol **26**:254-261.

53. **Gevers, D., G. Huys, and J. Swings.** 2001. Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. FEMS Microbiol Lett **205**:31-36.

54. **Gil-Lamaignere, C., E. Roilides, J. Hacker, and F. M. Muller.** 2003. Molecular typing for fungi--a critical review of the possibilities and limitations of currently and future methods. Clin Microbiol Infect **9**:172-185.

55. **Giraffa, G., P. De Vecchi, P. Rossi, G. Nicastro, and M. G. Fortina.** 1998. Genotypic heterogeneity among *Lactobacillus helveticus* strains isolated from natural cheese starters. J Appl Microbiol **85**:411-416.

57. **Goodacre, R., E. M. Timmins, P. J. Rooney, J. J. Rowland, and D. B. Kell.** 1996. Rapid identification of *Streptococcus* and *Enterococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks. FEMS Microbiol Lett **140**:233-239.

58. **Goto, K., T. Omura, Y. Hara, and Y. Sadaie.** 2000. Application of the partial 16S rDNA sequence as an index for rapid identification of species in the genus *Bacillus*. J Gen Appl Microbiol **46**:1-8.

59. **Guillaume-Gentil, O., P. Scheldeman, J. Marugg, L. Herman, H. Joosten, and M. Heyndrickx.** 2002. Genetic heterogeneity in *Bacillus sporothermodurans* as demonstrated by ribotyping and repetitive extragenic palindromic-PCR fingerprinting. Appl Environ Microbiol **68**:4216-4224.

60. **Haag, H., H.-U. Gremlich, R. Bergmann, and J.-J. Sanglier.** 1996. Characterization and identification of actinomycetes by FT-IR spectroscopy. J. Mcrobiol. Methods **27**:157-163.

61. **Hamana, K., H. Hamana, M. Niitsu, K. Samejima, T. Sakane, and A. Yokota.** 1993. Tertiary and quaternary branched polyamines distributed in thermophilic *Saccharococcus* and *Bacillus*. Microbios **75**:23-32.

62. **Hazem, A. and A. Manar.** 2003. Genetic polymorphism by RAPD-PCR and phenotypic characteristics of isolated thermotolerant *Bacillus* strains from hot spring sources. New Microbiol **26**:249-256.

63. **Helgason, E., N. J. Tourasse, R. Meisal, D. A. Caugant, and A. B. Kolsto.** 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl Environ Microbiol **70**:191-201.

64. **Helm, D., H. Labischinski, G. Schallehn, and D. Naumann.** 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. J Gen Microbiol **137**:69-79.

65. **Herman, L. and M. Heyndrickx.** 2000. The presence of intragenically located REP-like elements in *Bacillus sporothermodurans* is sufficient for REP-PCR typing. Res Microbiol **151**:255-261.

66. **Heyrman, J., N. A. Logan, M. Rodriguez-Diaz, P. Scheldeman, L. Lebbe, J. Swings, M. Heyndrickx, and P. De Vos.** 2005. Study of mural painting isolates, leading to the transfer of *'Bacillus maroccanus*' and *'Bacillus carotarum*' to *Bacillus simplex*, emended description of *Bacillus simplex*, re-examination of the strains previously attributed to *'Bacillus macroides*' and description of *Bacillus muralis* sp. nov. Int J Syst Evol Microbiol **55**:119-131.

67. **Hill, K. K., L. O. Ticknor, R. T. Okinaka, M. Asay, H. Blair, K. A. Bliss, M. Laker, P. E. Pardington, A. P. Richardson, M. Tonks, D. J. Beecher, J. D. Kemp, A. B. Kolsto, A. C. Wong, P. Keim, and P. J. Jackson.** 2004. Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl Environ Microbiol **70**:1068-1080.

68. **Huys, G., R. Coopman, P. Janssen, and K. Kersters.** 1996. High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. Int J Syst Bacteriol **46**:572-580.

69. **Ibelings, M. S., K. Maquelin, H. P. Endtz, H. A. Bruining, and G. J. Puppels.** 2005. Rapid identification of *Candida* spp. in peritonitis patients by Raman spectroscopy. Clin Microbiol Infect **11**:353-358.

70. **Jackson, P. J., K. K. Hill, M. T. Laker, L. O. Ticknor, and P. Keim.** 1999. Genetic comparison of *Bacillus anthracis* and its close relatives using amplified fragment length polymorphism and polymerase chain reaction analysis. J Appl Microbiol **87**:263-269.

71. **Jann, B., K. Reske, and K. Jann.** 1975. Heterogeneity of lipopolysaccharides. Analysis of polysaccharide chain lengths by sodium dodecylsulfate-polyacrylamide gel electrophoresis. Eur J Biochem **60**:239-246.

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

73. **Janssen, P., K. Maquelin, R. Coopman, I. Tjernberg, P. Bouvet, K. Kersters, and L. Dijkshoorn.** 1997. Discrimination of *Acinetobacter* genomic species by AFLP fingerprinting. Int J Syst Bacteriol **47**:1179-1187.

74. **Johnston, N. C., H. Goldfine, and W. Fischer.** 1994. Novel polar lipid composition of *Clostridium innocuum* as the basis for an assessment of its taxonomic status. Microbiology **140**:105-111.

75. **Kampfer, P.** 1994. Limits and possibilities of total fatty acid analysis for classification and identification of *Bacillus* species. Systematic and Applied Microbiology **17**:86-98.

76. **Kampfer, P.** 2002. Whole-cell Fatty Acid Analysis in the Systematics of *Bacillus* and related Genera, p. 271-299. Blackwell Publishing, *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

77. **Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P. Jackson.** 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J Bacteriol **179**:818-824.

78. Todar, K. 2005. http://textbookofbacteriology.net/Anthrax.html.

79. Todar, K.. The Genus *Bacillus*. 2005. http://textbookofbacteriology.net/Bacillus.html.

80. **Kirschner, C., K. Maquelin, P. Pina, N. A. Ngo Thi, L. P. Choo-Smith, G. D. Sockalingum, C. Sandt, D. Ami, F. Orsini, S. M. Doglia, P. Allouch, M. Mainfait, G. J. Puppels, and D. Naumann.** 2001. Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J Clin Microbiol **39**:1763-1770.

81. **Knox, K. W. and A. J. Wicken.** 1973. Immunological properties of teichoic acids. Bacteriol Rev **37**:215-257.

82. **Koch, A. L.** 1990. Growth and form of the bacterial cell wall. Am. Scientist 327-341.

83. **Koort, J., P. Vandamme, U. Schillinger, W. Holzapfel, and J. Bjorkroth.** 2004. *Lactobacillus curvatus* subsp. *melibiosus* is a later synonym of *Lactobacillus sakei* subsp. *carnosus*. Int J Syst Evol Microbiol **54**:1621-1626.

84. **Lee, J. S., Y. K. Shin, J. H. Yoon, M. Takeuchi, Y. R. Pyun, and Y. H. Park.** 2001. *Sphingomonas aquatilis* sp. nov., *Sphingomonas koreensis* sp. nov., and *Sphingomonas taejonensis* sp. nov., yellow-pigmented bacteria isolated from natural mineral water. Int J Syst Evol Microbiol **51**:1491-1498.

85. **Logan, N. A.** 2002. Modern methods for identification, p. 123-140. Blackwell Publishing, *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

86. **Logan, N. A., G. Forsyth, L. Lebbe, J. Goris, M. Heyndrickx, A. Balcaen, A. Verhelst, E. Falsen, A. Ljungh, H. B. Hansson, and P. De Vos.** 2002. Polyphasic identification of *Bacillus* and *Brevibacillus* strains from clinical, dairy and industrial specimens and proposal of *Brevibacillus* invocatus sp. nov. Int J Syst Evol Microbiol **52**:953-966.

87. **Logan, N. A., L. Lebbe, B. Hoste, J. Goris, G. Forsyth, M. Heyndrickx, B. L. Murray, N. Syme, D. D. Wynn-Williams, and P. De Vos.** 2000. Aerobic endospore-forming bacteria from geothermal environments in northern Victoria Land, Antarctica, and Candlemas Island, South Sandwich archipelago, with the proposal of *Bacillus fumarioli* sp. nov. Int J Syst Evol Microbiol **50**:1741-1753.

88. **Long, D. A.** 1977. Raman spectroscopy. McGraw-Hill New York, NY.

89. **Madigan, M. T., J. M. Martinko, and. J. Parker.** 2000. Brock Biology of Microorganisms, p. 29-47. Prentice Hall International. Ninth Edition.

90. **Magee, J. T., and R. Goodacre.** 2002. Fingerprint Spectrometry Methods in *Bacillus* Systematics, p. 255-270. *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

91. **Mangin, I., N. Bourget, and B. Decaris.** 1996. Ribosomal DNA polymorphism in the genus *Bifidobacterium*. Res Microbiol **147**:183-192.

92. **Maquelin, K.** 2002. Phd Scription: "Confocal Raman Microspectroscopy: a novel diagnostic tool in medical microbiology".

93. **Maquelin, K., L. P. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H. A. Bruining, and G. J. Puppels.** 2000. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. Anal Chem **72**:12-19.

94. **Maquelin, K., L. Dijkshoorn, T. J. van der Reijden, and G. J. Puppels.** 2005. Rapid epidemiological analysis of *Acinetobacter* strains by Raman spectroscopy. J Microbiol Methods, *in press*

95. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. A. Ngo-Thi, T. van Vreeswijk, M. Stammler, H. P. Endtz, H. A. Bruining, D. Naumann, and G. J. Puppels.** 2003. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. J Clin Microbiol **41**:324-329.

96. **Masco, L., G. Huys, D. Gevers, L. Verbrugghen, and J. Swings.** 2003. Identification of *Bifidobacterium* species using rep-PCR fingerprinting. Syst Appl Microbiol **26**:557-563.

97. **Maszenan, A. M., R. J. Seviour, B. K. Patel, P. Schumann, J. Burghardt, R. I. Webb, J. A. Soddell, and G. N. Rees.** 1999. *Friedmanniella spumicola* sp. nov. and *Friedmanniella capsulata* sp. nov. from activated sludge foam: gram-positive cocci that grow in aggregates of repeating groups of cocci. Int J Syst Bacteriol **49**:1667-1680.

98. **Mesbah, M., U. Premachandran, and W.B. Whitman.** 1989. Precise measurement of the G + C content of deoxyribonucleic acid by high performance liquid chromatography. International Journal of Systematic Bacteriology **39**:159-167.

99. **Moore, L. V., D. M. Bourne, and W. E. Moore.** 1994. Comparative distribution and taxonomic value of cellular fatty acids in thirty-three genera of anaerobic gram-negative bacilli. Int J Syst Bacteriol **44**:338-347.

100. **Moran, A. P., I. M. Helander, and T. U. Kosunen.** 1992. Compositional analysis of *Helicobacter pylori* rough-form lipopolysaccharides. J Bacteriol **174**:1370-1377.

101. **Morris, J. A. and R. W. Park.** 1973. A comparison using gel electrophoresis of cell proteins of campylobacters (vibrios) associated with infertility, abortion and swine dysentery. J Gen Microbiol **78**:165-178.

102. **Moschetti, G., G. Blaiotta, M. Aponte, G. Mauriello, F. Villani, and S. Coppola.** 1997. Genotyping of *Lactobacillus delbrueckii* subsp. *bulgaricus* and determination of the number and forms of rrn operons in *L. delbrueckii* and its subspecies. Res Microbiol **148**:501-510.

103. **Schleifer, K.H.** 1985. Analysis of the chemical composition and primary structure of mureine Methods Microbiol **18**: 123-156.

104. **Naumann, D., D. Helm, and H. Labischinski.** 1991. Microbiological characterizations by FT-IR spectroscopy. Nature **351**:81-82.

105. **Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev, and M. V. Ivanov.** 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, G. th. Int J Syst Evol Microbiol **51**:433-446.

106. **Nielsen, P., D. Fritze, and F. G. Priest.** 1995. Phenetic diversity of alkalophilic *Bacillus* strains - Proposal for 9 new species. Microbiolgy-UK **141**: 1745-1761.

107. **Nielsen, P., F. A. Rainey, H. Outtrup, F.G. Priest, and D. Fritze.** 1994. Comparative 16S rDNA sequence analysis of some alkaliphilic bacilli and the establishment of a sixth rRNAgroup whithin the genus *Bacillus*. FEMS Microbiology letters **117**:61-66.

108. **Noguchi, H., M. Uchino, O. Shida, K. Takano, L. K. Nakamura, and K. Komagata.** 2004. *Bacillus vietnamensis* sp. nov., a moderately halotolerant, aerobic, endospore-forming bacterium isolated from Vietnamese fish sauce. Int J Syst Evol Microbiol **54**:2117-2120.

109. **Nohynek, L. J., E. L. Nurmiaho-Lassila, E. L. Suhonen, H. J. Busse, M. Mohammadi, J. Hantula, F. Rainey, and M. S. Salkinoja-Salonen.** 1996. Description of chlorophenol-degrading *Pseudomonas* sp. strains KF1T, KF3, and NKF1 as a new species of the genus *Sphingomonas*, *Sphingomonas subarctica* sp. nov. Int J Syst Bacteriol **46**:1042-1055.

110. **Oberreuter, H., J. Charzinski, and S. Scherer.** 2002. Intraspecific diversity of *Brevibacterium linens, Corynebacterium glutamicum* and *Rhodococcus erythropolis* based on partial 16S rDNA sequence analysis and Fourier-transform infrared (FT-IR) spectroscopy. Microbiology **148**:1523-1532.

111. **Oberreuter, H., H. Seiler, and S. Scherer.** 2002. Identification of coryneform bacteria and related taxa by Fourier-transform infrared (FT-IR) spectroscopy. Int J Syst Evol Microbiol **52**:91-100.

112. **Olive, D. M. and P. Bean.** 1999. Principles and applications of methods for DNA-based typing of microbial organisms. J Clin Microbiol **37**:1661-1669.

113. **On, S. L.** 1996. Identification methods for *Campylobacters*, *Helicobacters*, and related organisms. Clin Microbiol Rev **9**:405-422.

114. **Osteras, M., J. Stanley, and T. M. Finan.** 1995. Identification of *Rhizobium*-specific intergenic mosaic elements within an essential two-component regulatory system of *Rhizobium* species. J Bacteriol **177**:5485-5494.

115. **Otto, C., N. M. Sijtsema, and J. Greve.** 1998. Confocal Raman microspectroscopy of the activation of single neutrophilic granulocytes. Eur Biophys J **27**:582-589.

116. **Oust, A., T. Moretro, C. Kirschner, J. A. Narvhus, and A. Kohler.** 2004. FT-IR spectroscopy for identification of closely related lactobacilli. J Microbiol Methods **59**:149-162.

117. **Owen, R. J. and P. J. Jackman.** 1982. The similarities between *Pseudomonas paucimobilis* and allied bacteria derived from analysis of deoxyribonucleic acids and electrophoretic protein patterns. J Gen Microbiol **128**:2945-2954.

118. **Pajvani, U. B. and P. E. Scherer.** 2003. Adiponectin: systemic contributor to insulin sensitivity. Curr Diab Rep **3**:207-213.

119. **Pelletier, M. J.** 1999. Introduction to Applied Raman Spectroscopy, p. 1-52. Blackwell Science, Analytical Applications of Raman Spectroscopy.

120. **Placzek, G.** 1934. Rayleigh-Streuung und Raman-Effect. p. 205-374. Handbuch der Radiologie.

121. **Pot, B., P. Vandamme, and K. Kersters.** 1994. Analysis of electrophoretic whole-organism protein fingerprints, p. 493-521. John Wiley & Sons, Ltd., Chichester, England, Modern Microbial methods. Chemical methods in prokaryotic systematics.

122. **Puppels, G. J., F. F. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D. J. Arndt-Jovin, and T. M. Jovin.** 1990. Studying single living cells and chromosomes by confocal Raman microspectroscopy. Nature **347**:301-303.

123. **Puppels, G. J., H. S. Garritsen, G. M. Segers-Nolten, F. F. de Mul, and J. Greve.** 1991. Raman microspectroscopic approach to the study of human granulocytes. Biophys J **60**:1046-1056.

124. **Puppels, G. J., W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. de Mul, and J. Greve.** 1991. Description and performance of a highly sensitive confocal Raman spectrometer. J. Raman Spectros. **22**: 217-225.

125. **Rademaker, J. L., B. Hoste, F. J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, and F. J. de Bruijn.** 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. Int J Syst Evol Microbiol **50**:665-677.

126. **Rajala, M. W. and P. E. Scherer.** 2003. Minireview: The adipocyte--at the crossroads of energy homeostasis, inflammation, and atherosclerosis. Endocrinology **144**:3765-3773.

127. **Raman, C. V., and K. S. Krishnan.** 1928. A new type of secundary radiation. Nature **121**:501-502.

128. **Randles, J. W., G. Steger, and D. Riesner.** 1982. Structural transitions in viroid-like RNAs associated with cadang-cadang disease, velvet tobacco mottle virus, and *Solanum nodiflorum* mottle virus. Nucleic Acids Res **10**:5569-5586.

129. **Ripabelli, G., J. McLauchlin, V. Mithani, and E. J. Threlfall.** 2000. Epidemiological typing of *Bacillus cereus* by amplified fragment length polymorphism. Lett Appl Microbiol **30**:358-363.

130. **Romano, I., L. Lama, B. Nicolaus, A. Gambacorta, and A. Giordano.** 2005. *Bacillus saliphilus* sp. nov., isolated from a mineral pool in Campania, Italy. Int J Syst Evol Microbiol **55**:159-163.

131. **Ronimus, R. S., L. E. Parker, N. Turner, S. Poudel, A. Ruckert, and H. W. Morgan.** 2003. A RAPD-based comparison of thermophilic bacilli from milk powders. Int J Food Microbiol **85**:45-61.

132. **Ruiz-Garcia, C., E. Quesada, F. Martinez-Checa, I. Llamas, M. C. Urdaci, and V. Bejar.** 2005. *Bacillus axarquiensis* sp. nov. and *Bacillus malacitensis* sp. nov., isolated from river-mouth sediments in southern Spain. Int J Syst Evol Microbiol **55**:1279-1285.

133. **Salkinoja-Salonen, M. S., R. Vuorio, M. A. Andersson, P. Kampfer, M. C. Andersson, T. Honkanen-Buzalski, and A. C. Scoging.** 1999. Toxigenic strains of *Bacillus licheniformis* related to food poisoning. Appl Environ Microbiol **65**:4637-4645.

134. **Sanglier, J. J., D. Whitehead, G. S. Saddler, E. V. Ferguson, and M. Goodfellow.** 1992. Pyrolysis mass spectrometry as a method for the classification, identification and selection of actinomycetes. Gene **115**:235-242.

135. **Satokari, R. M., E. E. Vaughan, H. Smidt, M. Saarela, J. Matto, and W. M. de Vos.** 2003. Molecular approaches for the detection and identification of bifidobacteria and lactobacilli in the human gastrointestinal tract. Syst Appl Microbiol **26**:572-584.

136. **Scherer, P. and H. Kneifel.** 1983. Distribution of polyamines in methanogenic bacteria. J Bacteriol **154**:1315-1322.

137. **Scherer, P. G. and J. Seelig.** 1987. Structure and dynamics of the phosphatidylcholine and the phosphatidylethanolamine head group in L-M fibroblasts as studied by deuterium nuclear magnetic resonance. EMBO J **6**:2915-2922.

138. **Scherer, P. W., I. I. Hahn, and M. M. Mozell.** 1989. The biophysics of nasal airflow. Otolaryngol Clin North Am **22**:265-278.

139. **Schleifer, K. H. and O. Kandler.** 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev **36**:407-477.

140. **Seelig, J., P. M. Macdonald, and P. G. Scherer.** 1987. Phospholipid head groups as sensors of electric charge in membranes. Biochemistry **26**:7535-7541.

141. **Segers, P., M. Vancanneyt, B. Pot, U. Torck, B. Hoste, D. Dewettinck, E. Falsen, K. Kersters, and P. De Vos.** 1994. Classification of *Pseudomonas diminuta* Leifson and Hugh 1954 and *Pseudomonas vesicularis* Busing, Doll, and Freytag 1953 in *Brevundimonas* gen. nov. as *Brevundimonas diminuta* comb. nov. and *Brevundimonas vesicularis* comb. nov., respectively. Int J Syst Bacteriol **44**:499-510.

142. **Serban, D., J.M. Benevides, and G. J. Thomas**. 2003. HU protein employs similar mechanisms of minor-groove recognition in binding to different B-DNA sites: Demonstration by Raman spectroscopy. Biochemistry **42**:7390-7399.

143. **Shida, O., H. Takagi, K. Kadowaki, and K. Komagata.** 1996. Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. Int J Syst Bacteriol **46**:939-946.

144. **Shute, L. A., C. S. Gutteridge, J. R. Norris, and R. C. Berkeley.** 1984. Curie-point pyrolysis mass spectrometry applied to characterization and identification of selected *Bacillus* species. J Gen Microbiol **130**:343-355.

145. **Shute, L. A., C. S. Gutteridge, J. R. Norris, and R. C. Berkeley.** 1988. Reproducibility of pyrolysis mass spectrometry: effect of growth medium and instrument stability on the differentiation of selected *Bacillus* species. J Appl Bacteriol **64**:79-88.

146. **Sisson, P. R., J. M. Kramer, M. M. Brett, R. Freeman, R. J. Gilbert, and N. F. Lightfoot.** 1992. Application of pyrolysis mass spectrometry to the investigation of outbreaks of food poisoning and non-gastrointestinal infection associated with *Bacillus* species and *Clostridium* perfringens. Int J Food Microbiol **17**:57-66.

147. **Smekal, A.** 1923. Zur Quantentheorie der Dispersion. Naturwiss. **11**:873-875.

148. **Sotgia, F., S. E. Woodman, G. Bonuccelli, F. Capozza, C. Minetti, P. E. Scherer, and M. P. Lisanti.** 2003. Phenotypic behavior of caveolin-3 R26Q, a mutant associated with hyperCKemia, distal myopathy, and rippling muscle disease. Am J Physiol Cell Physiol **285**: 1150-60.

149. **Spring S., Ludwig W. Marquez M. C. Ventosa A. Schleifer K. H.** 1996. *Halobacillus* gen. nov., with descriptions of *Halobacillus litoralis* sp. nov. and *Halobacillus trueperi* sp. nov. and transfer of *Sporosarcina halophila* to *Halobacillus halophilus* comb. nov. Int J Syst Bacteriol, **46**: 492-496.

150. **Stackebrandt, E., and W. Liesack.**1993. Nucleic acids and classification, p. 151-194. London Academic Press, In: Modern Approaches in Bacterial Systematics (M. Goodfellow and A. O'Donell, eds.)

151. **Stackebrandt, E., and J. Swiderski.** 2002. From Phylogeny to Systematics, p. 8-22. Blackwell Publishing, *In*: Applications and Systematics of *Bacillus* and Relatives. (R. Berkeley, M. Heyndrickx, N. Logan and P. De Vos, eds.).

152. **Suihko, M. L. and E. Stackebrandt.** 2003. Identification of aerobic mesophilic bacilli isolated from board and paper products containing recycled fibres. J Appl Microbiol **94**:25-34.

153. **Suzuki, K., M. Goodfellow and A. G. O'Donnell.** 1993. Cell enveloppes and classification, p. 195-250. Academic Press., Ltd., London.

154. **Tamura, T., Y. Nakagaito, T. Nishii, T. Hasegawa, E. Stackebrandt, and A. Yokota.** 1994. A new genus of the order *Actinomycetales*, *Couchioplanes* gen. nov., with descriptions of *Couchioplanes caeruleus* (Horan and Brodsky 1986) comb. nov. and *Couchioplanes caeruleus* subsp. azureus subsp. nov. Int J Syst Bacteriol **44**:193-203.

155. **Tamura, T., M. Takeuchi, and A. Yokota.** 1994. *Luteococcus japonicus* gen. nov., sp. nov., a new gram-positive coccus with LL-diaminopimelic acid in the cell wall. Int J Syst Bacteriol **44**:348-356.

156. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **33**:2233-2239.

157. **Ticknor, L. O., A. B. Kolsto, K. K. Hill, P. Keim, M. T. Laker, M. Tonks, and P. J. Jackson.** 2001. Fluorescent Amplified Fragment Length Polymorphism Analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* Soil Isolates. Appl Environ Microbiol **67**:4863-4873.

158. **Tsuboi, M., J. M. Benevides, P. Bondre, and G. J. Thomas Jr.** 2005. Structural Details of the Thermophilic Filamentous Bacteriophage PH75 Determined by Polarized Raman Microspectroscopy. Biochemistry **44**:4861-4869.

159. **Turnbull, P. C. B., and J. M. Kramer.** 1991. *Bacillus*, p. 188-210. American Society for Microbiology, Washington, D.C., In: Manual of clinical Microbiology, (P.R. Murray, ed.).

160. **Tynkkynen, S., R. Satokari, M. Saarela, T. Mattila-Sandholm, and M. Saxelin.** 1999. Comparison of ribotyping, randomly amplified polymorphic DNA analysis, and pulsed-field gel electrophoresis in typing of Lactobacillus rhamnosus and L. casei strains. Appl Environ Microbiol **65**:3908-3914.

161. **Uzunbajakava, N., A. Lenferink, Y. Kraan, B. Willekens, G. Vrensen, J. Greve, and C. Otto.** 2003. Nonresonant Raman imaging of protein distribution in single human cells. Biopolymers **72**:1-9.

162. **Van Der Zwet, W. C., G. A. Parlevliet, P. H. Savelkoul, J. Stoof, A. M. Kaiser, A. M. Van Furth, and C. M. Vandenbroucke-Grauls.** 2000. Outbreak of *Bacillus cereus* infections in a neonatal intensive care unit traced to balloons used in manual ventilation. J Clin Microbiol **38**:4131-4136.

163. **Vandamme, P., B. Pot, M. Gillis, P. de Vos, K. Kersters, and J. Swings.** 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev **60**:407-438.

164. **Vauterin Luc, R. J. S. J.** 2000. Synopsis on the Taxonomy of the Genus *Xanthomonas*. Phytopatholgy **90**:677-682.

165. **Versalovic, J., M. Schneider, F.J. de Bruijn, and J. R. Lupski.** 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol Cell Biol **5**:25-40.

166. **Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and a. l. et.** 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res **23**:4407-4414.

167. **Wayne, L. G., D.J. Brenner, R.R. Colwell, P.A.D. Grimont, O. Kandler, M.I.Krichevsky, L.H. Moore, W.E.C. Moore, R.G.E. Murray, E. Stackebrandt, M.P. Starr, and H.G. Truper.** 1987. Report of the ad hoc comittee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol **37**:463-464.

168. **Whitfield, C.** 1988. Bacterial extracellular polysaccharides. Can J Microbiol **34**:415-420.

169. **Wiegel, J.** 1981. Distinction between the Gram reaction and the Gram type of bacteria. Int J Syst Bacteriol **31:** 88-88

170. **Wilkinson, S. G. and M. E. Bell.** 1971. The phosphoglucolipid from *Pseudomonas* diminuta. Biochim Biophys Acta **248**:293-299.

171. **Wilkinson, S. G. and L. Galbraith.** 1979. Polar lipids of *Pseudomonas vesicularis*. Presence of a heptosyldiacylglycerol. Biochim Biophys Acta **575**:244-54.

172. **Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res **18**:6531-6535.

173. **Winder, C. L., E. Carr, R. Goodacre, and R. Seviour.** 2004. The rapid identification of *Acinetobacter* species using Fourier transform infrared spectroscopy. J Appl Microbiol **96**:328-339.

174. **Wisotzkey, J. D., P. Jurtshuk Jr, G. E. Fox, G. Deinhard, and K. Poralla.** 1992. Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. Int J Syst Bacteriol **42**:263-269.

175. **Woese, C. R.** 1987. Bacterial evolution. Microbiol Rev **51**:221-271.

176. **Wood, R. W.** 1928. Wavelength shifts in scattered light. Nature **122**:349.

177. **Xu, D. and J. C. Cote.** 2003. Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. Int J Syst Evol Microbiol **53**:695-704.

178. **Yamamoto, S., M. A. Chowdhury, M. Kuroda, T. Nakano, Y. Koumoto, and S. Shinoda.** 1991. Further study on polyamine compositions in *Vibrionaceae*. Can J Microbiol **37**:148-153.

179. **Yang, S. and R. E. Rothman.** 2004. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. Lancet Infect Dis **4**:337-348.

180. **Yasuzawa, K., N. Hayashi, N. Goshima, K. Kohno, F. Imamoto, and Y. Kano.** 1992. Histone-like proteins are required for cell growth and constraint of supercoils in DNA. Gene **122**:9-15.

181. **Yokota, A., M. Takeuchi, T. Sakane, and N. Weiss.** 1993. Proposal of six new species in the genus *Aureobacterium* and transfer of *Flavobacterium esteraromaticum* Omelianski to the genus *Aureobacterium* as *Aureobacterium esteraromaticum* comb. nov. Int J Syst Bacteriol **43**:555-564.

182. **Yoon, J. H., K. C. Lee, N. Weiss, Y. H. Kho, K. H. Kang, and Y. H. Park.** 2001. *Sporosarcina aquimarina* sp. nov., a bacterium isolated from seawater in Korea, and transfer of *Bacillus globisporus* (Larkin and Stokes 1967), *Bacillus psychrophilus* (Nakamura 1984) and *Bacillus pasteurii* (Chester 1898) to the genus *Sporosarcina* as *Sporosarcina globispora* comb. nov., *Sporosarcina psychrophila* comb. nov. and *Sporosarcina pasteurii* comb. nov., and emended description of the genus *Sporosarcina*. Int J Syst Evol Microbiol **51**:1079-1086.

183. **Yumoto, I., K. Yamazaki, T. Sawabe, K. Nakano, K. Kawasaki, Y. Ezura, and H. Shinano.** 1998. *Bacillus horti* sp. nov., a new gram-negative alkaliphilic bacillus. Int J Syst Bacteriol **48:**565-571.

Evaluation of an Accurate Calibration and Spectral Standardization Procedure for Raman Spectroscopy

Didier Hutsebaut

Peter Vandenabeele

Luc Moens

The Analyst, 2005, 130, 1204 - 1214

Chapter II

II. Evaluation of an Accurate Calibration and Spectral Standardization Procedure for Raman Spectroscopy

Due to modern developments Raman spectroscopy has evolved into a fast vibrational technique. Detailed fingerprints in combination with non-destructivity and minimal sample preparation has allowed the construction of reference libraries in a variety of research fields. Long-term stability and comparability are important characteristics when developing reference libraries. In addition, small shifts in highly similar spectra of different samples may limit the full potential of Raman spectroscopy. Since libraries often contain a large number of different and/or highly similar spectra, it is important that each data point in all the spectra corresponds to the exact Raman wavenumber. This is often not the case, due to shifts in optical pathway and/or shifts in laser wavelength.

This paper describes a complete calibration protocol (wavelength and intensity) and evaluates the procedure for both short and long term stability, by means of 60 randomly selected measurement sessions spread over a period of nine months. A two-step standardization procedure is proposed to deal with spectral shifts.

Introduction

Modern technical developments have made Raman spectroscopy a fast, powerful and easy-to-use analytical technique, combining detailed spectral fingerprints, non-destructivity and minimal sample preparation. It used to be a complex and slow technique, plagued with fluorescence interference, while nowadays, high quality spectra from even one single cell can be collected in less than three minutes, provided the Raman system is fully optimized for the particular field of interest[13]. Hence, it is no surprise that reference libraries are constructed in all different kinds of research fields as art analysis[3], pharmaceutical sciences[12] and clinical microbiology[9,10].

However, the stability of reference libraries, especially over longer periods, demands some special precautions. Spectra of highly related samples sometimes differ only by small Raman shifts; these differences may not be obscured by calibration errors. This sets some important requirements towards the wavelength calibration. Also, the spectra collected on different days and over an infinite period of time, of the same, inert sample, should -under ideal circumstances- be identical, except for noise. An intensity calibration procedure ensures such long-term stability and provides reliable relative intensities which are valuable for spectral library searches and quantitative analysis[7, 8].

Finally, prior to automated data-analysis, spectra should be merged in a large data matrix (reference library) such that each column represents exactly the same Raman wavenumber. The latter implies a procedure for spectral standardization.

This paper describes a calibration protocol and reports on short- and long-term stability of both wavenumber and intensity calibration. In addition, a method to deal with shifted Raman spectra is presented, commented and evaluated by data collected during the last nine months. The described protocol aims to contribute to reliable, comparable databases with a long-term stability.

Experimental

Reagents. 1,4 bis(2-methylstyryl)benzene ((Bis(MSB), scintillation grade, 99.5%), Naphthalene (scintillation grade, 99 + %), 4-Acetamidophenol (98%), epsilon-caprolactone (*p.a.*) were all obtained from Acros Organics. Polystyrene pellets (secondary standard) was obtained from Aldrich. Acetonitrile (HPLC grade) / Toluene (*p.a.*), obtained from Panreac and UCB respectively, were mixed in 50/50 volume %. Sulphur (precipitated, high purity) was obtained from UCB. Cyclohexane was used as obtained from Kaiser.

Software. Calibration and standardization was performed by in-house written software developed within the Matlab software package, version 6.5 (The Mathworks inc. Natick, MA) and the PLS toolbox, version 3.0 (Eigenvector Research inc, Manson, Wash.).

Spectrometer. All experiments were performed using a Kaiser System Hololab 5000R modular Raman microspectrometer (f/1.8) (KOSI, Ecully, France). The microscope was fitted with a 100x objective (PL Fluotar L, N.A. 0.75, W.D. 4.7 mm, Leica). Samples were excited using 45 - 50 mW of 785 nm laser light from a diode laser (Toptica Photonics AG, Martinsried/Munich, Germany). The scattered light is guided to the spectrograph by means of a confocal, 15 µm aperture collection fiber. A back illuminated deep depletion Pelletier cooled CCD detector (Andor, Belfast, Northern Ireland) operating at –70°C was used for the detection of the scattered light. Raman signal was collected in the spectral interval of 150 cm^{-1} – 3100 cm^{-1} with a spectral resolution of 4 cm^{-1}. Calibration of the absolute wavelength-axis was performed using the known wavelengths of the atomic lines from neon. The reference spectrum of a 200 W tungsten band lamp (Optronic Laboratories, Inc, Orlando, USA) operated at 6.500 A was used to correct for the wavelength dependent signal detection efficiency of the Raman set-up[13].

Neon spectra. The neon lamp as obtained from Kaiser was used. The neon lamp equipped with a diffuser is placed under the microscope objective. Neon spectra are collected with an accumulation time of 20 s in order to take advantage of the weaker lines of neon. The accumulation time was determined such that the neon line images were as intense as possible, short of saturation[2]. To prevent from calibration errors due to sample positioning, the neon lamp was moved thoroughly and relative to the microscope objective during accumulation time. Hence, complete and reproducible filling of the aperture of the objective was ensured. During the measurement of samples, additional neon spectra were periodically collected to validate the stability of the setup.

White light spectra. A calibrated white-light source (Quartz Tungsten Halogen (QTH) lamp, Optronic laboratories, Inc., Orlando, USA) operating at 6.500 A was used, with an accuracy of ± 0.5 % relative to the NIST scale over the relevant wavelength range (250 nm – 2500 nm). A microprocessor controlled precision DC source (OL 65 A, Optronic laboratories, Inc., Orlando, USA) is used to operate the lamp standard precisely: the error on the current is less than 0.01% at 6.50 A. In order to simulate sampling geometry of the Raman sample, such that both Raman scattering and emission of the QTH lamp follow the same light path trough the spectrometer, a reflection standard (Optical-grade, Labsphere, Inc. North Sutton) was used. The optical-grade Spectralon® reflectance material is characterized by a high (> 99%) reflectance in the wavelength region 400 nm – 1500 nm. To further reduce potential errors, 30 emission spectra of a white light source, collected at random locations at the Spectralon® and in identical conditions as the Raman sample, are averaged out. This procedure prevents interference of chromatic aberrations, which are known to yield different response curves for different sample positions[7]. Prior to averaging, all white light spectra were dark corrected and point-normalized to correct for multiplicative effects.

Dark signal. The collected spectra (e.g. emission spectra of a white source, Raman spectra of routine samples, optics,...) were dark corrected to remove the signal contribution of spontaneously generated electrons of the CCD detector. This non-constant contribution is independent of the transmission efficiency of the Raman set-up and should consequently be subtracted from all spectra before intensity calibration. Prior to the collection of the dark signal, the laser was switched off and the room was completely darkened to prevent incident light on the detector. Thirty dark signals were collected in identical conditions as the routine samples and averaged out.

Optics spectra. All optical components in a Raman spectrometer show small contributions to the final spectrum of routine samples. Thirty optics spectra were collected under similar conditions as the routine samples with that difference that no sample is in focus. Before subtraction from the Raman spectra of routine samples, optics spectra are dark corrected, averaged out and shifted towards the reference axes as will be discussed further on. A representative dark signal and Raman spectrum of the optical background are illustrated in Figure 1.

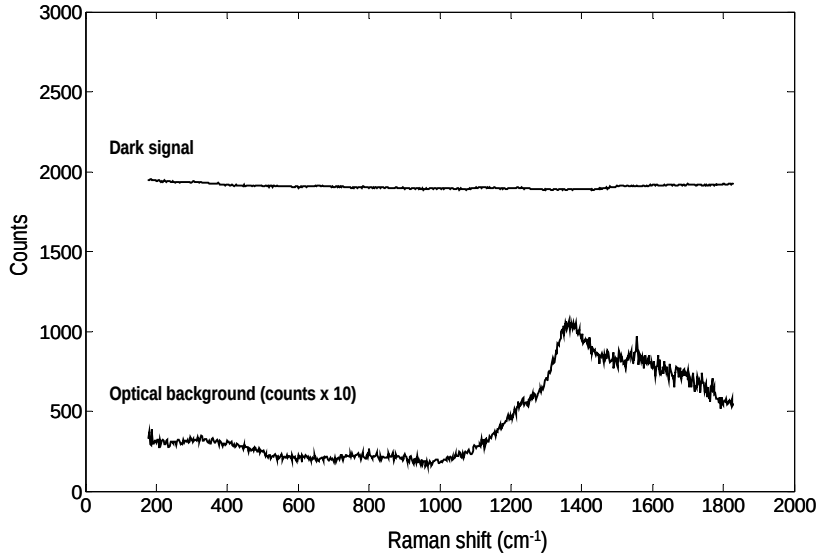


Figure 1: Representative dark signal and Raman spectrum of optical background, both collected with an accumulation time of 60s. The Raman spectrum of the optical background was corrected for dark signal and transmission efficiency of the set-up. For clarity, the optical background is multiplied by a factor 10.

Mathematical procedures

Estimation of the positions of neon lines and Raman bands. In order to perform an accurate wavelength calibration, it is important to focus on the procedure to determine the exact peak position of Raman bands or neon emission lines on the CCD (charge-coupled device). To do so, several approaches are possible. One could accept the pixel position (data point) corresponding with the maximum intensity of the band, or even better, fit a polynomial function to the (sharp) band. Although the second idea is to be preferred since interpolation of pixelposition is involved, it still has some disadvantages. Indeed, Raman bands and even to a higher extent neon emission lines, cover few pixel positions (data points) and the spectral window trough which the function should be fitted depends on the band shape. This operation is hampered by an abrupt change of slope, which makes the operation sensitive to noise and to variations in the band shape[4].

The procedure used in this work to measure the pixel positions (data points) of both irregularly shaped Raman bands and neon lines was adapted from Tseng *et al.*[5]. Briefly, it involves transformation of a spectral window (8 data points), containing the maximum of the peak (Figure 2A), to the Fourier domain where it is extended 64-fold by zero filling. This results in a new spectral window of 512 points.

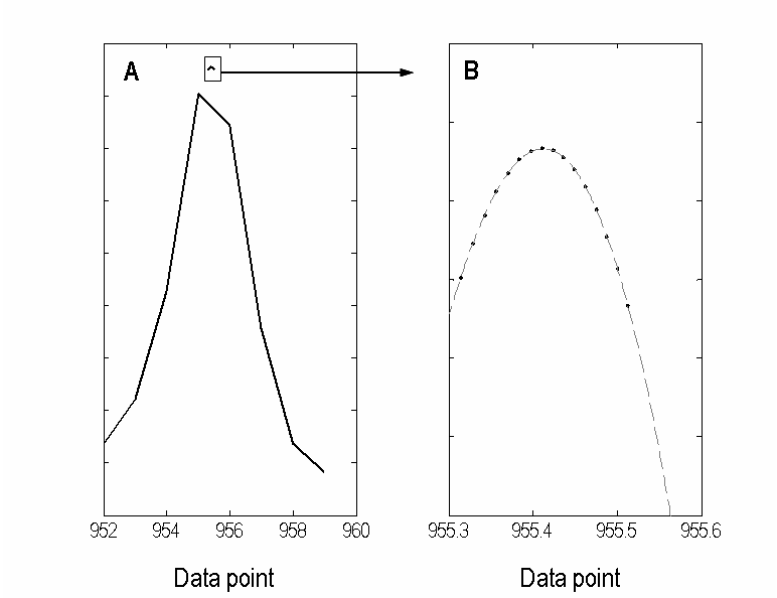


Figure 2: A) Detail of the spectral window (8 data points) containing the maximum of a neon emission line, prior to interpolation in the Fourier domain. B) Detail of the window enclosed in 2A showing the result after zero filling and calculation of the inverse transform.

After the inverse transform, a fraction of the original spectral window is defined such that the peak maximum is centered. The new spectral window contains 16 data points (Figure 2B). The position of the peak maximum is subsequently estimated by fitting a 2nd order polynomial function to the window and the determination of the interpolated pixel position for which the first derivative of this function becomes zero.

Outlier detection. Outlier detection was performed on the Raman wavenumber standards to remove possibly erroneous determined band positions. Therefore, for each calibration sample, the pixel positions (data points) for all Raman bands used during calibration were determined. For all but one calibration sample (i.e. sulphur) a forth-order polynomial fit was used to correlate Raman wavenumber to pixel position (data point). For sulphur, a second-order polynomial fit was used because only three Raman bands were available (Table 1). Next, the Raman wavenumbers were recursively shifted in the interval [–0.2 cm^{-1}, 0.2 cm^{-1}] to maximize the number of Raman bands that fall within the reported values ± one standard deviation. To obtain a high precision in Raman band position, the bands with an absolute error larger than one standard deviation in their Raman wavenumbers are excluded from the calibration. Finally, the original, not-shifted Raman wavenumbers of the included bands are used for calibration.

Overview of the Calibration Procedure

The calibration procedure described in this work, aims at performing an appropriate calibration of the Raman spectrometer, both in terms of the Raman wavenumber-axis and spectral intensity.Wavelength calibration finds the relation between the data point or pixel-position on the CCD and a physical variable, which expresses the energy of the radiation. Since the Raman wavenumber is an expression of the difference between the laser wavenumber (cm^{-1}) and the wavenumber of the scattered radiation, two procedures are required for the full calibration. First, absolute wavelength calibration is performed, relating the absolute wavelength of the scattered radiation to pixel position (or data point). Absolute wavelength calibration is necessary for intensity calibration, which corrects for the wavelength-dependent transmission efficiency of the optics and the spectrometer, as well as for the wavelength-dependent sensitivity of the detector. Second, pixel position (or data point) can be related to the Raman wavenumber axis by Raman wavenumber standards. In this work, Raman wavenumber standards are preferred to absolute wavelength standards (e.g. neon) for the calibration of the Raman wavenumber axis, since the emission wavelength of the diode laser in not accurately known. The calibration procedure (intensity as well as wavenumber axis) is completed with a mathematical standardization procedure. This procedure corrects the collected Raman spectra for minute variations in the optical pathway of the instrumentation and shifts in laser wavelength. In the next paragraphs, the different aspects of this calibration procedure will be discussed in detail.

Absolute wavelength calibration. Absolute wavelength calibration correlates the pixel position on the CCD with the absolute wavelength (nm scale) by means of the emission lines of a neon bulb. In Figure 3 a neon spectrum is shown; to visualize the weaker neon emission lines, the natural logarithm of the intensity is plotted. Peaks used to determine the absolute wavelength axis are labeled. Although the spectrometer is calibrated in the range 794.3181 nm - 948.6680 nm, the range used is restricted to 808.31 nm – 909.02 nm, to avoid extrapolation. Positions of neon-lines were determined as described (cf. *Mathematical procedures*) and the relation between the absolute wavelength axis and pixel position (data point) was determined using linear interpolation. As an indicative value for the goodness of fit, the root mean square error (RMSE) was calculated. This value was generally < 0.03 nm.

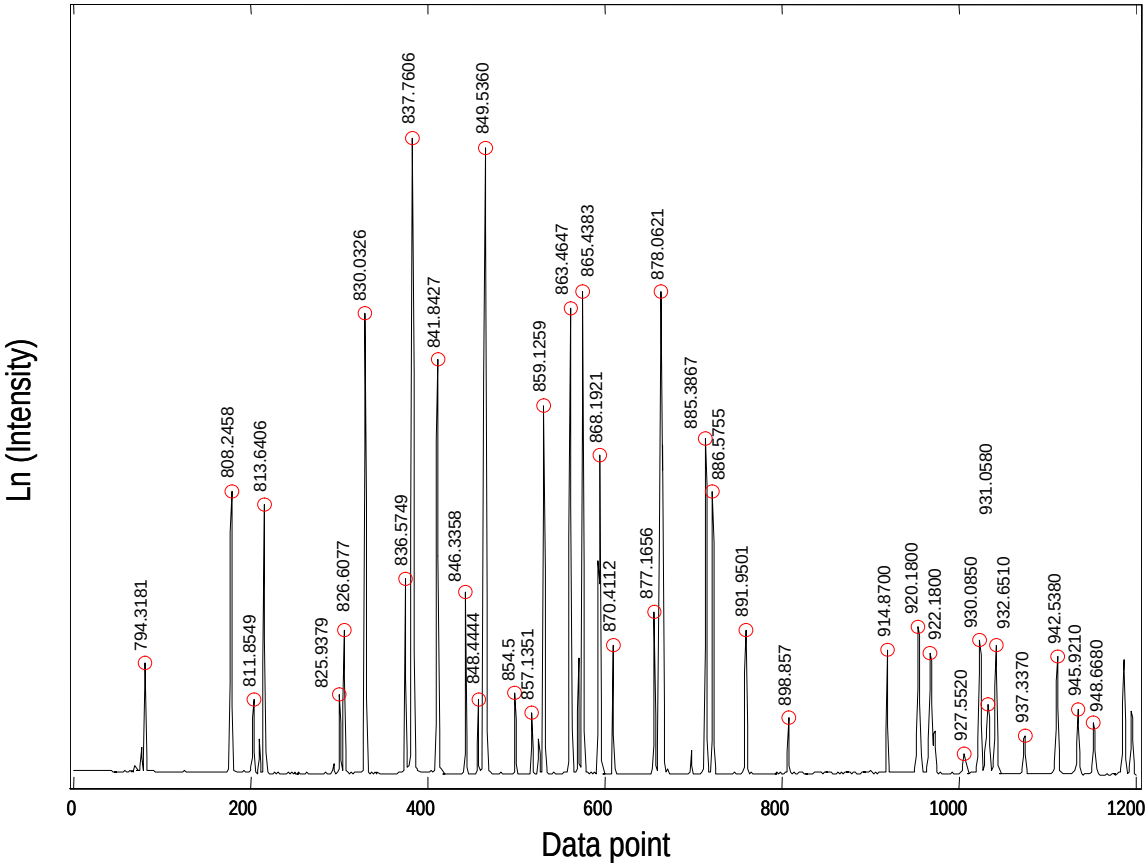


Figure 3: Neon spectrum (not corrected for the transmission efficiency of the set-up). Emission lines (nm) used for wavelength calibration are labeled. To show the less intense lines, the natural logarithm (Ln) of the intensity is plotted. The selected lines cover the complete spectral window used for analysis.

Raman wavenumber calibration. In the following, a '*measurement session*' is defined as a series of independent Raman measurements, collected under identical conditions, during which at the beginning and at the end of this session, as well as at well-defined time intervals during the session, a Raman spectrum of a control sample and a neon emission spectrum are recorded. Hence, during one measurement session at least three Raman spectra of a control sample and three emission spectra of neon are recorded. Each day the spectrometer is fully calibrated and several measurements sessions can take place during one day. The laser may be switched off in-between these measurement sessions.

Table 1: Raman wavenumbers (Raman shifts) used for calibration of standard samples. Selected Raman wavenumbers as reported by ASTM E1840-96. [a] Standard deviations as reported after analysis of six independent laboratories.

Component	*Raman Shift (± standard deviation)* [a]	*Component*	*Raman Shift (± standard deviation)* [a]	*Component*	*Raman Shift (± standard deviation)* [a]
4-acetamidophenol	329.2 ± 0.52	bis(MSB)	456.0 ± 0.56	Polystyrene	620.9 ± 0.69
	390.9 ± 0.76		642.4 ± 0.12		795.8 ± 0.78
	465.1 ± 0.30		841.6 ± 0.28		1001.4 ± 0.54
	504.0 ± 0.60		950.1 ± 0.13		1031.8 ± 0.43
	651.6 ± 0.50		978.0 ± 0.16		1155.3 ± 0.56
	710.8 ± 0.68		1104.1 ± 0.31		1450.5 ± 0.56
	797.2 ± 0.48		1177.7 ± 0.56		1583.1 ± 0.86
	834.5 ± 0.46		1290.7 ± 0.28		1602.3 ± 0.73
	857.9 ± 0.50		1316.9 ± 0.94		
	968.7 ± 0.60		1334.5 ± 0.16	Sulphur	153.8 ± 0.50
	1105.5 ± 0.27		1555.2 ± 0.19		219.1 ± 0.57
	1168.5 ± 0.65		1593.1 ± 0.44		473.2 ± 0.49
	1236.8 ± 0.46		1627.9 ± 0.23		
	1278.5 ± 0.45			Naphthalene	513.8 ± 0.31
	1323.9 ± 0.46	Cyclohexane	384.1 ± 0.78		763.8 ± 0.31
	1371.5 ± 0.11		426.3 ± 0.41		1021.6 ± 0.49
	1515.1 ± 0.70		801.3 ± 0.96		1147.2 ± 0.34
	1561.5 ± 0.52		1028.3 ± 0.45		1382.2 ± 0.31
	1648.4 ± 0.50		1157.6 ± 0.94		1464.5 ± 0.29
			1266.4 ± 0.58		1576.6 ± 0.29
Acetonitrile / Toluene (50/50)	378.5 ± 0.92		1444.4 ± 0.30		
	521.7 ± 0.34				
	786.5 ± 0.40				
	919.0 ± 0.40				
	1003.6 ± 0.37				
	1030.6 ± 0.36				
	1211.4 ± 0.32				
	1605.1 ± 0.47				
	2253.7 ± 0.42				
	2292.6 ± 0.89				

Although the Raman set-up is free of moving parts, expansions of optical components with temperature and optical aberrations can generate errors of several wavenumbers. Even when the laser line is not susceptible to shifts or errors, one-point calibration is inaccurate for most dispersive systems. Ideally, a large number of accurately known frequencies is dispersed across the spectrum[11]. Since it is well known that the laser wavelength of diode lasers may vary with time and temperature, Raman wavenumber standards were preferred to calibrate the Raman instrumentation. This only makes sense if the laser line remains at the same pixel position during the measurement sessions of that day. Unless the laser was realigned or switched off, the measured laser wavelength proved to be constant during the timeframe of a measurement session.

Raman wavenumber standards have the advantage of duplicating sample geometry, but result in a lower density of available frequencies compared with atomic emission lines. Therefore, seven Raman wavenumber standards were used to determine the Raman wavenumber axis (153.8 cm^{-1} – 2292.6 cm^{-1}) while only the spectral window in-between 370 cm^{-1} and 1800 cm^{-1} was used during data analysis. This has three advantages: (1) no Raman shift errors are introduced by extrapolation to an 'unknown' spectral region (2) a large number of Raman bands (67) is available to calibrate the spectral window used, resulting in a relatively high density of Raman bands and (3) small shifts originating from changing sample geometry are reduced since seven different standards are used.

Raman wavenumbers of the different standards used during calibration are listed in Table 1. The standards have been examined under authority of the ASTM (American Society for Testing and Materials) by at least six independent laboratories using both dispersive and FT-Raman spectrometers[1,11]. These Raman wavenumbers (± 1 standard deviation) were used as reference values. The pixel positions of the Raman bands were determined as mentioned above.

It has to be noted that the reported values (Table 1) were determined using 1064 and 514.5 nm lasers and that, due to resonance effects, these values may differ when the standards are excited at different wavelengths (e.g. 785 nm). In addition, procedures for determination of Raman wavenumbers were not standardized during the reported study and the laboratories did not use a standard calibration procedure[1]. Finally, noise may contribute to erroneous peak positioning. Therefore, prior to Raman wavenumber calibration, outlier detection was performed to reject less accurate determined band positions.

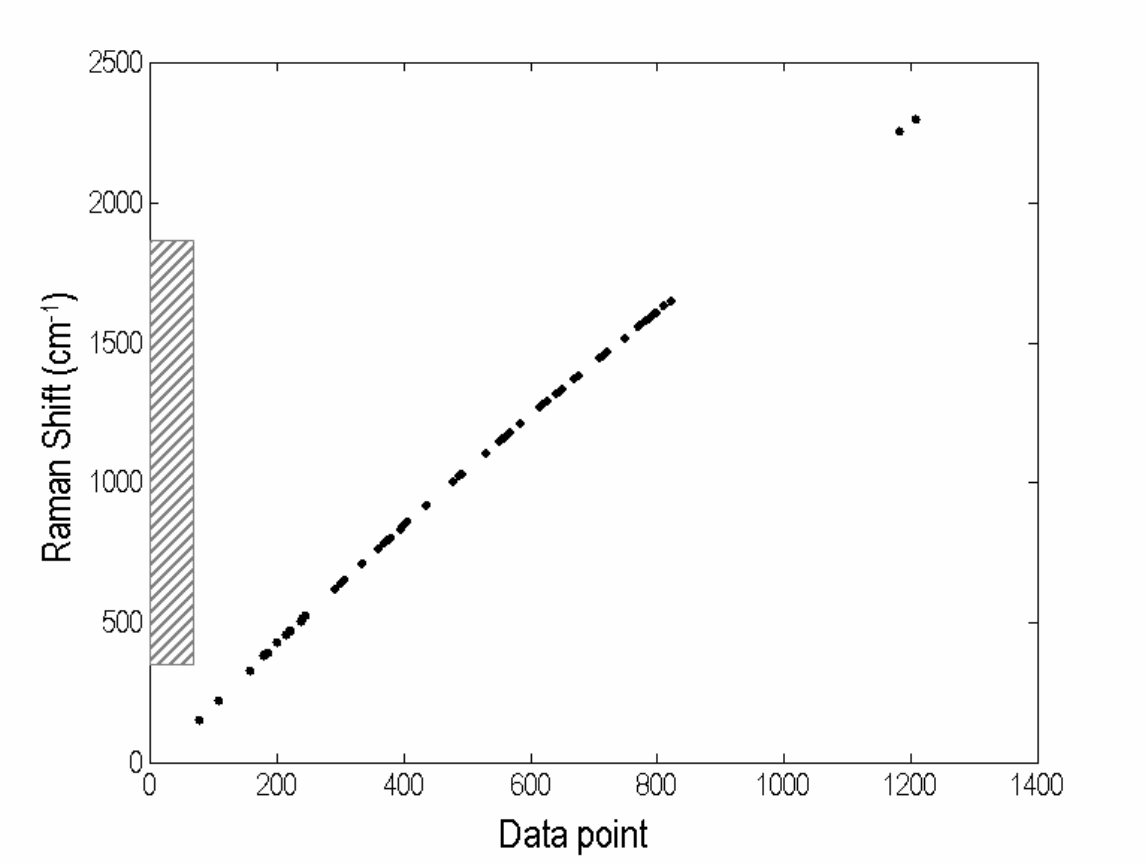


Figure 4: Relation between the 67 Raman wavenumbers (cm^{-1}) of reference materials used during calibration and their pixel position (data point). The spectral window used during analysis (shaded bar) is covered by a large number of peaks.

To emulate a high Raman wavenumber accuracy the accepted error interval was restricted to ± one standard deviation. Indeed, in contrast to the conditions under which the reference Raman wavenumbers (Table 1) were reported, only one excitation wavelength (˜ 785 nm), one Raman set-up and a standardized protocol are used. One could argue that given this stringent condition of acceptance, some Raman bands might be falsely rejected. However, we found that bands were rarely rejected with exception of few Raman bands of BISMSB (e.g. 978.0 cm^{-1} and 1334.5 cm^{-1}) and 4-acetamidophenol (e.g. 1105.5 cm^{-1} and 1371.5 cm^{-1}). However, these bands were not always rejected and could be traced back to erroneous determination of pixel position (data point). The Raman wavenumber axis was calculated by linear interpolation of the Raman wavenumbers in function of the pixel position (data point) of the corresponding Raman bands (Figure 4).

This approach is in accordance with some important issues brought up in literature earlier by Carter *et al.*[6] concerning wavelength calibration: (1) A sufficiently large number of evenly spaced Raman bands is available for calibration, covering the region of interest more or less completely, (2) the measurement and sampling conditions resemble those used for the samples and (3) the frequencies of the standards are accurately known and experimentally reproducible to within 0.5 cm^{-1}.

Intensity calibration. Raman spectroscopy is a single-beam technique lacking the internal response calibration inherent in absorbance measurements. Different spectrometers will vary in quantum-efficiency of the CCD, laser wavelength, grating efficiency, reflection and/or transmission efficiencies of the optical components (e.g. objective(s), grating, mirrors, optics…). As a result, spectra of the same sample, measured at different instruments, even at the same laser wavelength, can deviate importantly because of different response functions[8]. The response function is responsible for changes in relative intensities whenever the Raman set-up has changed (e.g. replacement of an objective, notch filter or detector). Such variations greatly complicate useful spectroscopic operations as library search and affect long-term stability of reference libraries. Therefore, during the calibration procedure dark signal, optics and white light spectra are recorded, to allow correction for their contributions.

Standardization. Once the Raman set-up is fully calibrated and the reproducibility of peak positions and relative intensity is ensured, spectral data are merged into one data matrix (reference library), prior to multivariate analysis. In this data matrix, all values in a particular column are supposed to represent a measurement at exactly the same Raman wavenumber. Due to variations in optical pathway and measured laser wavelength - the laser was switched off between the calibration and the start of a particular measurement session - spectra will be projected on different pixels. Therefore, the collected spectra need to be standardized towards a reference axis. With the application of such a standardization procedure, the collected Raman spectra are recalculated as if they were all collected at identical conditions (optical pathway and laser wavelength).

The Kaiser Hololab spectrometer is a fixed-grating system without any moving parts and is less sensitive to temperature changes, vibrations and other environmental conditions than systems with a longer focal length or movable gratings. However, shifts in the projection of the Raman spectra relative to the detector can still occur, because of varying ambient conditions and shifts in laser wavelength. Small but significant shifts (˜ 0.2 – 1.0 cm^{-1}) in the Raman spectrum were noticed whenever the laser was switched on. The latter seemed to be responsible for significant variations between different measurement sessions. These shifts have a major impact on multivariate analyses, especially when the spectra are highly similar, which is often the case in bio-spectroscopy. The procedure described in this work is two-fold and assumes that no significant shifts (nor absolute as relative) occur during the measurement sessions. It will be shown that this assumption is fulfilled.

Corrections and recalculation to reference wavelength axis. The first phase of standardization takes absolute shifts (e.g. as a result of changes in the optical pathway) into account. The average emission spectrum of at least three scaled neon spectra, collected during the measurement sessions and at regular time intervals, is calculated. This average neon emission spectrum (N_{avg}) is shifted by linear interpolation over few data points in the range [-5, 5] to minimize the sum of squared differences between the average neon emission spectrum and the reference neon emission spectrum.

The latter is expressed in the following performance function:

$$\Sigma(\mathbf{N}_{avg} - \mathbf{N}_{ref})^2 \qquad (1)$$

in which $\mathbf{N}_{avg}$ is the average emission spectrum of the normalized and shifted (i.e. by linear interpolation) neon spectra collected during the measurement session and $\mathbf{N}_{ref}$ is the normalized reference neon emission spectrum, which was recorded at a particular time (e.g. at he introduction of the Raman set-up in the laboratory). This reference spectrum used is the same for all measurement sessions.

By minimizing the sum of squared differences between the average neon spectrum and the reference neon spectrum, it is guaranteed that the neon spectra during all measurement sessions are identical to the reference neon spectrum. Hence, differences in pixel position (or data points) due to changes in the optical pathway are corrected for. It is noted that the procedure can be extended by minimizing the sum of squared differences between each individual neon spectrum and the reference neon spectrum allowing correction of absolute wavelength shifts during the measurement sessions. However, such shifts were not found during our experiments and hence for simplicity the procedure was used as described.

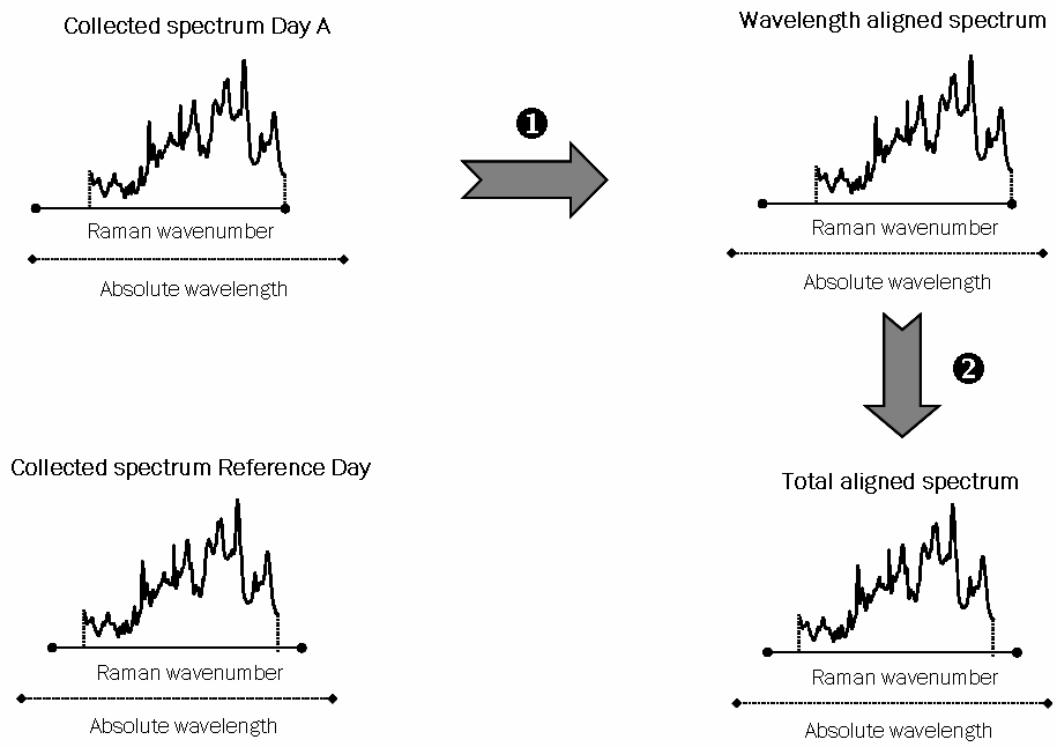


Figure 5: Illustration of the two-fold spectral alignment procedure. ❶ The collected Raman spectrum is shifted towards the reference wavelength axis to correct for changes in the optical pathway of the Raman set-up. ❷ The resulting Raman spectrum is shifted towards the reference Raman wavenumber axis to correct for changes in laser wavelength. The result is identical to the spectrum collected on the reference day.

Once the number of data points for which the function reaches a minimum (shift Δ) is known, all spectra (e.g. emission spectra of a white light source and Raman spectra of optical background, control sample, routine samples…) collected during a specific measurement session are recalculated by means of linear interpolation from the wavelength axis during the measurement session towards the reference wavelength axis. This is illustrated in Figure 5, step 1. In doing so, all spectra are represented as if the optical pathway of the Raman set-up was identical to the one on the reference day.

Laser wavelength correction. The second phase of standardization takes care of shifts in laser wavelength. The average was calculated of at least three Raman spectra of the control sample (ε-caprolactone), collected at regular time intervals and each shifted towards the reference absolute wavelength axis. The first derivative of the average control Raman spectrum was calculated ($\mathbf{C}_{avg}$) and shifted by linear interpolation over few data points to minimize the sum of squared differences between the average, derivative spectrum and the derivative reference spectrum. The latter is expressed in the following performance function:

$$\sum (\mathbf{C}_{avg} - \mathbf{C}_{ref})^2 \qquad (2)$$

in which $\mathbf{C}_{avg}$ is the average, normalized first derivative spectrum of the control sample, collected during the measurement session and $\mathbf{C}_{ref}$ is the normalized, first derivative spectrum of the reference material (ε-caprolactone). First derivative spectra (Savitzky-Golay, 9 points window, 3th order polynomial for interpolation) were used during laser wavelength correction to avoid background contributions interfering with the minimization. The second phase of standardization is illustrated in Figure 5, step 2.

After standardization, the Raman spectra are corrected for the contribution of the background generated by the optics and divided by the response curve to yield the intensity corrected Raman spectra of the samples.

Discussion

To evaluate the calibration procedure as discussed so far, two issues are of main importance. First, since the Raman set-up is daily calibrated, both set-up and laser wavelength need to be stable over the measurement sessions of the day. This concerns the short-term stability of the set-up and will be assessed by an absolute calibration standard (neon) and a control sample (ε-caprolactone) to validate the stability of the optical pathway and laser wavelength, respectively. Both cases include errors introduced by adjustment of the sample position.

Second, the equipment will be validated over a longer period of time by means of calibration data that were collected in our lab during nine months. To do so, 60 measurement sessions were randomly selected. The latter concerns all kind of changes that are common sense in a Raman laboratory: daily recalibration of the equipment, sample repositioning, alignment of the instrumentation, re-coupling of the laser, etc. By evaluation of both short- and long-term stability, it is intended to provide an accurate indication of the performance of the calibration protocol.

Short-term reproducibility

The short term reproducibility of both the absolute and relative wavelength calibration was validated by comparing spectra, absolute wavelengths (neon) and Raman wavenumbers (ε-caprolactone). Neon emission spectra and ε-caprolactone Raman spectra were both taken at regular time intervals during the measurement sessions on ten randomly selected days.

Precision of Raman wavenumbers. For each of the 10 measurement sessions, Raman wavenumbers for ε-caprolactone (after calibration and prior to standardization) are reported in Table 2. Standard deviations for *n* measurements, including sample replacement, are given between brackets. The Raman spectra of the control samples show high reproducibility as illustrated by the standard deviations of the Raman wavenumbers. However, some of the standard deviations show relatively high values (e.g. 1253.7 cm^{-1} ± 2.12). These values are not indicative for a shift in laser wavelength, since these high standard deviations are not found for the other peak positions of the same sample. Spectra of the control samples taken during a particular session and their difference spectra (Figure 6) confirm that no shifts in laser wavelength occur during the measurements, contributing to the short-term reproducibility. Although no shifts were detected, it is noticed that some small bands remain after subtraction, while a random noise band is expected. This is explained by a different contribution of the glass vessel containing the ε-caprolactone to the collected Raman spectrum, affecting the mean of the SNV scaled spectra. This is not an artifact of the intensity calibration (cf. *Accuracy in determining the Response curve***)** and is entirely depending on sample focusing.

Table 2: Overview of the mean (for *n* measurements) Raman band positions of ε-caprolactone during 10 randomly selected measurement sessions. Standard deviations for *n* replicate measurement with sample replacement are given between brackets. Values as found after full wavelength calibration, without spectral alignment.

Session 1 (*n* = 5)	Session 2 (*n* = 4)	Session 3 (*n* = 4)	Session 4 (*n*= 4)	Session 5 (*n* = 4)	Session 6 (*n* = 5)	Session 7 (*n* = 4)	Session 8 (*n* = 5)	Session 9 (*n* = 4)	Session 10 (*n* = 4)
489.0 (±0.39)	488.5 (±0.62)	488.9 (±0.43)	489.7 (±0.21)	488.7 (±0.44)	488.6 (±0.14)	488.5 (±0.16)	488.7 (±0.17)	488.7 (±0.22)	488.4 (±0.29)
572.0 (±0.17)	571.8 (±0.19)	571.5 (±0.35)	571.7 (±0.26)	571.6 (±0.28)	571.2 (±0.05)	571.3 (±0.06)	571.4 (±0.10)	571.5 (±0.13)	571.3 (±0.05)
696.8 (±0.26)	696.1 (±0.24)	696.8 (±0.22)	696.3 (±0.13)	696.8 (±0.19)	696.8 (±0.09)	696.6 (±0.06)	696.8 (±0.03)	696.7 (±0.09)	696.6 (±0.10)
735.2 (±0.30)	734.9 (±0.25)	735.4 (±0.23)	735.0 (±0.18)	735.0 (±0.08)	735.2 (±0.13)	734.9 (±0.15)	735.1 (±0.05)	735.1 (±0.08)	734.9 (±0.10)
818.4 (±0.24)	818.2 (±0.15)	818.0 (±0.36)	818.0 (±0.27)	818.4 (±0.13)	818.4 (±0.17)	818.2 (±0.12)	818.5 (±0.17)	818.3 (±0.16)	818.4 (±0.16)
849.6 (±0.23)	849.4 (±0.13)	849.2 (±0.33)	849.1 (±0.32)	849.3 (±0.20)	849.0 (±0.10)	849.0 (±0.14)	849.2 (±0.02)	849.2 (±0.13)	849.1 (±0.07)
1016.8 (±0.22)	1016.7 (±0.12)	1016.3 (±0.16)	1016.3 (±0.13)	1016.6 (±0.14)	1016.4 (±0.05)	1016.3 (±0.06)	1016.4 (±0.07)	1016.2 (±0.11)	1016.2 (±0.11)
1056.8 (±0.59)	1056.2 (±0.80)	1057.0 (±0.33)	1056.7 (±0.67)	1056.9 (±0.31)	1056.5 (±0.36)	1056.4 (±0.07)	1056.8 (±0.22)	1056.8 (±0.17)	1056.6 (±0.09)
1088.8 (±0.24)	1088.4 (±0.16)	1088.6 (±0.09)	1088.7 (±0.14)	1088.5 (±0.20)	1088.3 (±0.10)	1088.3 (±0.36)	1088.9 (±0.15)	1088.5 (±0.29)	1088.5 (±0.24)
1165.7 (±1.25)	1165.0 (±1.38)	1164.2 (±1.33)	1165.7 (±0.92)	1165.4 (±0.64)	1164.7 (±1.06)	1164.8 (±0.29)	1164.7 (±0.81)	1164.0 (±0.40)	1164.0 (±0.68)
1226.9 (±0.27)	1226.3 (±0.51)	1225.9 (±0.74)	1226.2 (±0.51)	1226.1 (±0.71)	1226.0 (±0.23)	1225.5 (±0.35)	1226.1 (±0.68)	1225.9 (±0.39)	1226.0 (±0.33)
1254.9 (±1.47)	1253.8 (±2.12)	1253.7 (±2.15)	1253.7 (±2.12)	1253.9 (±2.02)	1253.1 (±0.43)	1252.5 (±0.41)	1253.2 (±0.52)	1253.1 (±0.48)	1252.9 (±0.38)
1328.8 (±0.26)	1328.5 (±0.34)	1328.7 (±0.24)	1328.4 (±0.35)	1328.9 (±0.34)	1328.9 (±0.20)	1328.4 (±0.31)	1328.9 (±0.48)	1329.0 (±0.23)	1328.9 (±0.31)
1363.9 (±0.70)	1363.6 (±0.63)	1363.1 (±1.04)	1362.7 (±1.50)	1363.0 (±1.31)	1362.9 (±0.59)	1362.5 (±0.65)	1363.9 (±0.76)	1363.7 (±0.62)	1364.0 (±1.12)
1448.1 (±0.61)	1449.5 (±1.12)	1449.6 (±1.20)	1450.2 (±0.41)	1448.3 (±0.52)	1448.2 (±0.17)	1449.4 (±0.88)	1449.1 (±0.35)	1449.0 (±0.40)	1450.1 (±0.33)

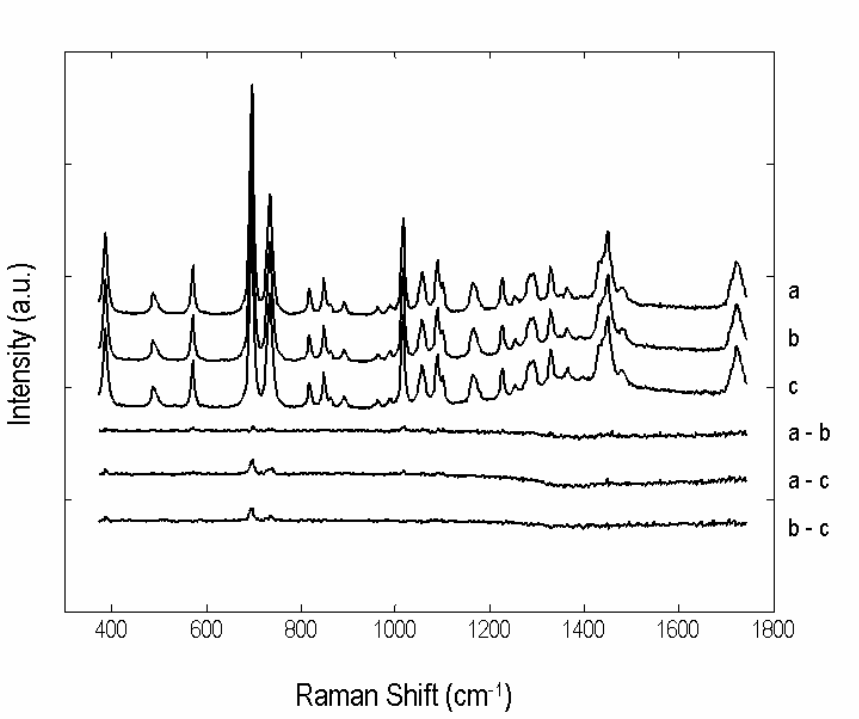


Figure 6: Representative Raman spectra of ε-caprolactone collected during randomly selected measurement sessions (a, b and c). Their difference spectra (a – b, a – c and b - c) reveal that no shifts occur during the measurement sessions. Spectra were scaled (SNV) prior to subtraction. For clarity, spectra were shifted vertically. (a.u.) = arbitrary units.

Precision in determining atomic line wavelengths. Values for neon emmisionlines for each of ten randomly selected measurement sessions are reported in Table 3. Standard deviations for *n* measurements including sample replacement are given between brackets. The optical pathway shows high stability as illustrated by the standard deviations of the peak positions for neon. This is due to the absence of moving parts in the spectrometer. Accuracy of determination of the neon emission wavelengths (considering the tabulated values as "real" values) fluctuates between < 0.01 and 0.05 nm and standard deviations on replicate, independent determinations lie between 0.02 and < 0.005 nm.

The validation of short term reproducibility shows that laser wavelength and Raman set-up are stable within the measurement sessions.

Table 3: Overview of neon emission lines during 10 randomly selected measurement sessions. Standard deviations for *n* replicate measurement with sample replacement are given between brackets. Values as found after full wavelength calibration, without spectral alignment. Thirteen neon lines were randomly selected within the spectral window for clarity of the table.

Emissionline (nm)	***session 1 (n = 4)***	***session 2 (n = 2)***	***session 3 (n = 3)***	***session 4 (n = 2)***	***session 5 (n = 3)***	***session 6 (n = 3)***	***session 7 (n = 3)***	***session 8 (n = 4)***	***session 9 (n = 2)***	***session 10 (n = 3)***
813.6406	813.63 (±0.03)	813.63 (±0.03)	813.62 (±0.02)	813.62 (±0.03)	813.62 (±0.03)	813.61 (±0.01)	813.58 (±0.01)	813.60 (±0.01)	813.61 (±0.02)	813.61 (±0.02)
830.0326	830.04 (±0.01)	830.04 (±0.01)	830.04 (±0.00)	830.04 (±0.01)	830.04 (±0.01)	830.05 (±0.01)	830.04 (±0.00)	830.04 (±0.01)	830.04 (±0.01)	830.04 (±0.00)
837.7606	837.75 (±0.01)	837.74 (±0.01)	837.74 (±0.01)	837.74 (±0.00)	837.74 (±0.01)	837.75 (±0.01)	837.76 (±0.00)	837.74 (±0.01)	837.76 (±0.02)	837.76 (±0.02)
841.8427	841.82 (±0.01)	841.82 (±0.01)	841.80 (±0.01)	841.80 (±0.01)	841.79 (±0.02)	841.79 (±0.01)	841.81 (±0.01)	841.80 (±0.01)	841.82 (±0.02)	841.83 (±0.02)
846.3358	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.00)	846.34 (±0.01)	846.34 (±0.01)	846.34 (±0.02)
849.536	849.54 (±0.01)	849.54 (±0.01)	849.53 (±0.00)	849.53 (±0.01)	849.52 (±0.00)	849.52 (±0.01)	849.53 (±0.00)	849.52 (±0.00)	849.53 (±0.01)	849.53 (±0.01)
859.1259	859.13 (±0.01)	859.12 (±0.01)	859.10 (±0.02)	859.12 (±0.02)	859.10 (±0.03)	859.12 (±0.03)	859.14 (±0.00)	859.11 (±0.02)	859.14 (±0.02)	859.14 (±0.02)
863.4647	863.48 (±0.01)	863.48 (±0.01)	863.47 (±0.01)	863.48 (±0.02)	863.47 (±0.01)	863.47 (±0.02)	863.48 (±0.00)	863.46 (±0.01)	863.48 (±0.02)	863.48 (±0.01)
865.4383	865.47 (±0.01)	865.47 (±0.01)	865.47 (±0.01)	865.48 (±0.01)	865.47 (±0.01)	865.48 (±0.01)	865.46 (±0.01)	865.48 (±0.01)	865.47 (±0.01)	865.47 (±0.01)
868.1921	868.22 (±0.01)	868.22 (±0.02)	868.22 (±0.02)	868.23 (±0.02)	868.22 (±0.02)	868.23 (±0.00)	868.19 (±0.01)	868.22 (±0.02)	868.20 (±0.01)	868.20 (±0.01)
878.0621	878.09 (±0.02)	878.08 (±0.02)	878.08 (±0.02)	878.08 (±0.02)	878.07 (±0.01)	878.00 (±0.00)	878.06 (±0.00)	878.06 (±0.00)	878.06 (±0.01)	878.06 (±0.01)
885.3667	885.42 (±0.02)	885.41 (±0.02)	885.41 (±0.01)	885.41 (±0.01)	885.40 (±0.01)	885.40 (±0.01)	885.40 (±0.00)	885.39 (±0.00)	885.40 (±0.01)	885.40 (±0.01)
891.9501	891.98 (±0.01)	891.97 (±0.01)	891.96 (±0.03)	891.98 (±0.01)	891.97 (±0.02)	891.98 (±0.03)	892.00 (±0.00)	891.98 (±0.01)	892.00 (±0.01)	892.00 (±0.01)

Long-term reproducibility

Precision of Raman wavenumbers. To evaluate the long-term stability of the Raman wavenumber calibration, the Raman wavenumbers of naphthalene and bis(MSB) (after calibration and prior to standardization) were determined on 60 measurement sessions (Table 4). The reported values include daily recalibration, small shifts in laser wavelength and so forth. The calibration data of the 60 sessions were randomly selected and cover a period of approximately 9 months of intensive use.

Table 4: Raman wavenumbers for naphthalene and bis(MSB). Raman wavenumbers and standard deviations as previously reported (left column) and mean values as determined over 60 randomly selected measurement sessions as determined in our laboratory (right column). Standard deviations illustrate the long term stability of the wavelength calibration.

	Raman wavenumber (cm^{-1}) ± *Standard deviation*[a]	*Mean (n = 60)* ± *Standard deviation*
Bis(MSB)	456.0 ± 0.56	456.0 ± 0.02
	642.4 ± 0.12	642.4 ± 0.12
	--	780.1 ± 0.12
	841.6 ± 0.28	841.6 ± 0.02
	978.0 ± 0.13	978.1 ± 0.19
	--	1049.0 ± 0.20
	1104.1 ± 0.16	1104.0 ± 0.15
	--	1205.1 ± 0.20
	1290.7 ± 0.28	1290.7 ± 0.07
	1316.9 ± 0.94	1316.9 ± 0.07
	1334.5 ± 0.16	1334.6 ± 0.18
	1555.2 ± 0.19	1555.3 ± 0.15
	1593.1 ± 0.44	1593.1 ± 0.21
	1627.9 ± 0.23	1627.9 ± 0.06
Naphthalene	513.8 ± 0.31	513.8 ± 0.10
	763.8 ± 0.31	763.8 ± 0.01
	1021.6 ± 0.49	1021.6 ± 0.03
	1147.2 ± 0.34	1147.2 ± 0.11
	1382.2 ± 0.31	1382.2 ± 0.02
	1464.5 ± 0.29	1464.5 ± 0.13
	1576.6 ± 0.29	1576.6 ± 0.03

[a] *Values as reported by ASTM E1840-96*

For Bis(MSB) three Raman wavenumbers of bands not included during calibration are reported (780.1 cm^{-1} ± 0.12, 1049.0 cm^{-1} ± 0.19 and 1205.1 cm^{-1} ± 0.15 cm^{-1}) exhibiting small standard deviations. Raman wavenumbers for Bis(MSB) are in congruence with the values determined by the ASTM, with standard deviations in the same order of magnitude, but generally smaller. This is not surprisingly since the reported standard deviations concern measurements on different instruments using different calibration protocols. In addition, Raman wavenumbers for naphthalene are in perfect congruence with the ones previously reported and highly reproducible with standard deviations between 0.13 and 0.01 cm^{-1}.

Precision in determining atomic line wavelengths. Table 5 shows the wavelengths of thirteen evenly spaced neon lines, their means and standard deviations as determined on 60 independent measurements. Standard deviations fluctuate between 0.007 and 0.023 nm, illustrating the high reproducibility of the absolute wavelength calibration. Comparison with the reported results in Table 3, demonstrate that the error on the absolute wavelength calibration between different days is not larger than the error on the absolute calibration on the same day.

Table 5: Wavelength of neon emission lines as determined over 60 randomly selected measurement sessions as determined in our laboratory (right column). Standard deviations given between brackets illustrate the long term stability of the wavelength calibration.

Emission line (nm)	*Mean (n = 60)*
813.6406	813.6247 (±0.020)
830.0326	830.0442 (±0.007)
837.7606	837.7457 (±0.012)
841.8427	841.8031 (±0.017)
846.3358	846.3415 (±0.007)
849.5360	849.5285 (±0.009)
859.1259	859.1193 (±0.023)
863.4647	863.4769 (±0.015)
865.4383	865.4766 (±0.010)
868.1921	868.2257 (±0.014)
878.0621	878.0746 (±0.017)
885.3867	885.4063 (±0.013)
891.9501	891.9764 (±0.022)

Precision in determining the response curve. The procedure to collect white light spectra is evaluated by comparison of the collected white light spectra between different days. Figure 7A illustrates three (randomly selected) white light spectra a, b and c as they are used to calculate the response curve prior to the correction for the wavelength dependent transmission efficiency of the Raman set-up. The white light spectra result from the average of 30 normalized spectra. The ratios of the three emission spectra (a/b, a/c and b/c) are also shown.

Figure 7B zooms in on the calculated ratios showing three noise bands with an overall mean of one. Noise bands were shifted with 0.1 (a/b) and –0.1 (b/c) for clarity. The constant ratio of the normalized white light spectra, equalling unity, shows that the relative intensities are not affected by daily recalibration of the Raman set-up and hence illustrates the reproducibility of the intensity calibration as it is used, including the determination of the peak positions of the neon emission lines.

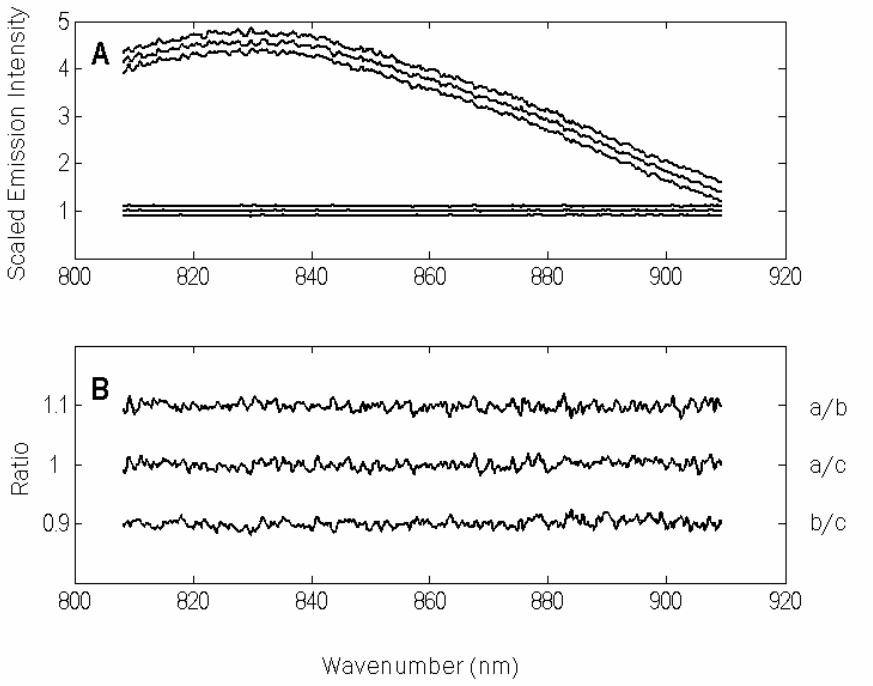


Figure 7: A. Emission spectra of a white light source during three randomly selected measurement sessions and their ratio's below. B. Detail of the ratio's of the emission spectra of a white light source. The ratios were shifted +0.1 (a/b) and –0.1 (b/c) for clarity. The ratios have a constant value equaling one and hence the relative intensities are not affected in between two independent intensity calibrations.

Spectral standardization procedure

Figure 8 illustrates the progress of performance function (2) in function of the shift values. The function gradually declines with smaller negative shift and increases again with larger positive shift. The minimum is reached when the control spectrum is shifted over –0.23039 data points. If no shift occurred the minimum of the performance function is expected at zero and hence the found minimum (Δ = –0.23039) indicates the need for standardization.

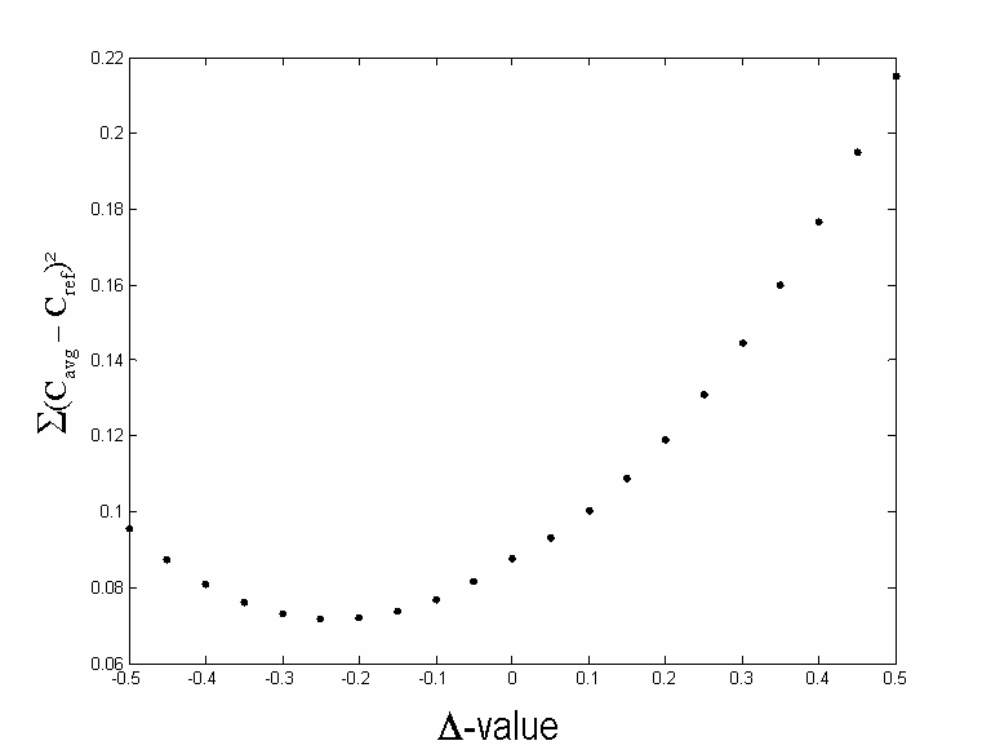


Figure 8: Evolution of the performance function (Σ ($\mathbf{C}_{avg}$ – $\mathbf{C}_{ref}$)²) in function of shift-value (Δ-value). The function reaches a minimum at a Δ-value different of zero, indicating the need for standardization.

Figure 9A shows the difference of the SNV-scaled spectra (reference – control sample) for shift values (Δ-values) of –0.5, 0.23039, 0 and +0.2, respectively. Spectral subtraction for Δ = 0 results in a first derivative kind of difference spectrum, indicative for small shifts between the Raman spectra. These shifts have a large impact on multivariate data-analysis, especially when spectra of different samples are complex and highly similar. When the evaluation function reaches its minimum (Δ = –0.23039) these 'first derivative effects' become negligible (Figure 9B). Difference spectra for Δ = -0.5 and Δ = +0.2 confirm the effect of small shifts on the Raman spectra.

After spectral standardization, the difference spectra (reference – control sample) show a flat noise band with some small artefacts (Figure 9B). The latter results from (1) different contribution of the glass vessel containing the control sample and (2) the presence of spikes in the spectrum of the control sample. Both affect the overall mean of the Raman spectra and result in small intensity differences after subtraction. It is noticed that the presence of spikes does not interfere with the minimization of the evaluation function. The difference spectrum demonstrates the performance of the spectral alignment procedure and confirms the performance of both the wavelength and intensity calibration as described.

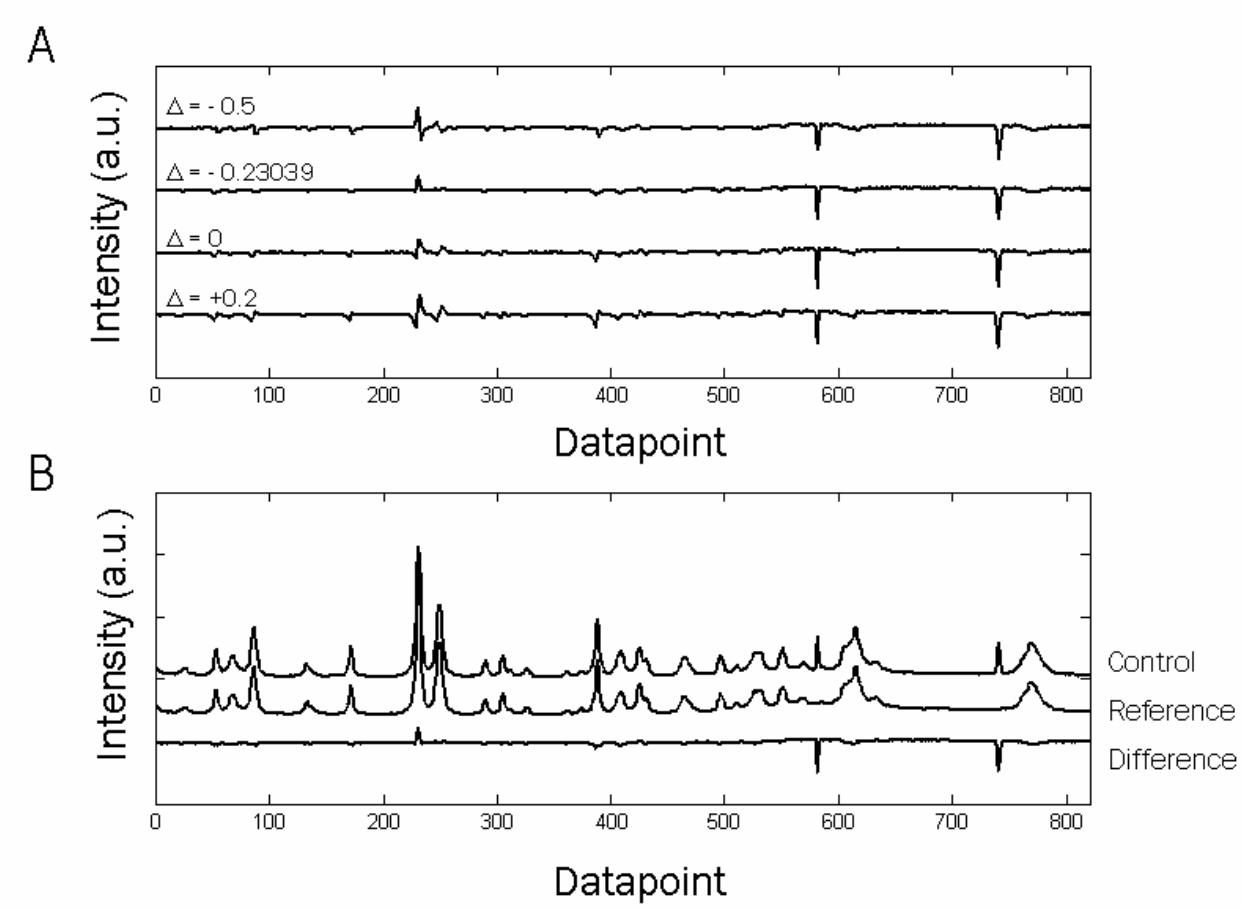


Figure 9: A. Difference spectra for multiple Δ-values (-0.5, -0.23039, 0 and + 0.2). The 'first-derivative look' indicates small shifts in laser wavelength. B. Raman spectrum of the control sample (ε-caprolactione) after shift over the optimal number of data points (Δ = -0.23039), reference spectrum and difference spectrum (Reference – Control). No shifts are present after standardization. Spikes do not interfere with the optimization.

Conclusion

This paper gives an overview of absolute wavelength (nm), Raman wavenumber (cm^{-1}) and intensity calibration and reports on a spectral standardization procedure to improve long-term stability of reference libraries. The long-term stability is extremely important when constructing reference libraries. When the Raman spectra show few discrete bands and/or are visually distinguishable, calibration artefacts are easily tolerated during multivariate analyses. However, when Raman spectra are highly similar and visual discrimination becomes hard to realize (e.g. in the case of biological samples), the impact of these artefacts will hamper the full potential of Raman spectroscopy.

It is shown that the calibration procedure, as described, is highly accurate in the determination of band positions, both on a short-term and long-term (9 months) period. Standard deviations for Raman band positions between 0.2 cm^{-1} and 0.01 cm^{-1} are reported, based on 60 independent calibrations, indicating a high precision of the applied calibration protocol. Furthermore, it is found that the protocol for intensity calibration is performing well as evaluated by the constant ratios of different response curves as obtained after independent (absolute wavelength) calibration since the relative intensities were not affected. Intensity calibration ensures the long-term stability of the reference library, whenever the optical pathway of the Raman set-up is changed (replacement of objective, notch/edge filter, detector…).
Finally, the Raman spectra were standardized towards a reference system by means of a twofold procedure. The procedure is evaluated by comparing difference spectra of randomly selected samples and proved to be efficient when compared with difference spectra of the same Raman sample before the spectral alignment procedure.

The reported approach may be applied for all kind of Raman set-ups and offers long term stability within the same laboratory. A more difficult problem is the standardization of spectra from different instruments. For instance, spectral resolution and relative intensities (due to resonance effects) are known to affect compatibility between spectral libraries. While higher spectral resolutions can easily be ‘downgraded’ by mathematical procedures, the problem of resonance is less easily tackled. This however, is out of the scope of this manuscript.

Acknowledgements

The authors are grateful to the Fund for Scientific Research - Flanders (FWO-Vlaanderen) for their financial support (Grant G.0156.02) of this research project. D.H. greatly acknowledges the institute for science and technology (IWT) for his doctoral grant and the reviewers for their constructive remarks. P.V. wishes to thank the FWO-Vlaanderen for his postdoctoral fellowship.

Reference List

1. ASTM. ASTM E1840-96. 2002, " Standard Guide for Raman Shift Standards for Spectrometer Calibration", ASTM International.

2. **Fountain, A. W., T. J. Vickers, and C. K. Mann**. 1998. Factors that Affect the Accuracy of Raman Shift Measurements on Multichannel Spectrometers. Applied Spectroscopy **52**: 462-468.

3. **Burgio, L., and R. J. Clark.** 2001. Library of FT-Raman spectra of pigments, minerals, pigment media and varnishes, and supplement to existing library of Raman spectra of pigments with visible excitation. Spectrochim Acta A Mol Biomol Spectrosc **57**:1491-1521.

4. **Shen, C.B., T. J. Vickers, and C. K. Mann**. 1992. Abscissa Error Detection and Correction in Raman Spectroscopy. Applied Spectroscopy **46**: 772-777.

5. **Tseng, C.H., J. F. Ford, C. K. Mann and T. J. Vickers**. 1993. Wavelength Calibration of a Multichannel Spectrometer. Applied Spectroscopy **47**:1808 -1813.

6. **Carter, D.A., R. W. Thompson, C. E. Taylor, and J. E. Pemberton**. 1995. Frequency/Wavelength Calibration of Multipurpose Multichannel Raman Spectrometers. Part II: Calibration Fit Considerations and Calibration Standards. Applied Spectroscopy **49**: 1561-1576.

7. **Ray, K.G. and R.L. McCreery**. 1997. Simplified Calibration of Instrument Response Function for Raman Spectrometers Based on Luminescent Intensity Standards. Applied Spectroscopy **51**: 108-115.

8. **Fryling, M., C. J. Frank, and R. L. McCreery**. 1993. Intensity Calibration and Sensitivity Comparisons for CCD/Raman Spectrometers. Applied Spectroscopy **47**:1965-1974.

9. **Maquelin, K., L. P. Choo-Smith, H. P. Endtz, H. A. Bruining, and G. J. Puppels.** 2002. Rapid identification of Candida species by confocal Raman microspectroscopy. J Clin Microbiol **40**:594-600.

10. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, and G. J. Puppels.** 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J Microbiol Methods **51**:255-271.

11. **McCreery, R.L.** 2001. Raman Spectroscopy for Chemical Analysis. Wiley Interscience, New York

12. **McCreery, R. L., A. J. Horn, J. Spencer, and E. Jefferson.** 1998. Noninvasive identification of materials inside USP vials with Raman spectroscopy and a Raman spectral library. J Pharm Sci **87**:1-8.

13. **Puppels, G. J., W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. de Mul, and J. Greve**. 1991. Description and performance of a highly sensitive confocal Raman spectrometer. J Raman Spectros **22**: 217-225.

Growth Conditions disclosed by three *Bacillus* species and how they affect the achievable taxonomic resolution of Raman Spectroscopy

D. Hutsebaut

K. Maquelin

P. De Vos

P. Vandenabeele

L. Moens

G.J. Puppels

Chapter III

Analytical Chemistry, 2004, 76, 6274 - 6281

III. Growth Conditions disclosed by three *Bacillus* species and how they affect the achievable taxonomic resolution of Raman spectroscopy

Confocal micro-Raman spectroscopy requires a minimum of sample handling and no reagents and allows fast identification of microorganisms. Since it reflects the overall molecular composition of the cells, it provides much more information than classical, microbial analyses. However, since the molecular set-up of a cell depends on culture conditions, it can be argued that this will affect the reproducibility and discrimination ability of Raman spectroscopy. We used *Bacillus cereus*, *Bacillus pumilus* and *Bacillus licheniformis* which are known to be clearly distinct from each other and each displaying important phenotypic heterogeneity, in a wide variety of culture conditions to analyze this. It is illustrated that the influence of culture conditions on the identification accuracy and taxonomic resolution of Raman spectroscopy is important though the effect on the final identification is limited within the set of strains studied. Furthermore, some conditions even allow for better discrimination than others. From a practical point of view, it is especially important that differences in culturing time (and culturing temperature) can be accommodated.

Introduction

Vibrational spectra of microorganisms consist of signal contributions of all components present in the cells and therefore reflect their overall molecular composition. Naumann *et al.*[1-3] using FT-IR spectroscopy, showed that such spectra could serve to identify a wide range of microorganisms. More recently, Raman spectroscopy was shown to have similar capabilities[4,5]. The high discriminatory power of vibrational spectroscopies is illustrated by the accurate differentiation of closely related bacterial species such as *Enterococcus faecium, Enterococcus hirae, Enterococcus faecalis, Enterococcus gallinarum, Enterococcus casseliflavus*[6]. Various publications have confirmed the applicability of these methods to discriminate between bacteria at the species level[7-9] and even strain level[2,4,10].
Simplicity of sample preparation and speed of analysis are important advantages when vibrational methods are compared to classical microbial identification tools. Furthermore, the development of highly optimized Raman spectrometers enables measurement of good quality spectra of even one single mammalian cell[11,12]. Together this has set the stage for the development of Raman spectroscopy into a powerful tool for rapid and inexpensive routine microbial analysis.

Since culture conditions influence the molecular makeup of a cell, it can be argued that they will affect the reproducibility of the Raman spectra. Earlier, Choo-Smith *et al.*[14] described the influence of some growth parameters on the heterogeneity within older colonies. The authors found that spectra of older colonies depend on the location within the colony at which the spectra were obtained, while the absence of spectral differences in 6-h colonies suggests homogeneity in terms of molecular composition. Based on these findings it was concluded that spectra of 6-h colonies are more suitable for building of spectral libraries for identification purposes using microbial (micro)colonies. Other studies focused on variation of the chemical composition of cells (*Clostridium*) in the different growth and production stages of the culture[14,15].

In the study presented here, we analyzed the effect of culture conditions on the identification accuracy of closely related *Bacillus* strains.

Materials and methods

Strains & culture conditions. The genus *Bacillus* was chosen as the model system. The taxonomy of *Bacillus* species is in general very complex. The three species selected (*Bacillus cereus*, *Bacillus licheniformis* and *Bacillus pumilus)* are known to be clearly distinct from each other, each displaying an important phenotypic internal heterogeneity. In this study a selection of 30 strains was used to represent biological variation at species level (Table 1). *Bacillus* strains obtained from the BCCM™ / LMG (Belgian Coordinated Collections of Micro-organisms, LMG Bacteria collection) culture collection were included as well as well-characterized freshly isolated strains from other studies of the laboratory of microbiology. These research strains (R-strains) were identified at the species level by 16S-rDNA sequencing and repetitive element genomic fingerprinting (Rep-PCR)[16]. Pure cultures were stored at –80°C in Microbank vessels containing 15% glycerol (cryoprotectant) until use.

The purity of the cultures was checked on solid culture medium, in identical conditions as applied for Raman analysis (see below). One colony was then recultured and subsequently analyzed according to the conditions as depicted in Figure 1. The media brain heart infusion (BHI), tryptone soya agar (TSA) and gelatin agar (GA) were obtained from Tritium Microbiologie (Veldhoven, The Netherlands).

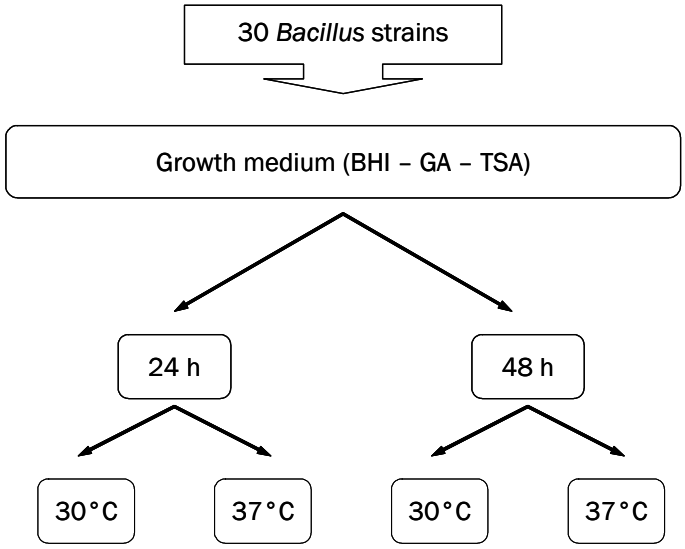


Figure 1: Overview of the different experiments performed in this study.

Table 1: An overview of the different strains used in this study. The selected strains are known to be assigned correctly at the species level and exhibit genotypic, phenotypic and ecological variation.

Name	*No.*	*Other No.*	*Source*
Bacillus cereus	LMG 12334	Logan B0358	Vomit of monkey with implicated rice associated illness
Bacillus cereus	LMG 12335	DSM 4313; NCTC 11145	Meat loaf causing a typical diarrhoeal type of food poisoning
Bacillus cereus	LMG 14742	CIP 78.3	Contaminant of pharmaceutical preparation
Bacillus cereus	LMG 17602	DSM 4282	Brain abscess
Bacillus cereus	LMG 17607	CCM 2543	Pig
Bacillus cereus	LMG 17612	Logan B0374	Faeces
Bacillus cereus	R-13605		Gelatin contaminant
Bacillus cereus	LMG 6923^T	ATCC 14579^T ; CCUG7414^T ; DSM 31^T ; IAM 12605^T ; JCM 2152^T ; NCIB 9373^T ; NRRL B-3711^T	
Bacillus cereus	R-13614		Gelatin contaminant
Bacillus cereus	LMG 8943	Moureau C13	Soil over subsurface golddeposit
Bacilus licheniformis	LMG 7560	ATCC 12759 ; IFO 12198 ; NCIB 8550 ; NRS 1415	*Manihot esculenta* , tuber
Bacilus licheniformis	LMG 7562	ATCC 11944 ; DSM 603 ; IFO 12199 ; NCIB 8874 ; NRS 1330	Septic wound
Bacilus licheniformis	LMG 12244	ATCC 102	Milk
Bacilus licheniformis	LMG 12363^T	ATCC 14580T ; CCUG 7422^T ; DSM 13^T ; JCM 2505^T ; LMG 12407^T ; NCIMB 9375^T; NRS 1264^T	
Bacilus licheniformis	LMG 17334	DVL 8400227	Aborted bovine, placenta
Bacilus licheniformis	LMG 17659	ATCC 6598 ; NRS 745	Horse bload stream
Bacilus licheniformis	LMG 6934	DSM 394	Garden soil
Bacilus licheniformis	R - 13577		Gelatin contaminant
Bacilus licheniformis	LMG 7559	ATCC 9945 ; IFO 12197 ; NCIB 8062 ; NRS 712	Flour
Bacillus pumilus	LMG 10642	Suparyono 3	*Oryza sativa* , leaves
Bacillus pumilus	LMG 10826	Suparyono 9002	*Oryza sativa* , stem
Bacillus pumilus	LMG 12259	Gibson 1036 ; Lister 2812	
Bacillus pumilus	LMG 12372	Gibson 1130	
Bacillus pumilus	LMG 12373		
Bacillus pumilus	LMG 12374	Gibson 47	
Bacillus pumilus	LMG 12376	Gibson 604	
Bacillus pumilus	LMG 3455	ATCC 15716 ; IAM 12050	Airborne isolate
Bacillus pumilus	LMG 18928^T	ATCC 7061^T ; DSM 27^T ; IAM 12469^T ; JCM 2508^T ; LMG 7132^T ; NCIB 9369^T ; NRS 272^T	
Bacillus pumilus	LMG 3586	Simpson B12	
Bacillus pumilus	R-13435		Gelatin contaminant

Sample Preparation. Smears were prepared from 24- or 48-h cultures on different media (BHI, GA, TSA) at distinct temperatures (30°C or 37°C). Biomass (60 – 100 µg) of several well-isolated colonies was harvested using a platinum loop and smeared on a CaF_2 substrate (polished, UV-grade, Crystran Ltd., Poole, United Kingdom). Prior to Raman analysis, samples were dried in a desiccator over drying beads at room temperature for 15 minutes.

Data were collected in 12 measuring sessions over a 3-month period. To eliminate the potential interference of uncontrollable day-to-day variations, planning of the experiments was randomized using a randomized block design. The randomization of the analyses was organized in such a way that discrimination between species could not be erroneously based on variations linked with day-to-day artifacts but with biochemical compositions.

Confocal Raman microspectroscopy. Raman spectroscopic measurements were performed as described previously[6,17]. Briefly, the CaF_2 substrate containing the dried smears was placed directly under the microscope of a Renishaw System 1000 Raman microspectrometer (Renishaw plc, Gloucestershire, UK). The microscope was fitted with an 80x near-infrared objective (MIR Plan 80x/0.75, Olympus). Samples were excited using 100-150 mW of 830 nm laser light from a titanium-sapphire laser (model 3900, Spectra Physics, Mountain View, CA) pumped by an argon ion laser (series 2000, Spectra Physics). The spatial resolution (measurement volume) of the setup was ~ 1.5 µm in the lateral direction and 7-8 µm along the optical axis.

Spectral analyses

Data pre-processing. The constant background signal contribution originating from optical elements in the laser light delivery pathway was subtracted from all spectra. The reference spectrum of a tungsten band lamp operated at known temperature was used to correct for the wavelength dependent signal detection efficiency of the Raman set-up[18,19]. Calibration of the wavenumber-axis was performed using the known wavelengths of the emission lines of neon and argon calibration lamps.

Spectra were collected at 10 randomly chosen locations within each smear, to compensate for potential biological variance within smears. These 10 spectra were averaged for further analysis. Each averaged spectrum represented a unique experiment, i.e. one strain cultured on a specific combination of medium, temperature and growth time. Further data pre-processing consisted of (1) calculating the first derivative of the Raman spectra by using a Savitzky-Golay algorithm with a nine-point filter and a second order polynomial for interpolation, (2) cutting out the fingerprint region (400-1800 cm^{-1}), and (3) scaling to standard normal variance (i.e. zero mean and unit variance).

Multivariate analysis. First data reduction was performed using principal component analysis; this is a well-known method for reducing the dimensionality in a data set[20,21]. The maximum number of *n*-1 principal components (PCs) was calculated (*n* being the number of spectra in the analysis), accounting for 99 - 100% of the variation in the data set.

Hierarchical cluster analysis (HCA) was performed on the *n*-1 PC scores obtained for each spectrum, using Ward's clustering algorithm and the Euclidean distance measure to generate a dendrogram. All data analysis was performed using the Matlab software package, version 6.1 (The Mathworks inc. Natick, MA) with the statistics toolbox (The Mathworks) and the PLS toolbox, version 2.0 (Eigenvector Research inc., Manson, Wash.).

For Linear discriminant analysis (LDA)[22], only PC scores accounting for more than 1% of the variance in the data set were retained. A two-sided *t*-test was used to individually select PC scores that showed the highest significance in discriminating the different microbial groups presented. The number of PC scores that was used as an input for an LDA model was kept at least two times smaller than the number of spectra in the smallest model group to prevent overfitting in the LDA model (i.e. ½**n*-1, where *n* is the number of samples in the smallest group). Generated models were evaluated by a leave-one-strain-out evaluation[23]. Briefly, the model is made using the spectra of all but one strain. All spectra of the strain left out are then predicted using the latter model. By iteratively leaving the spectra of each strain out in turn, a percentage of correct identification based upon the dataset is obtained.

Results and discussion

The aim of this study was to analyze the effect of culturing conditions on the identification accuracy of microorganisms by Raman spectroscopy. In this first phase of exploration, we investigated *B. cereus, B. pumilus and B. licheniformis* represented by 10, 11 and 9 strains, respectively, that were known to be assigned correctly at the species level and that exhibit genotypic, phenotypic, and ecological variation. Furthermore, the three species are phylogenetically well positioned versus each other: *B. pumilus* and *B. licheniformis* are close relatives within the *B. subtilis* group, while *B. cereus* is outside this group and close to *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides* and *B. weihenstephanensis*[24,25].

All strains (Table 1) were cultured several times under different conditions. From the 360 planned Raman analyses (30 strains, each with 12 culture conditions), 248 spectra were finally included in the dataset. Deviating data was omitted from further analysis and could be traced back to culturing artifacts. For instance, some culture conditions caused extensive formation of extra cellular slime (e.g. *B. pumilus* LMG 10826 on GA at 37°C, 24h; *B. licheniformis* LMG 6934 on BHI at 37°C, 24h), others did not generate sufficient biomass (e.g. *B. pumilus* LMG 12376 on GA at 37°C, 48h). In still other cases no data could be obtained because colonies stuck to the agar media and could not be harvested properly (e.g. *B. licheniformis* LMG 6934 on GA at 37°C, 48h; *B. licheniformis* LMG 17334 on TSA at 30°C, 24h; *B. pumilus* LMG 12259 on TSA at 37°C, 24h and *B. licheniformis* LMG 12363^T on GA at 37°C, 48h). This last observation is a strong argument in favor of collecting Raman spectra of microorganisms directly on solid culture medium as described previously[17]. No such difficulties were noticed for *B. cereus* strains. The heterogeneity in growth behavior of the selected *Bacillus* strains thus immediately point to some difficulties that may occur when analyzing microorganisms with Raman spectroscopy. These growth specific complications were not reported when representatives of other microorganisms such as *Enterococci*, *Streptococcus* and *Candida* [5,18] were analyzed and certainly strengthen the choice for *Bacillus* as model system to evaluate the possibilities of Raman spectroscopy for taxonomic and identification applications with significant variance in the dataset.

Identification at the species level.

When the complete dataset was subjected to HCA, there was no clear separation of the three species in separate clusters, nor were clusters formed based on culture medium, temperature or incubation time. This can most likely be explained because HCA uses all the variance in a dataset. The variance introduced by the various culture conditions is obviously larger than (or equal to) the interspecies variance. Hence, the information for the species discrimination could be distorted by the spectral variation resulting from differences in the other culture parameters. Therefore, after the comparison of different standardized culture conditions, we first analyzed subsets of the data with limited variations in the culture conditions.

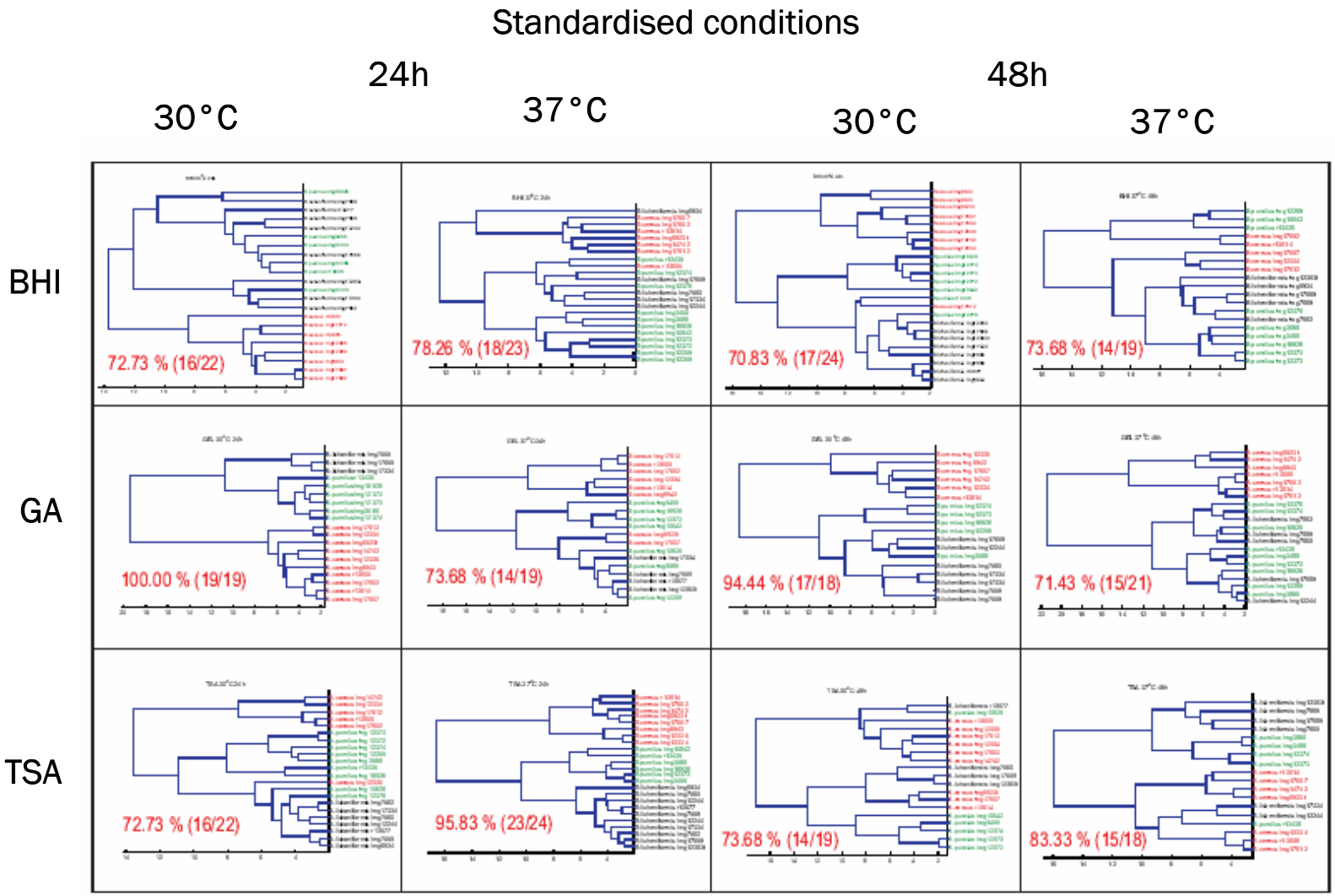


Figure 2: HCA as explained in Materials and Methods of Raman spectra of *Bacillus* strains cultured under various conditions. Also shown for each specific culture condition are the results of the leave-one-strain-out evaluation of the LDA (percentage correct prediction and number of strains correctly predicted/total number of strains between brackets). (•) *B. cereus* (red), (O) *B. pumilus* (green) and (*) *B. licheniformis* (black)

Identification performance under various standardized culture conditions. Figure 2 presents the dendrograms obtained for all standardized culture conditions. When the strains were cultured on TSA for 24 hours at 37°C or GA for 24 hours and 30°C, well-defined clusters are found at the species level. In both cases, the Raman spectra of *B. licheniformis* strains are found to be more similar to those of *B. pumilus* than to those of *B. cereus*. This observation supports the currently known phylogenetic relation of the three species based on the 16S rDNA sequence[25].

However, an optimal taxonomic resolution and the applicability in taxonomic studies seems to be dependent on the chosen culture medium, time and temperature. This effect is well illustrated by comparison of the different dendrograms in Figure 2.

The identification accuracy at species level is also evaluated by means of LDA on the 12 possible culture conditions within the dataset. The evaluation results of the various LDA models are also shown in Figure 2 and are at some points remarkably different from the HCA results. The most likely explanation for this finding is the strict limitation on the number of variables in the LDA models (½*n* - 1). Because the number of samples per species group is small, two to five variables are maximally allowed. This, in combination with the earlier mentioned heterogeneity in the cultures, could be too low for capturing all discriminating information.

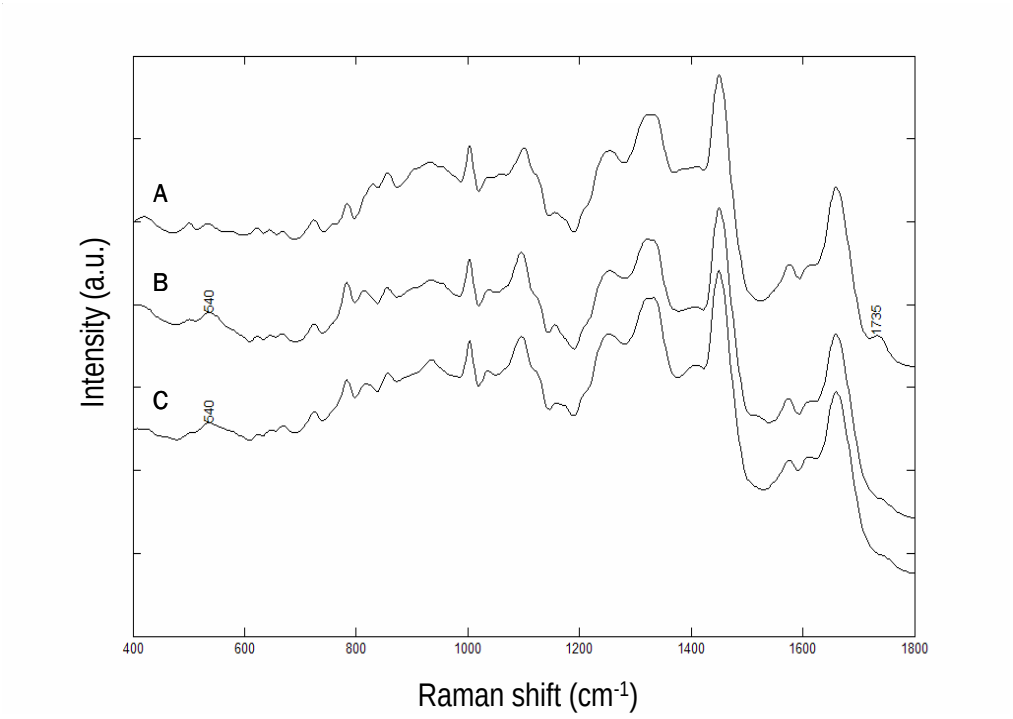


Figure 3: Representative Raman spectra of each species obtained by averaging out all Raman spectra of strains obtained after growth at BHI, 24h, 30°C. (A) *B .cereus*, (B) *B. pumilus* and (C) *B. licheniformis*. The spectra have been displaced vertically on the intensity axis; a.u. arbitrary units. Discriminative peaks are labelled.

The superior identification accuracies for some growth conditions can be related to differences in the Raman spectra. Figures 3 and 4 show representative Raman spectra of each species, grown under "less-optimal" (BHI, 24h, 30°C; 72.73% accuracy) and "near-optimal" (GA, 24h, 30°C; 100% accuracy) culture conditions, respectively. When the strains are grown on BHI, 24h, 30°C the Raman spectra of *B. licheniformis* and *B. pumilus* are very similar and the delineation of both species becomes very difficult. *B. cereus* strains are discriminated from the other two species by Raman bands in the spectral region near 540 cm^{-1} and 1735 cm^{-1}. These bands can be assigned to COC glycosidic ring deformation and >C=O ester stretching[4].

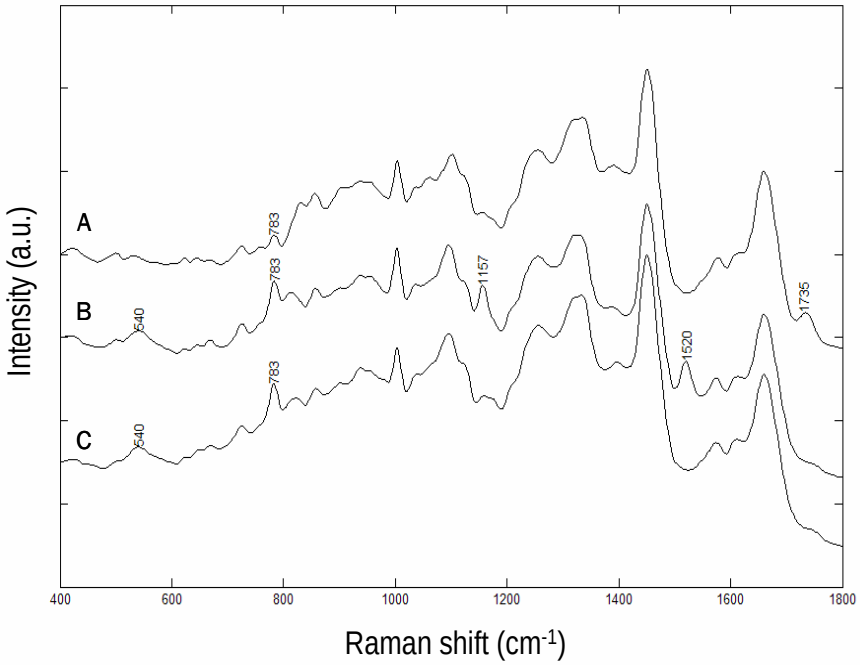


Figure 4: Representative Raman spectra of each species obtained by averaging out all Raman spectra of strains obtained grown on GA, 24h, 30°C. (A) *B .cereus,* (B) *B. pumilus* and (C) *B. licheniformis.* The spectra have been displaced vertically on the intensity axis; a.u. arbitrary units. Discriminative peaks are labelled.

In earlier studies, it was noticed that signal contributions from growth phase-related storage compounds, such as poly(β-hydroxy-buteric acid) granules and dipicolinic acid from endospores, could be detected in FT-IR spectra of *Bacillus* strains[26]. In this study, however, we did not find spectral evidence for these compounds nor were spores found during microscopical examination (results not shown). When the strains are grown on GA, 24h, 30°C the Raman spectra of *B. pumilus* and *B. licheniformis* are clearly distinct from each other by Raman bands near 783 cm^{-1} (uracil), 1157 cm^{-1} and 1520 cm^{-1}. The latter two peaks, in combination with the one at 1004 cm^{-1}, are indicative of an increased carotenoid level.

Hence, the differences in taxonomic resolution under different culture conditions can be related to differences in the biochemical composition of the cells. This indicates that the taxonomical resolution could be increased by selecting the right culture conditions.

Effect of varying a single culturing parameter. In a next step, the robustness of bacterial identification was explored by changing one culturing parameter. Therefore, the complete dataset was split into different subgroups, in which only one parameter varied (e.g. growth time), while the two remaining parameters (medium and temperature) were constant. For each subgroup, an LDA model was calculated. The results are summarized in Table 2. Clearly, changing only one cultural parameter does not affect LDA's ability to discriminate between the three *Bacillus* species importantly. Most of the growth conditions give rise to robust LDA-models with, relative to the large cultural variation, good identification abilities (> 80 %) as shown by leave-one-strain out evaluation. Here the LDA models seem to perform better than with standardized culture conditions. Due to the higher number of samples per species group, there are more variables allowed in the models (½**n*-1), hence more variance can be accounted for in the models.

Table 2: Overview of the identification results obtained at the species level using linear discriminant analysis expressed in percentage (%) correct identification. The ratio correct classified/total number of cases between parenthes.

Altering one culture parameter	
Variable Culturing Temperature (30° - 37°C)	
BHI - 24u	80.00 % (36/45)
BHI - 48u	83.72 % (36/43)
GA - 24u	86.84 % (33/38)
GA - 48u	71.79 % (28/39)
TSA - 24 u	84.78 % (39/46)
TSA - 48 u	89.19 % (33/37)
Variable Culturing Time (24 – 48h)	
BHI - 30°C	80.43 % (37/46)
BHI - 37°C	90.48 % (38/42)
GA - 30°C	97.30 % (36/37)
GA - 37°C	75.00 % (30/40)
TSA - 30°C	82.93 % (34/41)
TSA - 37°C	97.62 % (41/42)
Variable Medium (GA – TSA –BHI)	
24h - 30°C	80.95 % (51/63)
24h - 37°C	93.94 % (62/66)
48h - 30°C	90.16 % (55/61)
48h - 37°C	79.31 % (46/58)

Of three key culture parameters, the robustness towards incubation time is interesting to consider from a practical point of view, since time is often difficult to control completely. Small differences (5 – 10 min) can easily occur during incubation or Raman analysis. When large incubation time differences (24h) were introduced in the dataset, up to 97 % of the strains were still correctly identified at species level as evaluated by leave-one-strain-out analysis. The dendrograms underlining the robustness towards time variations for all six combinations of medium and temperature are shown in Figure 5. As shown, certain growth conditions (GA, 30°C) are more robust towards large time variations, since they do not introduce variances larger than the inter-species variances.

Effect of varying several growth parameters. Table 3 contains the LDA results in case two parameters were altered. The best identification (92.74 %) for this set is obtained when subgroups contain all spectra of colonies cultured at constant temperature of 30°C, irrespective of the medium and the incubation time. Also, when altering all parameters but culture medium, TSA and BHI result in slightly better identification results than GA. This demonstrates convincingly that species-specific information can be extracted from Raman spectra, even when changing the culture parameters.

Table 3: Overview of the identification results obtained at the species level using linear discriminant analysis expressed in percentage correct identification. The ratio correct classified/total number of cases between parentheses

Altering two culture parameters	
Constant medium (30 -37 °C / 24 - 48 h)	
BHI	89.77 % (79/88)
GA	80.52 % (62/77)
TSA	91.57 % (76/83)
Constant culturing temperature (24 – 48h, BHI – GA – TSA)	
30 °C	92.74 % (115/124)
37 °C	90.32 % (112/124)
Constant culturing time (30 – 37 °C / BHI-GA-TSA)	
24 h	90.70 % (117/129)
48 h	89.08 % (106/119)

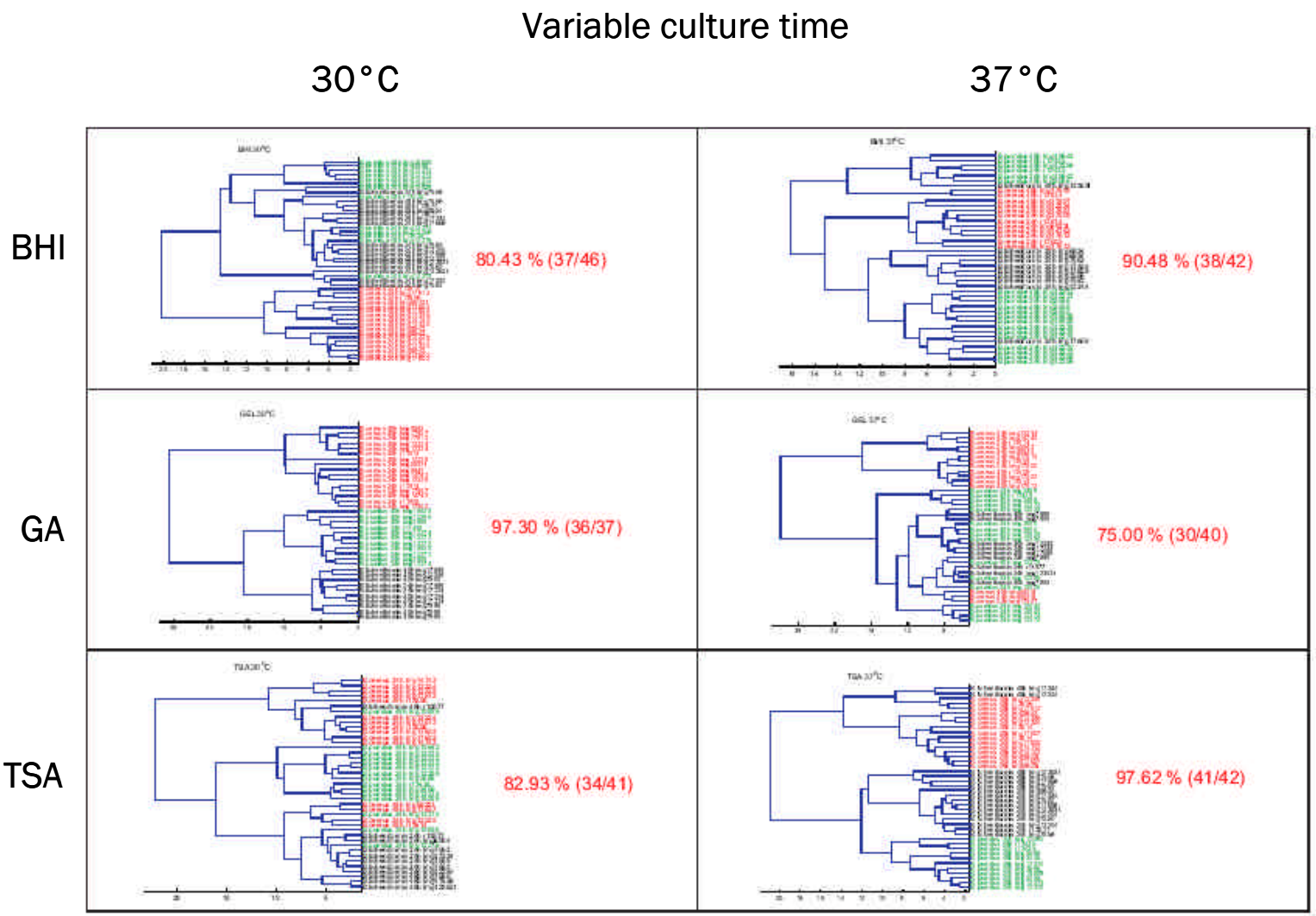


Figure 5: HCA as explained in Materials and Methods of Raman spectra of *Bacillus* strains cultured during 24 or 48h. Also shown for the different sets of culture conditions condition is the leave-one-strain-out evaluation of the LDA (percentage correct prediction and number of strains correctly predicted/total number of strains between brackets). (•) *B. cereus* (red), (O) *B. pumilus* (green) and (*) *B. licheniformis* (black)

When none of the parameters considered in this study (medium composition, culture temperature and incubation time) were kept constant, i.e. data of all experiments were combined, LDA was still able to discriminate successfully between *B. cereus*, *B. licheniformis* and *B. pumilus*. In this situation, 92.34 % (229/248) of the strains were then recognized correctly at the species level as evaluated by leave-one-strain-out evaluation. Figure 6 visualizes the group separation between the three species by means of the discriminant functions. *B. cereus* strains are clearly distinct from *B. licheniformis* and *B. pumilus*, indicating that the biochemical composition from the latter two species is more similar. This is in correspondence with the phylogenetic relationship of the three species. Using scatter plots to visualize the taxonomic relationships could serve as an alternative to the unsupervised HCA, when the identity of a subset in the data is known.

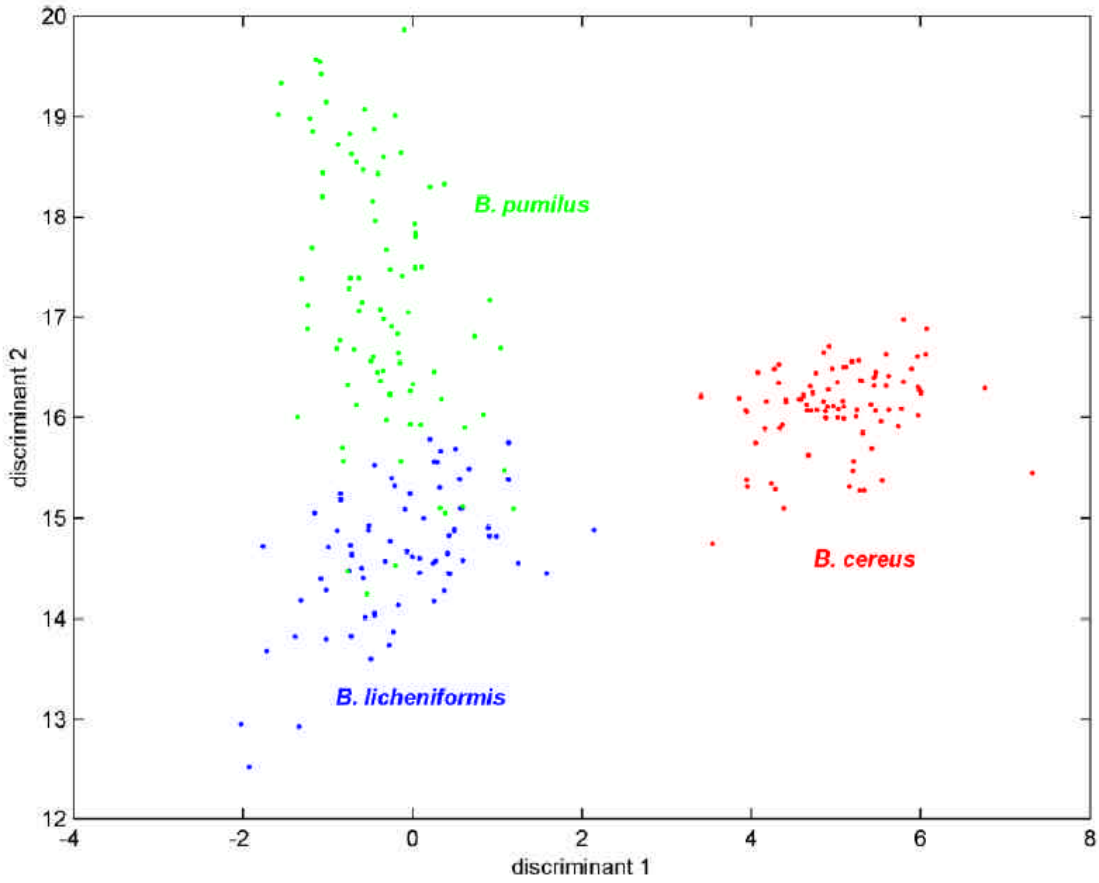


Figure 6: Group separation between the three *Bacillus* species by means of linear discriminant analysis. As a whole, *B. cereus* strains are better separated than *B. licheniformis* and *B. pumilus* strains, indicating that the spectra from the last two species are more similar to each other than to those of *B. cereus*. This is not in contradiction with the more separate phylogenetic position of *B. cereus*. (•) *B. cereus* (red), (O) *B. pumilus* (green) and (*) *B. licheniformis* (blue)

Conclusion and prospects

We demonstrated that Raman spectroscopy can be a robust method for the identification of microorganisms. Culturing *Bacillus* strains belonging to three well distinct species under different conditions, still allowed their discrimination using chemometric methods. The identification at species level as exhibited here was less sensitive to culture conditions, while grouping of the strains relying only on HCA was sometimes less in correspondence with the 16S-rDNA sequence analysis. The results bode well for the routine practical situation of clinical diagnostics, as they indicate, that the standardization of the Raman identification procedure could be easily implemented and maintained. From a practical point of view it is especially important that differences in culturing time (and culturing temperature) can be accommodated. Therefore, in addition to the fact that vibrational spectroscopy only requires a minimum of sample handling, and no reagents, the level of standardization of culturing conditions is also not very extreme.

It has to be noted that the combination of culture parameters selected in this work may not be the most optimal ones to build a reliable database. Indeed, speed and comfort of microbial analysis are imposing factors to determine the choice of incubation time, while a proper composition of the medium may facilitate the harvesting of bacterial colonies. Nevertheless, it is clear that appropriate growth conditions in general affect positively the implementation possibilities of Raman spectroscopy in microbial analysis.

More research is currently performed, to further improve and determine maximally achievable taxonomic resolution of Raman spectroscopy, by studying other closely related and well-characterized *Bacillus* species.

Acknowledgements

The authors are grateful to the Fund for Scientific Research - Flanders (FWO-Vlaanderen) for their financial support (Grants G.0156.02 and G.0343.03) of this research project. D.H. greatly acknowledges the institute for science and technology (IWT) for his doctoral grant. P.V. wishes to thank the FWO-Vlaanderen for his postdoctoral fellowship.

References

1. D. Naumann, D. Helm, H. Labischinski, *Nature*, 1991, **351**, 8-82.

2. D.Helm, H. Labischinski, D. Naumann, *J Microbio. Method*, 1991, **14**, 127-142.

3. D.Helm, H. Labischinski, G.Schallehn, D. Naumann, *J Gen Microbiol*, 1991, **137**, 69–79.

4. K. Maquelin, C. Kirschner, L. Choo-Smith, N. van den Braak, H. Endtz, D. Naumann, G.J. Puppels. *J Microbiol Methods*, 2002, **51**, 255-271.

5. K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N.A. Ngo-Thi, T. van Vreeswijk, M. Stämmler, H.P. Endtz, H.A. Bruining, D. Naumann, G.J. Puppels, *J Clin Microbiol*, 2003, **41**, 324–329.

6. C. Kirschner, K. Maquelin, P. Pina, N.A. Ngo Thi, L.-P. Choo-Smith, G.D. Sockalingum, C. Sandt, D. Ami, F. Orsini, S.M. Doglia, P. Allouch, M. Manfait, G.J. Puppels, D. Naumann, *J Clin Microbiol*, 2001, **39**, 1763-1770.

7. S.H. Beattie, C. Holt, D. Hirst, A.G. Williams, *FEMS Microbiol. Lett.*, 1998, **164**, 201-206.

8. H. Oberreuter, H. Seiler, S. Scherer, *Int J Syst Evol Microbiol.*, 2002, **52**, 91-100.

9. C. Sandt, G. D.Sockalingum, D. Aubert, H. Lepan, C. Lepouse, M. Jaussaud, A. Leon, J. M. Pinon, M. Manfait, D. Toubas, *J Clin Microbiol*, 2003, **41**, 954-959.

10. M.C. Curk,; F. Peladan, J.C. Hubert, *FEMS Microbiol Lett*, 1994, **123**, 241-248.

11. G.J. Puppels, F.F.M. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D.J. Arndt-Jovin, T.M. Jovin, *Nature*, 1990, **347**, 301-303.

12. G.J. Puppels, In *Fluorescent and luminescent probes for biological activity*, 2nd ed.; W.T. Mason, Ed.; Academic Press: London, 1999, p.377-406.

13. L.P. Choo-Smith, K. Maquelin, T. van Vreeswijk, H.A. Bruining, G.J. Puppels, N.A. Ngo Thi, C. Kirschner, D. Nauman, D. Ami, A.M Villa,. F. Orsini, S.M. Doglia, H. Lamfarraj, G.D. Sockalingum, M. Manfait, P. Allouch, H.P Endtz,. *App and Environ Microbiol*, 2001, ***67***, 1461-1469.

14. K. C. Schuster, I. Reese, E. Urlaub, J. R. Gapes, B. Lendl, *Anal Chem*, 2000, **72**, 5529-5534.

15. M. Grube, J.R. Grapes, K.C. Schuster, *Anal Chim Acta*, 2002, 471, 127-133.

16. E. De Clerck, T. Vanhoutte, T. Hebb, J. Geerinck, J. Devos, P. De Vos, *Appl Environ Microb*, 2004, **70**, 3664 – 3672

17. K. Maquelin, L.P. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H.A. Bruining, G.J. Puppels, *Anal Chem*, 2000, **72**, 12-19.

18. G. J. Puppels, W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. de Mul, J. Greve, *J Raman Spectrosc*, 1991, **22**, 217-225.

19. R. Wolthuis, T. C. Bakker Schut, P. J. Caspers, H. P. J. Buschman, T. J. Römer, H.A. Bruining, G. J.Puppels, In *Fluorescent and luminescent probes for biological activity*, 2nd ed.; W.T. Mason,; Ed.; Academic Press: London, 1999; p. 431-455.

20. I.T. Jollife, *Principal component analysis*. Springer Verslag: New York, 1986.

21. E. M. Timmins, S. A. Howell, B.K. Alsberg, W.C. Noble, R. Goodacre, *J Clin Microbiol*, 1998, **36**, 367-364.

22. T. C. Bakker Schut, M. J. H. Witjes, H.J.C.M. Sterenborg, O.C. Speelman, J.L.N. Roodenburg, E.T. Marple, H.A. Bruining, G.J. Puppels, *Anal Chem*, 2000, ***72***, 6010-6018.

23. K. Maquelin, L.-P. Choo-Smith, H.P. Endtz, H.A. Bruining, G.J. Puppels, *J Clin Microbiol*, 2002, **40**, 594-600.

24. A. Cherif, L. Brusetti, S. Borin, A. Rizzi, A. Boudabous, H. Khyami-Horani, D. Daffonchio, *J Appl Microbiol*, 2003, **94**, 1108-1119.

25. K. Goto, T. Omura, Y. Hara, Y. Sadaie, *J Gen Appl Microbiol*, 2000**, 46**, 1 – 8.

Raman microspectroscopy as an identification tool within the phylogenetically homogeneous *'Bacillus subtilis'*-group

Didier Hutsebaut

Joachim Vandroemme

Jeroen Heyrman

Peter Dawyndt

Peter Vandenabeele

Luc Moens

Paul De Vos

Chapter IV

Systematic and Applied Microbiology, 2005, submitted

IV. Raman microspectroscopy as an identification tool within the phylogenetically homogeneous '*Bacillus subtilis*'-group.

Vibrational methods have multiple advantages compared to more classic, chemotaxonomic and even molecular microbial tools for the identification of bacteria. Nevertheless, their definite breakthrough in diagnostic microbiology laboratories is determined by their identification potential. This paper reports on the profound evaluation of Raman spectroscopy to identify closely related species by means of 69 *Bacillus* strains that are assigned or closely related to the phylogenetically homogeneous '*Bacillus subtilis*'-group (*sensu stricto*). These strains were chosen to represent biological variation within the selected species and to create a realistic view on the possibilities of this technique

The evaluation resulted in 49/54 correct identifications at the species level for intern and 15/19 for extern testing. The correct identification of strains that were not represented in the training set supports the potential as a general identification tool. Considering the vague borderline between the species studied, Raman spectroscopy can be regarded as a promising application for identifications at the species level.

Introduction

Vibrational spectroscopic methods are currently studied and developed as powerful new techniques for identification of micro-organisms. The multiple advantages of these methods, compared to more classic, chemotaxonomic and even molecular tools, make them of great interest for modern diagnostic laboratories. A wide range of micro-organisms can be discriminated at the species level by Fourier Transform Infrared (FT-IR) spectroscopy[3,24] and Raman spectroscopy[13,17]. Both techniques provide biochemical information regarding the complete molecular composition of the cells (carbohydrates, fatty acids, proteins, RNA/DNA)[21] and hence combine discriminatory abilities of various phenotypic characterizations (e.g. whole-cell fatty acid composition (FAME), whole-cell protein polyamide gel electrophoresis (SDS-PAGE), metabolic tests (API, Biolog, Vitek), differences in pigmentation, etc.).

Raman spectroscopy combines high information content, speed of analysis, minimal sample preparation and is, at the same time, less labour intensive than classic methods. Together, this has set the stage for the development of Raman spectroscopy into a powerful tool for rapid and inexpensive microbial analysis. However, before this vibrational technique can be used as a more general routine screening tool, the limits of the taxonomic resolution of the method have to be evaluated for various well-chosen groups of Gram positive and Gram negative bacteria. Indeed, since it concerns a chemotaxonomic method, the reproducibility and the taxonomic resolution is affected by the molecular composition of the cell, which in turn depends on variations of growth conditions (medium composition, incubation time and temperature, etc.)[11].

To evaluate the feasibility of Raman spectroscopy as an identification tool, the genus *Bacillus* was selected. This genus is characterized by aerobic or facultative anaerobic, rod-shaped, Gram-positive spore-forming bacteria that have quite different phenotypes[4]. During the last decades, this genus has been split at the generic level[2, 7,10, 22, 33, 38, 41] while new relatives were discovered[23, 26, 32, 34, 35, 37, 40] and therefore this group of bacteria now comprises more than fifteen genera. Due to new species descriptions, the authentic genus *Bacillus* still contains more than 100 species. Among them, a group of eight species *Bacillus axarquiensis*, *Bacillus malacitensis*, *Bacillus mojavensis*, *Bacillus vallismortis*, *Bacillus amyloliquefaciens*, *Bacillus atrophaeus*, *Bacillus velezensis* and *Bacillus subtilis* (with two subsp. *subtilis* and subsp. *spizizenii*)[19] exists that is homogenous at the phenotypic and phylogenetic level. These species represent the '*Bacillus subtilis*'-group.

Three additional species, *Bacillus licheniformis*, *Bacillus sonorensis* and *Bacillus pumilus* are also closely related, yet they are phylogenetically clearly separated from the others. The majority of phenotypic surveys report that members of the '*B. subtilis*'-group are very difficult to discriminate. For example, Roberts *et al.*[28,29] demonstrated that, for the traits tested, fatty acid composition was the only phenotypic character that distinguished *B. mojavensis* and *B. vallismortis* from one another or from *B. subtilis*. Further, phenotypic discrimination between *B. atrophaeus* and *B. subtilis* was limited to pigmentation[20], fatty acid composition, and a positive oxidase test[29]. In another extensive study, *B. amyloliquefaciens* could only be distinguished from *B. subtilis* by three of the 75 phenotypic traits tested 15. Recently, Ruiz Garcia *et al.*[42] tested the type strains of all species closely related to *B. subtilis* on 112 characteristics. These authors found clear differences between the type strains of all species, e.g. four of the traits enabled the differentiation of *B. amyloliquefaciens* and *B. subtilis*. However, since this study included a single strain per species it did not consider the variation within the species and more extensive studies are needed to verify whether these differences are stable identifications for species differentiations. Palmisano *et al.*[25] described *B. sonorensis* as a novel species and could phenotypically only distinguish it from *B. licheniformis* by salt tolerance and pigmentation.

In modern bacterial classification, 16S rDNA sequence analysis is standardly used to allocate strains at various taxonomic levels. However, 16S rDNA sequences often show limited or no variation if closely related taxa are considered[8]. This is also the case for certain members of the '*B. subtilis*'-group. Indeed, the 16S rDNA sequences of the type strains of *B. subtilis* (subsp. *subtilis* and subsp. *spizizenii*), *B. amyloliquefaciens*, *B. atrophaeus*, *B. mojavensis*, *B. vallismortis* and the recently described *B. velezensis*[30] and *B. malacitensis*[31], show similarities ranging from 99.2-99.8% (pairwise similarities based on the UPGMA algorithm of 16S rDNA sequences obtained from the GenBank/EMBL/DDBJ database). *B. axarquiensis* shows 97.4% similarity with *B. mojavensis* and *B. subtilis* subsp. *spizizenii*[31]. In comparison, when the sequences of six different strains designated to *B. subtilis* are compared using the same method, similarities between 99.4 - 99.9% are obtained. This clearly reflects the limitations of 16S rDNA sequence analysis to discriminate members of the *'B. subtilis'*-group. In addition, the type strains of *B. licheniformis* and *B. sonorensis* show a 16S rDNA sequence similarity of 99.6%. A comparative study of the housekeeping genes *gyrA*, *polC* and *rpoB*[28] resulted in a combined clustering that clearly allows the discrimination of *B. mojavensis*, *B. amyloliquefaciens*, *B. atrophaeus*, *B. subtilis* and *B. licheniformis* which indicates the potential power of this sequencing approach.

Based on the outlined taxonomic situation, the '*B. subtilis*'-group is considered as an excellent model in the evaluation of Raman spectroscopy as a technique to identify closely related species. The current classification of the strains which were used, is well established due to polyphasic characterization. The study aims at testing the feasibility of identification by Raman spectroscopic analysis against a (limited) database. In order to evaluate the identification results in a realistic way, a blind approach is used.

Materials and methods

Strains and culture conditions. The 69 *Bacillus* strains used and their sources are listed in Table 1. The strains were selected to represent biological variation of the species. Pure subcultures were temporarily stored at –20 °C in Microbank vessels under 15% v/v glycerol cryoprotection before use.

Strains were grown on Brain Heart Infusion (BHI) that was freshly prepared (29.6 g BHI (BD) and 16,0 g Bacto™Agar (BD) were mixed in 800 ml distilled water, stirred for 15 min, autoclaved for 20 min at 121 °C and poured into Petri dishes) to comprise the effect of potential variation in medium preparation on the Raman spectra. Before use, Petri dishes were incubated overnight at 28 °C, checked for contamination and stored in a closed plastic bag at room temperature for maximum 1 week.

Prior to Raman analysis, bacteria were grown overnight on BHI at 37 °C. A single well-isolated colony was picked and streaked again on BHI agar plate using the third quadrant method. The Petri dishes were then incubated at 37 °C (+/- 1 °C) for exactly 7 hours.

Sample preparation. Confluently grown cells were harvested from the third quadrant with a disposable plastic inoculation loop and manually spotted on a CaF_2 substrate (polished, UV-grade, Crystran, Poole, United Kingdom). If no confluent growth was present in the third quadrant, biomass was taken from the second quadrant. The spotted biomass was allowed to dry at room temperature for 20 minutes in a desiccator containing drying beads.

Planning of the experiments. To accommodate potential sample heterogeneity, four spectra were collected at randomly chosen locations within the spotted biomass, with a signal collection time of 60s, each. An ‘experiment’ is considered as the mean of four Raman measurements of a strain grown on a freshly prepared culture plate (BHI) for 7 hours at 37 °C on a particular day. At least 3 replicate experiments were recorded for each strain on different days, using different medium batches. To eliminate the potential interference of unpredictable day-to-day variations, planning of the experiments was completely randomized. Data were collected over a 5-month period encompassing 19 measuring sessions. Each measuring session consisted of 10 – 20 experiments. The complete set contained 219 experiments.

Table 1: Strains used, with their strain number and origin.

Name	No.	Other No.	Source
B.amyloliquefaciens	LMG 12325	ATCC 23844	Pfizer Fabrizyme HC Concentrate
B. amyloliquefaciens	LMG 12329	ATCC 23843	Standard Brand's Diastafor alpha-amylase concentrate
B. amyloliquefaciens	LMG 12330	ATCC 23845;Campbell strain N;IAM 1522;LMG 12328;Logan B0171;Logan B0176;Logan B0253;Logan B0255	
B. amyloliquefaciens	LMG 12331		
B. amyloliquefaciens	LMG 12384	Carr M67;Logan B0500	Cocoa
B. amyloliquefaciens	LMG 12385	Hartman 1A35;Logan B0620	
B. amyloliquefaciens	LMG 17599	Logan B0425;strain F938/78	Insect bite
B. amyloliquefaciens	LMG 17600	Logan B0568;strain B(ii)	Anti-corrosion coating
B. amyloliquefaciens	LMG 17601	Keates GO/1;Logan B0781	Sterility test
B. amyloliquefaciens	LMG 9814^T	ATCC 23350;CIP 103265;DSM 7;IFO 15535;LMG 12234;NCIMB 12077;NRRL B-14393;VTT E-80124	Soil
B. atrophaeus	LMG 16797^T	ATCC 49337;DSM 7624;IFO 15539;JCM 9070;NCIMB 12899;NRRL NRS-213	Soil
B. atrophaeus	LMG 17795	DSM 5551;IFO 15407;NRRL NRS-253;NRS 253	Air
B. atrophaeus	LMG 17796	NRRL NRS-748;NRS 748	Decomposed wheat
B. atrophaeus	LMG 8199	DSM 2277;NCIB 8649;NCTC 10073;NRC B-467	
B. licheniformis	LMG 12247	Gibson 1153;Logan B0090	
B. licheniformis	LMG 12248	Gibson 1160;Logan B0091	
B. licheniformis	LMG 12360	Gibson 307;Logan B0242	
B. licheniformis	LMG 12362	Gibson 1174;Logan B0244	
B. licheniformis	LMG 12363^T	ATCC 14580;CCM 2145;CCUG 7422;CECT 20;CIP 52.71;DSM 13;IFO 12200;JCM 2505;KCTC 1753;KCTC 1918;LMG 12407;LMG 6933;NCFB 1772;NCIMB 9375;NCTC 10341;NRS 1264;	
B. licheniformis	LMG 17334	DVL 8400227	Aborted bovine, placenta
B. licheniformis	LMG 17337	DVL 9315375	Aborted bovine foetus, lung

B. licheniformis	LMG 17339	DVL 9410670-3	Silage
B. licheniformis	LMG 17340	DVL 9410670-4	Potato pulp for cattle feeding
B. licheniformis	LMG 17649		
B. licheniformis	LMG 17652	Logan B0755;NIH 21;NRS 551	Clinical pathogen
B. licheniformis	LMG 17654		
B. licheniformis	LMG 17656		
B. licheniformis	LMG 17657	Logan B0451;strain SM11	Salt mash
B. licheniformis	LMG 17658	Logan B0698;strain C/T6	Yoghurt
B. licheniformis	LMG 17659	ATCC 6598;Logan B0758;NRS 745	Horse, blood stream
B. licheniformis	LMG 6934	Claus IV;DSM 394	Garden soil
B. licheniformis	LMG 7559	ATCC 9945;BU 169;CCM 2206;IFO 12197;NCFB 793;NCIB 8062;NRS 712;OUT 8369	Flour
B. licheniformis	LMG 7560	ATCC 11560;ATCC 12759;Damodaran P-8;Hankey B133;IFO 12198;NCIB 8549;NCIB 8550;NRS 1415	*Manihot esculenta*, tuber
B. licheniformis	LMG 7630	Bonde 69	Seawater
B. licheniformis	LMG 7634	Bonde 127	Chinchilla, faeces
B. mojavensis	LMG 17797^{T}	ATCC 51516;DSM 9205;IFO 15718;NRRL B-14698	Soil
B. mojavensis	LMG 17798		
B. pumilus	LMG 10826		
B. pumilus	LMG 12258	Logan B0052;UB 8/1/6/8	
B. pumilus	LMG 12259	Gibson 1036;Lister 2812;LMG 12373;Logan B0102;Logan B0235	
B. pumilus	LMG 12372	Gibson 1130;Logan B0234	
B. pumilus	LMG 12373	Gibson 1036;Lister 2812;LMG 12373;Logan B0102;Logan B0235	
B. pumilus	LMG 12374	Gibson 47;Logan B0237	
B. pumilus	LMG 12375	Gibson 67;Logan B0238	
B. pumilus	LMG 12376	Gibson 604;Logan B0239	
B. pumilus	LMG 18517	ATCC 14884;CCM 2218;CIP 76.18;DSM 361;IFO 12102;NCIB 8982;NCTC 8241	
B. pumilus	LMG 18928^{T}	ATCC 7061;CCEB 639;CCM 2144;CECT 29;CIP 52.67;CN 2200;DSM 27;IAM 12469;IFO 12092;JCM 2508;LMG 7132;Logan B0019;NCFB 1766;NCIB 9369;NCTC 10337;NRS 272;OUT 8376	

B. pumilus	LMG 3455	ATCC 15716;IAM 12050;RH 2185;Tulecke A1/13TI	Airborne isolate
B. pumilus	LMG 8942	Moureau C12	Soil over subsurface golddeposit
B. sonorensis	LMG 21636^T		
B. subtilis	LMG 12260	Logan B0044	
B. subtilis	LMG 12261	Logan B0046;NCTC 2116;UB 8/1/6/2	
B. subtilis	LMG 12262	ATCC 9943;Ford 2;NCTC 2587	
B. subtilis	LMG 12263	Gibson 1111;Logan B0093	
B. subtilis	LMG 12264	Gibson 1007;Lister 2592;Logan B0098	
B. subtilis	LMG 12379	ATCC 7003;Logan B0223;NRS 730	Soil
B. subtilis	LMG 12380	Gibson 66;Logan B0226	
B. subtilis	LMG 12381	Gibson 1156;Logan B0228	
B. subtilis	LMG 17722		
B. subtilis	LMG 17725		
B. subtilis	LMG 17726		
B. subtilis	LMG 17727		
B. subtilis	LMG 19154		
B. subtilis	LMG 19155	NRRL B-14820	
B. subtilis	LMG 7135^T	ATCC 6051;BGSC 3A1;CCEB 642;CCM 2216;CCUG 163B;CECT 39;CIP 52.65;CN 5638;CNCTC Bs 18/66;DSM 10;FIRDI 255;IAM 12118;IEM BS1866;IFO 12210;JCM 1465;NCFB 1769;NCIMB 3610;NCTC 3610;NRS 1315;NRS 744;USCC 1414	
B. subtilis	LMG 8197	ATCC 6633;CCM 1999;DSM 347;IFO 13720;IFO 3134;JCM 2499;NCIB 8054;NCIMB 8566;NCTC 10400;NRS 231	
B. vallismortis	LMG 17800	Cohan DV7-C-1;NRRL B-14894	Arroyo in alluvial fan with desert holly bushes
B. vallismortis	LMG 18725^T	DSM 11031;NRRL B-14890;	Soil from sand dune with mesquite tree

Raman microspectroscopy. Raman spectroscopic measurements were performed similarly to Maquelin *et al.*[16]. In brief, the CaF_2 substrate with the dried smears was placed directly under the microscope of a Kaiser System Hololab 5000R modular Raman microspectrometer (f/1.8) (KOSI, Ecully, France). The microscope was fitted with a 100x objective (PL Fluotar L, 100x/0.75, W.D. 4.7 mm, Leica). Samples were excited using 45 - 50 mW of 785 nm laser light from a diode laser (Toptica Photonics AG, Martinsried/Munich, Germany). The scattered light is guided to the spectrograph by means of a confocal, 15 µm aperture collection fiber. A back illuminated deep depletion CCD camera (Andor, Belfast, Northern Ireland) was used for the detection of the scattered light. Raman signal was collected in the spectral interval of 150 cm^{-1} – 3100 cm^{-1} with a spectral resolution of 4 cm^{-1}. Calibration of the absolute wavelength-axis was performed using the known wavelengths of the atomic lines from neon[12]. The reference spectrum of a tungsten band lamp (Optronic laboratories, Orlando, USA) operated at known fixed current was used to correct for the wavelength dependent signal detection efficiency of the Raman set-up[12,27].

Analysis of Raman data

Data pre-processing. Dark signal and constant optical background arising from optical components in the Raman set-up were subtracted from each spectrum. All four spectra collected from the same strain were standardized towards a reference spectrum[12] and the first derivatives were calculated to minimize the influence of background signal caused by slight sample fluorescence. Consequently, the data were averaged for further analysis and treated as one single spectrum (one experiment = one average spectrum). Further data preprocessing consisted of cutting out the fingerprint region (370 cm^{-1} – 1750 cm^{-1}) and scaling to zero mean and unit variance (SNV). The resulting data for each strain were used in the multivariate analysis.

Multivariate data analysis. After autoscaling, data reduction was first performed using principal component analysis (PCA) which is performed with the PLS toolbox (Eigenvector Research Inc., Manson, WA, USA) for Matlab software (Mathworks Inc., Natick, MA, USA). PCA is a well-known method to reduce the dimensionality of the data[5,36,36]. The maximum number of n-1 principal components (PCs) was calculated (n = the number of spectra in the analysis), and those PC's accounting for 99.9 % of the variation in the data set were retained.

An *F*-test was used to individually select the most significant PC's ($\alpha < 0.05$) allowing discrimination among species. The selected scores served as input for Linear Discriminant Analysis[36] and *j - 1* canonical (*j* = number of different species) variates were retained in the subsequent identification. Linear Discriminant Analysis (LDA) was used only to calculate the canonical variates (CV's) minimizing the within-species variance, while maximizing the variation between different species. In the following, the CV's thus obtained will be referred to as '*pattern*'. Since the LDA model was trained at the species level, strain specific information was neglected.

Description of dataset and identification. During internal evaluation, the complete dataset (219 experiments) was randomly splitted in a training set (3/4) and a test set (1/4). The test set, an assemblage of 54 experiments, was treated as if it concerned independent, 'unknown' samples. Therefore, the data of the test set were projected on the PCA-space as determined by the training set and the significant PC's ($\alpha < 0.05$) were used to calculate the patterns of the test set. Based on the resulting patterns, the similarities (Euclidean distance) between the 54 experiments of the test set and all experiments in the training set were calculated. The identification was regarded as reliable at the species level when at least three out of five matches pointed to the same species. Since only the first five matches were considered, this corresponds to a relative proportion of identification = 0.6. The training set consisted of 165 (= 219 - 54) experiments.

For external evaluation a set of 19 strains was assembled independently and in a blind way. These tester strains encompass (1) strains that were already included in the training set, (2) strains not represented in the training set, but known to be members of species represented in the training set and (3) one strain belonging to a species other than those represented in the training set. Strains of this blind study were analyzed in duplicate, using freshly prepared BHI growth media. These 19 strains of which the identity was unknown for the experimenter, were identified by comparison with the members of the training set exhibiting the highest similarities (Euclidean distance).

Results

Comparison of Raman spectra. Representative Raman spectra of some investigated type strains are shown in Figure 1(a). Correspondingly their high genetic similarity, the Raman spectra of these strains are highly similar under the specified growth conditions, with the exception of striking bands near 1156 cm^{-1} and 1520 cm^{-1} present in the Raman spectra of LMG 18725^T. These bands together with the band at 1004 cm^{-1} can be attributed to β-carotene[39]. Figure 1(b) illustrates that these characteristic bands are not reliable for species discrimination by the Raman spectra of two *B. vallismortis* strains (LMG 18725^T and LMG 17800). Strain dependent carotenoid levels, as measured by Raman analysis, were also reported for *Enterococcus casseliflavius* and *Enterococcus hirae*[14].

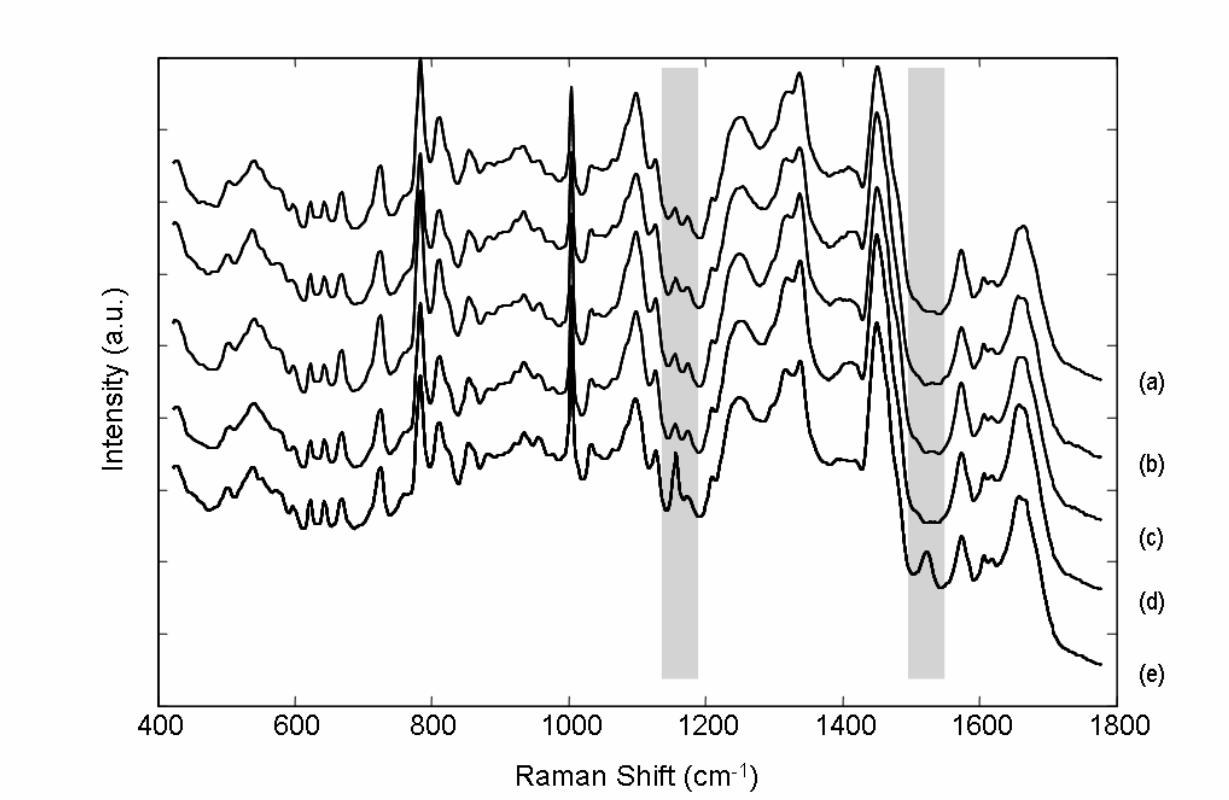


Figure 1(a): Representative Raman spectra obtained by averaging all Raman spectra of (a) *B. amyloliquefaciens* LMG 9814T, (b) *B. atrophaeus* LMG 16797T, (c) *B. mojavensis* LMG 17797T, (d) *B. subtilis* LMG 7135T and (e) *B. vallismortis* LMG 18725T. All strains were grown on BHI at 37 °C for 7 hours. Raman bands caused by carotenoid structures are highlighted. The spectra have been displaced vertically on the intensity axis for (a.u.: arbitrary units).

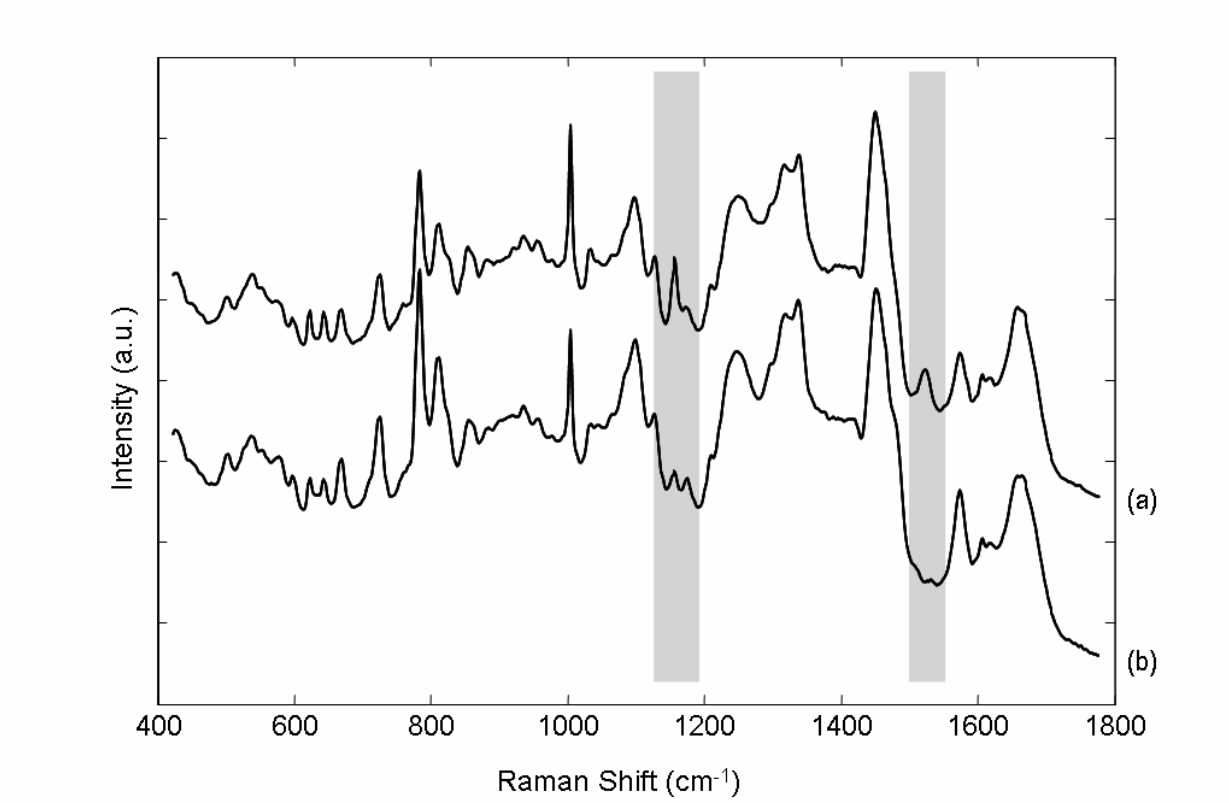


Figure 1(b): Representative Raman spectra obtained by averaging all Raman spectra of two *B. vallismortis* strains (a) LMG 18725^T and (b) LMG 17800. All strains were grown on BHI at 37 °C for 7 hours. Striking Raman bands at 1520 cm^{-1}, 1156 cm^{-1} (highlighted) together with the intense band at 1004 cm^{-1} are caused by increased levels of β-carotene. The spectra have been displaced vertically on the intensity axis for clarity (a.u.: arbitrary units).

Due to their complexity, it is generally difficult to evaluate Raman fingerprints visually. Yet, small biochemical differences between closely related species can be visualized by means of their difference (derivative) spectra. As outlined in the experimental section, not the Raman spectra as such, but their first derivative spectra are used for analysis. Figure 2 shows the derivative spectra for two replicate experiments (*B. amyloliquefaciens* LMG 9814^T) together with their difference spectrum, demonstrating the high similarity between these fingerprints. As a comparison, the derivative spectra of the same strain and *B. vallismortis* LMG 17800^T, are also shown (Figure 3). Here, subtle spectral differences between both fingerprints appear.

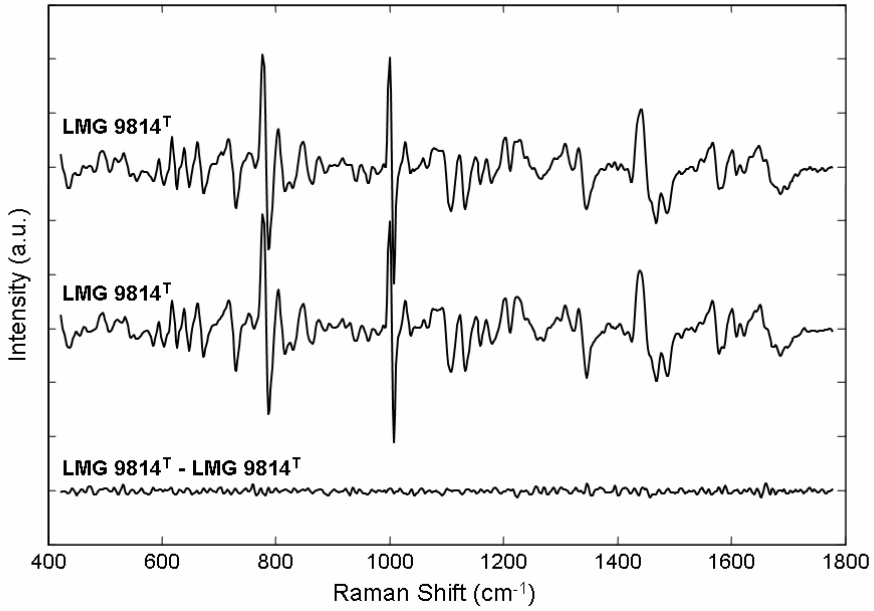


Figure 2: First derivative spectra of *B. amyloliquefaciens* LMG 9814^T as used for analysis (see text for details). The difference spectrum of these two independent experiments (LMG 9814^T - LMG 9814^T), reveals no important spectral differences. The strain was grown on BHI at 37 °C for 7 hours. The spectra have been displaced vertically on the intensity axis for clarity (a.u.: arbitrary units).

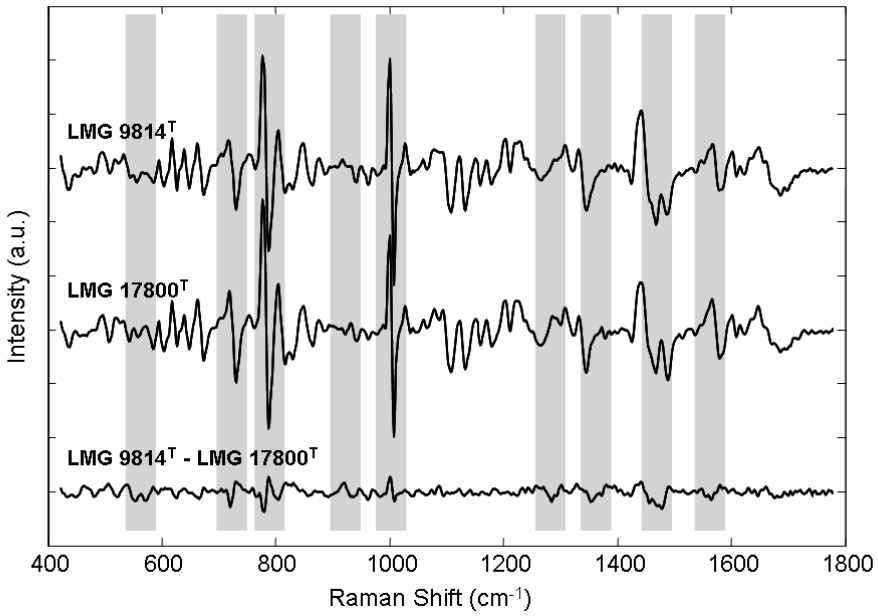


Figure 3: First derivative spectra of *B. amyloliquefaciens* LMG 9814^T and *B. vallismortis* LMG 17800^T as used for analysis. The difference spectrum (LMG 9814^T - LMG 17800^T) shows subtle differences allowing identification at the species level. The most striking visual distinctive Raman bands are highlighted. All strains were grown on BHI at 37 °C for 7 hours. The spectra have been displaced vertically on the intensity axis for clarity (a.u.: arbitrary units).

A detailed attribution of these spectral differences is beyond the scope of this work. Nevertheless, it is shown that Raman spectra provide information on biochemical differences allowing identification at the species level (see below). Since it concerns phenotypical information, a reliable discrimination (and taxonomic resolution) will only be achieved when experimental factors such as preparation and composition of growth media, incubation conditions and the protocols for measuring and analyzing the Raman spectra are optimal and reproducible[11].

Internal evaluation of the reference library

To evaluate the identification ability of Raman spectroscopy, 69 strains (Table 1) belonging to eight species were used for the construction of a database covering the biological variation per species. In some cases, the biological species variation is unknown because of the lack of sufficient strains. However, all strains were selected for their known correct assignment at the species level.

For internal evaluation, the database was randomly divided in a training set (3/4) and a test set (1/4). The identification of test entries is based on the highest five matches with entries of the training set at the species level. The number of times a given species name appears in these first five matches is represented by its relative proportion. This proportion equals one, if five out of five matches point towards the same species, 0.8 if four out of five matches point towards the same species, etc. The results are reported in Table 2. The relative proportions may be regarded as an appreciation of correct identification. Misidentifications are marked with an asterisk (*).

Table 2: Identification results of the test set (54 experiments). Taxonomic status of the species from the test set is horizontally shown. Species included by the training set are listed vertically. Final species identification follows from the relative proportion of identification. Misidentifications at species level are marked with an asterisk (*).

Number	Taxonomic status	*B.amyloliquefaciens*	*B. atrophaeus*	*B. licheniformis*	*B. mojavensis*	*B. pumilus*	*B. sonorensis*	*B.subtilis*	*B. vallismortis*	Final identification
1	*B. amyloliquefaciens* LMG 12325	0.2						0.8		*B. subtilis* *
2	*B. amyloliquefaciens* LMG 12325	1								*B. amyloliquefaciens*
3	*B. amyloliquefaciens* LMG 12325	1								*B. amyloliquefaciens*
4	*B. amyloliquefaciens* LMG 9814^T	0.8			0.2					*B. amyloliquefaciens*
5	*B. amyloliquefaciens* LMG 12330	0.8						0.2		*B. amyloliquefaciens*
6	*B. amyloliquefaciens* LMG 17599	0.4				0.2		0.4		***-Not identified-***
7	*B. amyloliquefaciens* LMG 9814^T	1								*B. amyloliquefaciens*
8	*B. amyloliquefaciens* LMG 12330	0.6						0.4		*B. amyloliquefaciens*
9	*B. amyloliquefaciens* LMG 12385	1								*B. amyloliquefaciens*
10	*B.atrophaeus* LMG 17795		1							*B.atrophaeus*
11	*B.atrophaeus* LMG 17795	0.2	0.4			0.2		0.2		***-Not identified-***
12	*B.atrophaeus* LMG 17796		1							*B.atrophaeus*
13	*B.atrophaeus* LMG 17795		1							*B.atrophaeus*
14	*B.atrophaeus* LMG 16797^T		1							*B.atrophaeus*
15	*B. licheniformis* LMG 12363^T			1						*B. licheniformis*
16	*B. licheniformis* LMG 6934			0.8		0.2				*B. licheniformis*
17	*B. licheniformis* LMG 6934			1						*B. licheniformis*
18	*B. licheniformis* LMG 7560			0.8						*B. licheniformis*
19	*B. licheniformis* LMG 7560			1						*B. licheniformis*
20	*B. licheniformis* LMG 7560			1						*B. licheniformis*
21	*B. licheniformis* LMG 7560			0.8						*B. licheniformis*
22	*B. licheniformis* LMG 7630			1						*B. licheniformis*
23	*B. licheniformis* LMG 17659			1						*B. licheniformis*
24	*B. licheniformis* LMG 17337			1						*B. licheniformis*
25	*B. licheniformis* LMG 17649			1						*B. licheniformis*
26	*B. licheniformis* LMG 17339			1						*B. licheniformis*
27	*B. licheniformis* LMG 7630			1						*B. licheniformis*
28	*B. licheniformis* LMG 12363^T			1						*B. licheniformis*
29	*B. mojavensis* LMG 17797				1					*B. mojavensis*
30	*B. mojavensis* LMG 17798	0.4			0.6					*B. mojavensis*
31	*B. pumilus* LMG 10826					1				*B. pumilus*
32	*B. pumilus* LMG 12372					1				*B. pumilus*
33	*B. pumilus* LMG 12372			0.2		0.8				*B. pumilus*
34	*B. pumilus* LMG 18928^T					1				*B. pumilus*
35	*B. pumilus* LMG 18928^T					1				*B. pumilus*
36	*B. pumilus* LMG 12372		0.4			0.6				*B. pumilus*
37	*B. pumilus* LMG 12376			0.2		0.8				*B. pumilus*
38	*B. pumilus* LMG 18517					1				*B. pumilus*
39	*B. pumilus* LMG 12259			0.2		0.8				*B. pumilus*
40	*B. sonorensis* LMG 21636^T					0.2	0.8			*B. sonorensis*
41	*B. sonorensis* LMG 21636^T		0.2			0.2	0.6			*B. sonorensis*
42	*B. subtilis* LMG 8197				0.6			0.2	0.2	*B. mojavensis* *
43	*B. subtilis* LMG 8198	0.6		0.2	0.2					*B. amyloliquefaciens* *
44	*B. subtilis* LMG 7135^T							1		*B. subtilis*
45	*B. subtilis* LMG 17725							1		*B. subtilis*
46	*B. subtilis* LMG 19154							1		*B. subtilis*
47	*B. subtilis* LMG 8198	0.4						0.6		*B. subtilis*
48	*B. subtilis* LMG 17725									*B. subtilis*
49	*B. subtilis* LMG 19154							1		*B. subtilis*
50	*B. subtilis* LMG 12261							1		*B. subtilis*
51	*B. subtilis* LMG 17725				0.4			0.6		*B. subtilis*
52	*B. vallismortis* LMG 18725			0.2					0.8	*B. vallismortis*
53	*B. vallismortis* LMG 17800^T					0.4			0.6	*B. vallismortis*
54	*B. vallismortis* LMG 18725								1	*B. vallismortis*

An overall identification accuracy at the species level of more than 90 % (49/54) was obtained. Only three patterns were misidentified and two could not be identified unequivocally. Misidentifications concerned strains attributed to *B. amyloliquefaciens, B. subtilis* and *B. mojavensis.* When grown as specified, these species are distinguished only by subtle spectral differences. Two patterns obtained from *B. amyloliquefaciens* (nr 6. in Table 2) and *B. atrophaeus* (nr. 11 in Table 2), could not be identified because the first five matches corresponded to at least three different species without a clear preference for one of them.

The internal evaluation did not reveal misidentifications for *B. atrophaeus* and *B. vallismortis* strains. Although some of the tested patterns matched with more than one species, the majority of the matches (3 out of 5) pointed to the same species of the *'B. subtilis'*-group and the satellite species *B. pumilus*, *B. licheniformis* and *B. sonorensis*. In this context (Table 2), two situations in which the relative proportion of identification equals 0.6 can occur (1) 0.6, 0.2, 0.2 (indicating a three species match) and (2) 0.6, 0.4 (indicating a two species match). Clearly, the first identification profile is more explicit and is probably more reliable. The results of the internal evaluation underline the potential of Raman spectroscopy to discriminate highly related species based even on subtle spectral differences.

External evaluation: blind study

One of the drawbacks of internal evaluations is that these are sometimes not realistic in terms of identification accuracy. As such, internal evaluations are known to exhibit higher identification accuracies than when independent, external evaluations are performed[24]. Therefore, a modest blind study was set-up, based on 19 tester strains that were analyzed in duplicate (i.e. in total 38 experiments) and identified towards the training set. To maximize the eventual effect of medium preparation and/or batches, spectra were collected from growth on independently prepared medium batches.

Identification results of the blind study (strains A – S) for each replicate experiment are shown in Table 3. From the eleven strains that *a posteriori* turned out to be included in the training set, all but one (Unknown F) were correctly identified. The identification accuracy of 91 % (10/11) supports the internal evaluation and herewith the reproducibility of the method following the protocol. The patterns of unknown F pointed to *B. subtilis* (relative proportion of identification = 0.6) for the first replicate experiment, but did not reach the threshold (relative proportion of identification = 0.6) for the second replicate experiment. Therefore, the cumulation of both experiments did not lead to convincing species identification, although strain LMG 12263 was included in the training set.

Table 3: Identification results for the 19 tester strains (A – S) of the blind study. Final species identification and strain assignment (after disclosure) are given. The upper part of the table reports the identification results of strains already included in the training set, while the lower part shows the identification results for strains that are new to the training set (marked with a superscript case N (^{N})). For each replicate experiment (I & II) the relative proportions of identification are shown.

	Replicate Measurement I								Replicate Measurement II									
	B.amyloliquefaciens	*B. atrophaeus*	*B. licheiformis*	*B. mojavensis*	*B. pumilus*	*B. sonorensis*	*B. subtilis*	*B. vallismortis*	*B.amyloliquefaciens*	*B. atrophaeus*	*B. licheiformis*	*B. mojavensis*	*B. pumilus*	*B. sonorensis*	*B. subtilis*	*B. vallismortis*	***Final species identification***	***Known strain assignment***
Unknown A			0.8				0.2				1						*B.licheniformis*	*B. licheniformis* LMG 7559
Unknown D			0.8				0.2				1						*B. licheniformis*	*B. licheniformis* LMG 12248
Unknown E					1								0.8			0.2	*B. pumilus*	*B. pumilus* LMG 12259
Unknown F							0.6	0.4			0.2		0.2		0.4	0.2	***-Not identified-***	*B. subtilis* LMG 12263
Unknown J				0.2			0.8		0.2						0.8		*B.subtilis*	*B. subtilis* LMG 12381
Unknown K		1								1							*B. atrophaeus*	*B. atrophaeus* LMG 16797
Unknown M			0.8				0.2				1						*B. licheniformis*	*B. licheniformis* LMG 17339
Unknown O							1								1		*B. subtilis*	*B. subtilis* LMG 19155
Unknown Q								1								1	*B. vallismortis*	*B. vallismortis* LMG 17800
Unknown R					1								1				*B. pumilus*	*B. pumilus* LMG 18928^{T}
Unknown S						1								1			*B. sonorensis*	*B. sonorensis* LMG 21636^{T}
Unknown B	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	***-Not identified-***	*B. atrophaeus* LMG 8198^{N}
Unknown C				0.2	0.8								1				*B. pumilus*	*B. pumilus* LMG 10825^{N}
Unknown G							1		0.6						0.4		***-Not identified-***	*B. amyloliquefaciens* LMG 12326^{N}
Unknown H	0.6						0.4		1								*B. amyloliquefaciens*	*B. amyloliquefaciens* LMG 12328^{N}
Unknown I								1								1	*B. pumilus*	*B. pumilus* LMG 12377^{N}
Unknown L			1								1						*B. licheniformis*	*B. licheniformis* LMG 12336^{N}
Unknown N							1								1		*B. subtilis*	*B. subtilis* LMG 17728^{N}
Unknown P					0.4		0.6		0.2				0.6		0.2		***-Not identified-***	*B. velezensis* LMG 22478^{T,N}

The blind study also reveals the potential of the method as a general identification tool because of the inclusion of eight strains that were not represented hitherto in the database. Five out of these eight strains were correctly identified at the species level, which is particular encouraging considered the small spectral differences between the species studied. Hence, it is concluded that even these subtle differences are sufficiently reproducible for species discrimination.

The other three strains, not included in the database (unknowns B, G and P), were misidentified for miscellaneous reasons. The colony of unknown B exhibited an orange pigmentation. Intense peaks at 1134 cm^{-1} and 1532 cm^{-1} in the Raman spectra (not shown) of this strain pointed to the presence of carotenoid structures[1]. Since the Raman spectra of this strain deviated importantly from the spectra encompassed by the training set, unknown B was categorized *a priori* as unknown. Unknown B was disclosed as *B. atrophaeus* LMG 8198 (= NCIB 8058, DSM 675, ATCC 9372), also named the 'red strain'[9].

For unknown G, the data obtained from the first replicate experiment point towards *B. subtilis* (proportion of identification = 1), while the second replicate experiment points towards *B. amyloliquefaciens* (proportion of identification = 0.6). During internal evaluation, inconsistent identifications for *B. subtilis*, *B. amyloliquefaciens* and *B. mojavensis* were also found.

Unknown P was disclosed as the recently described type strain of *B. velezensis* LMG 22478^{T} [30], a member of the homogeneous *'B. subtilis'*-group that was not represented by the training set used during the blind study. Hence, the non identification was a correct answer.

Discussion

To our best knowledge, it is the first time that Raman spectra of bacteria have been evaluated for their identification performance at the species level of so closely related bacteria as those belonging to the '*B. subtilis*'-group (*sensu stricto*) and its direct vicinity. Promising results were obtained if a strict standardized protocol is followed to exclude the potential experimental variations as much as possible. The combination of a randomized experimental set-up and a blind, external evaluation should provide a most realistic view on the discriminative ability of Raman spectroscopy for species identification. The study demonstrates that related and phenotypically similar *Bacillus* species can be discriminated by Raman spectroscopy.

Nevertheless, the Raman method described here should not be considered as a stand-alone technique for taxonomic purposes. It is more an additional tool both in the polyphasic characterization process of bacteria, as well as for screening purposes allowing fast identification at the species or perhaps even finer levels. Supplementary, it is noted that in this type of whole cell approaches the interpretation of the Raman spectra might be affected by over-expressed and/or strain dependent characteristics. As an example of this, the presence of carotenoid structures is given. These structures gives rise to intense Raman bands, but are known not to be species descriptive.

The observation that Raman spectra of closely related species are less suited for direct visual evaluation might be considered as a drawback, for example, in relation to SDS-PAGE or AFLP[6]. However, subtraction of derivated Raman spectra reveal small, nonetheless significant, differences that have shown to be sufficient for species discrimination. Indeed, both modest internal and external evaluations indicate that Raman spectroscopy applies well as a stable identification method since at least 90 % of the strains that were included in the reference dataset and most strains that were not included in the dataset were correctly identified, both at the species level. Difficulties and contradictions in the identifications were only observed within taxa that are most probably not clearly separated.

Another condition *sine qua non*, concerns the composition of the dataset that is used for identification. Indeed, only unambiguously classified strains covering the biological variation of the taxon, can be included. These results support previous studies on the evaluation of Raman spectroscopy for the identification of micro-organisms[18,14].

A final advantage concerns the very limited sample preparation and the short period in which the identification results are obtained (less than 8 hours, starting from incubation).

Acknowledgements

The authors are grateful to the Research Foundation - Flanders (FWO-Vlaanderen) for financial support (Grants G.0156.02 and G.0343.03) of this research project. D.H. greatly acknowledges the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for his doctoral grant. J.H. and P.V. wish to thank the FWO-Vlaanderen for their postdoctoral fellowships.

Reference List

1. **Armstrong, G. A.** 1997. Genetics of eubacterial carotenoid biosynthesis: a colorful tale. Ann Rev Microbiol **51**:629-659.

2. **Ash, C., F. G. Priest, and M. D. Collins**. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus *Paenibacillus*. Antonie Van Leeuwenhoek **64**:253-260.

3. **Behrendt, U., A. Ulrich, P. Schumann, D. Naumann, and K. Suzuki**. 2002. Diversity of grass-associated *Microbacteriaceae* isolated from the phyllosphere and litter layer after mulching the sward; polyphasic characterization of *Subtercola pratensis* sp. nov., *Curtobacterium herbarum* sp. nov. and *Plantibacter flavus* gen. nov., sp. nov. Int J Syst Evol Microbiol **52**:1441-1454.

4. **Claus, D., and R. C. W. Berkeley.** 1986. Genus *Bacillus* Cohn 1872, 174 AL, *In*: P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G. Holt (Eds): Bergey's Manual of Systematic Bacteriology. Williams & Wilkins, Baltimore, pp. 1105-1139.

5. **Crescenzi, M., and A. Giuliani**. 2001. The main biological determinants of tumor line taxonomy elucidated by a principal component analysis of microarray data. FEBS Lett **507**:114-118.

6. **De Vos, P.** 2002. Nucleic Acid Analysis and SDS-PAGE of Whole-cell Proteins in *Bacillus* taxonomy, p. 141-159. *in*: Applications and Systematics of *Bacillus* and Relatves (R. Berkeley, M. Heyndrickx, N. Logan, P. De Vos (eds.)).

7. **Fortina, M. G., R. Pukall, P. Schumann, D. Mora, C. Parini, P. L. Manachini, and E. Stackebrandt**. 2001. *Ureibacillus* gen. nov., a new genus to accommodate *Bacillus* thermosphaericus (Andersson et al. 1995), emendation of *Ureibacillus thermosphaericus* and description of *Ureibacillus terrenus* sp. nov. Int J Syst Evol Microbiol **51**:447-455.

8. **Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk Jr**. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol **42**:166-170.

9. **Fritze, D., and R. Pukall**. 2001. Reclassification of bioindicator strains *Bacillus subtilis* DSM 675 and *Bacillus subtilis* DSM 2277 as *Bacillus atrophaeus*. Int J Syst Evol Microbiol **51**:35-37.

10. **Heyndrickx, M., L. Lebbe, K. Kersters, B. Hoste, R. De Wachter, P. De Vos, G. Forsyth, and N. A. Logan**. 1999. Proposal of *Virgibacillus proomii* sp. nov. and emended description of *Virgibacillus pantothenticus* (Proom and Knight 1950) Heyndrickx et al. 1998. Int J Syst Bacteriol **49**:1083-1090.

11. **Hutsebaut, D., K. Maquelin, P. De Vos, P. Vandenabeele, L. Moens, and G. J. Puppels**. 2004. Effect of Culture Conditions on the Achievable Taxonomic Resolution of Raman Spectroscopy Disclosed by Three *Bacillus* Species. Anal Chem **76**:6274-6281.

12. **Hutsebaut, D., P. Vandenabeele, and L. Moens.** 2005. Evaluation of an Accurate Calibration and Spectral Standardization Procedure for Raman Spectroscopy. The Analyst **130**: 1204 – 1214.

13. **Ibelings, M. S., K. Maquelin, H. P. Endtz, H. A. Bruining, and G. J. Puppels**. 2005. Rapid identification of *Candida* spp. in peritonitis patients by Raman spectroscopy. Clin Microbiol Infect **11**:353-358.

14. **Kirschner, C., K. Maquelin, P. Pina, N. A. Ngo Thi, L. P. Choo-Smith, G. D. Sockalingum, C. Sandt, D. Ami, F. Orsini, S. M. Doglia, P. Allouch, M. Mainfait, G. J. Puppels, and D. Naumann**. 2001. Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J Clin Microbiol **39**:1763-1770.

15. **Logan, N. A. and R. C. Berkeley**. 1984. Identification of *Bacillus* strains using the API system. J Gen Microbiol **130**:1871-1882.

16. **Maquelin, K., L. P. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H. A. Bruining, and G. J. Puppels**. 2000. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. Anal Chem **72**:12-19.

17. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. A. Ngo-Thi, T. van Vreeswijk, M. Stammler, H. P. Endtz, H. A. Bruining, D. Naumann, and G. J. Puppels**. 2003. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. J Clin Microbiol **41**:324-329.

18. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, and G. J. Puppels**. 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J Microbiol Methods **51**:255-271.

19. **Nakamura, L. K., M. S. Roberts, and F. M. Cohan**. 1999. Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: a proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov. and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. Int J Syst Bacteriol **49**:1211-1215.

20. **Nakamura, N. K.** 1989. Taxonomic relationship of black-pigmented *Bacillus subtilis* strains and a proposal for *Bacillus atrophaeus* sp.nov. Int J of Syst Bacteriol **39**: 295-300

21. **Naumann, D., D. Helm, and H. Labischinski**. 1991. Microbiological characterizations by FT-IR spectroscopy. Nature **351**:81-82.

22. **Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev, and M. V. Ivanov**. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermodenitrificans*. Int J Syst Evol Microbiol **51**:433-446.

23. **Niimura Y., Koy E. Yanagida F. Suzuki K. I. Komagata K. Kozaki M.** 1990. *Amphibacillus xylanus* gen.nov., sp.nov., a facultatively anaerobic sporeforming xylan digesting bacterium which lacks cytochrome, quinone and catalase. Int J Syst Bacteriol **40**: 297-301.

24. **Oberreuter, H., H. Seiler, and S. Scherer**. 2002. Identification of coryneform bacteria and related taxa by Fourier-transform infrared (FT-IR) spectroscopy. Int J Syst Evol Microbiol **52**:91-100.

25. **Palmisano, M. M., L. K. Nakamura, K. E. Duncan, C. A. Istock, and F. M. Cohan**. 2001. *Bacillus sonorensis* sp. nov., a close relative of *Bacillus licheniformis*, isolated from soil in the Sonoran Desert, Arizona. Int J Syst Evol Microbiol **51**:1671-1679.

26. **Pikuta, E., A. Lysenko, N. Chuvilskaya, U. Mendrock, H. Hippe, N. Suzina, D. Nikitin, G. Osipov, and K. Laurinavichius**. 2000. *Anoxybacillus pushchinensis* gen. nov., sp. nov., a novel anaerobic, alkaliphilic, moderately thermophilic bacterium from manure, and description of *Anoxybacillus flavitherms* comb. nov. Int J Syst Evol Microbiol **50**:2109-2117.

27. **Puppels, G. J., W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. de Mul, and J. Greve.** 1991. Description and performance of a highly sensitive confocal Raman spectrometer. J Raman Spectros **22**, 217-225.

28. **Roberts, M. S., L. K. Nakamura, and F. M. Cohan**. 1994. *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. Int J Syst Bacteriol **44**:256-264.

29. **Roberts, M. S., L. K. Nakamura, and F. M. Cohan**. 1996. *Bacillus vallismortis* sp. nov., a close relative of *Bacillus subtilis*, isolated from soil in Death Valley, California. Int J Syst Bacteriol **46**:470-475.

30. **Ruiz-Garcia, C., V. Bejar, F. Martinez-Checa, I. Llamas, and E. Quesada**. 2005. *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Velez in Malaga, southern Spain. Int J Syst Evol Microbiol **55**:191-195.

31. **Ruiz-Garcia, C., E. Quesada, F. Martinez-Checa, I. Llamas, M. C. Urdaci, and V. Bejar**. 2005. *Bacillus axarquiensis* sp. nov. and *Bacillus malacitensis* sp. nov., isolated from river-mouth sediments in southern Spain. Int J Syst Evol Microbiol **55**:1279-1285.

32. **Schlesner, H., P. A. Lawson, M. D. Collins, N. Weiss, U. Wehmeyer, H. Volker, and M. Thomm**. 2001. *Filobacillus milensis* gen. nov., sp. nov., a new halophilic spore-forming bacterium with Orn-D-Glu-type peptidoglycan. Int J Syst Evol Microbiol **51**:425-431.

33. **Shida, O., H. Takagi, K. Kadowaki, and K. Komagata**. 1996. Proposal for two new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov. Int J Syst Bacteriol **46**:939-946.

34. **Spring, S., W. Ludwig, M. C. Marquez, A. Ventosa, and K. H. Schleifer.** 1996. *Halobacillus* gen. nov., with descriptions of *Halobacillus litoralis* sp. nov. and *Halobacillus trueperi* sp. nov. and transfer of *Sporosarcina halophila* to *Halobacillus halophilus* comb. nov. Int. J. Syst. Bacteriol. **46**: 492-496.

35. **Touzel, J. P., M. O'Donohue, P. Debeire, E. Samain, and C. Breton**. 2000. *Thermobacillus xylanilyticus* gen. nov., sp. nov., a new aerobic thermophilic xylan-degrading bacterium isolated from farm soil. Int J Syst Evol Microbiol **50**:315-320.

36. **Vandeginste, B. G. M., D. L. Massart, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers –Verbeke.** 1998. *In*: B.G.M. Vandeginste, S.C. Rutan (Eds.), Handbook of Chemometrics and Qualimetrics: Part B, Elsevier Science, Amsterdam.

37. **Waino, M., B. J. Tindall, P. Schumann, and K. Ingvorsen**. 1999. *Gracilibacillus* gen. nov., with description of *Gracilibacillus halotolerans* gen. nov., sp. nov.; transfer of *Bacillus dipsosauri* to *Gracilibacillus dipsosauri* comb. nov., and *Bacillus salexigens* to the genus *Salibacillus* gen. nov., as *Salibacillus salexigens* comb. nov. Int J Syst Bacteriol **49**:821-31.

38. **Wisotzkey, J. D., P. Jurtshuk Jr, G. E. Fox, G. Deinhard, and K. Poralla**. 1992. Comparative sequence analyses on the 16S rRNA (rDNA) of *Bacillus acidocaldarius, Bacillus acidoterrestris,* and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. Int J Syst Bacteriol **42**:263-269.

39. **Withnall, R., B. Z. Chowdhry, J. Silver, H. G. Edwards, and L. F. de Oliveira**. 2003. Raman spectra of carotenoids in natural products. Spectrochim Acta A Mol Biomol Spectrosc **59**:2207-2212.

40. **Yoon, J. H., K. H. Kang, and Y. H. Park**. 2002. *Lentibacillus salicampi* gen. nov., sp. nov., a moderately halophilic bacterium isolated from a salt field in Korea. Int J Syst Evol Microbiol **52**:2043-2048.

41. **Yoon, J. H., N. Weiss, K. C. Lee, I. S. Lee, K. H. Kang, and Y. H. Park**. 2001. *Jeotgalibacillus alimentarius* gen. nov., sp. nov., a novel bacterium isolated from jeotgal with L-lysine in the cell wall, and reclassification of *Bacillus marinus* Ruger 1983 as *Marinibacillus marinus* gen nov., comb. nov. Int J Syst Evol Microbiol **51**:2087-2093.

Exploration of Micro-Raman Spectroscopy as a taxonomic tool within the *'Bacillus cereus'* group: A comparison with FAFLP

Didier Hutsebaut
Leen Van Brandt
Marie-Hélène Guinebretière
Jeroen Heyrman
Marc Heyndrickx
Peter Vandenabeele
Luc Moens
Paul De Vos

Chapter V

Draft version

V. Exploration of Micro-Raman Spectroscopy as a taxonomic tool within the '*Bacillus cereus'* group: A comparison with FAFLP.

This study reports on the exploration of Raman spectroscopy as a taxonomic tool within the complex '*Bacillus cereus*'- group and close relatives (*B. mycoides, B. pseudomycoides*). Seventy-six strains, encompassing culture collection strains as well as food isolates, were selected based on their allocation to the '*B. cereus'*-group and analyzed by fatty acid methyl ester (FAME) analysis, Raman spectroscopy, Fluorescent Amplified Length Polymorphism (FAFLP) and rep-PCR.

Raman clusters were delineated at two different levels and revealed good correspondence to the clusters obtained by FAFLP analysis. Furthermore, the obtained Raman sub-clusters were generally in good agreement with the clusters obtained by rep-PCR fingerprinting, since they repeatedly group identical isolates together in the same Raman sub-cluster.

Since groupings based on cluster analysis are generally affected by the selected linkage algorithm (e.g. UPGMA, Ward) an additional correlation study was performed on the original similarity matrices obtained by FAFLP and Raman spectroscopy. A highly significant ($\alpha < 0.0001$) correlation could be established based on Kendall's Tau non-parametric statistic. The results indicate that Raman spectroscopy has potential as a tool in taxonomic studies.

Introduction

Vibrational methods (Fourier-Transform Infrared (FT-IR) and Raman spectroscopy) provide distinct molecular fingerprints of the overall bacterial cell composition (proteins, teichoic acids, lipolysaccharides, fatty acids, DNA/RNA, etc.)[23] that are claimed to be highly reproducible and strain specific[8,16,17,25]. However, when the number of strains studied increases, more advanced statistical techniques are needed for successful discrimination, even at the species level[3,25].

As such, Raman spectroscopy has been applied successfully for the identification of clinical relevant species[14,18,20,21]. The potential of this technique to identify closely related species was further elaborated within the phylogenetically homogeneous '*Bacillus subtilis*'-group[12].

For yeasts, FT-IR spectroscopy as typing method was investigated by Sandt *et al.* who obtained similar groupings for RAPD and FT-IR spectroscopy within *Candida albicans* strains[29]. Other studies report on the ability of FT-IR to distinguish multiple yeast isolates at the sub-genotype[17] and even the strain level[35].

For bacteria, it was shown that 25 *Actinetobacter* strains belonging to the *Acinetobacter* complex group according to Raman spectroscopic analysis (RSA) in a highly similar way as by AFLP analysis[19]. However, groupings as obtained by cluster analysis are affected by the linkage algorithm (UPGMA, Ward, etc.) and hence should be regarded as 'secondary data', while the original similarity matrices constitute 'primary data'[27]. As such, a correlation study should be performed to underpin the obtained clusters. The taxonomic value of these clusters needs further polyphasic investment. However, a significant correlation between vibrational and genotypic (fingerprinting) methods has not been demonstrated yet.

Oberreuter *et al.*[24] calculated the correlation between 16S rDNA sequence similarities and FT-IR spectral similarities for *Brevibacterium linens, Corynebacterium glutamicum* and *Rhodococcus erythropolis* but a correlation below the species level could not be observed, as could be expected.

As the phylogenetically homogenous '*Bacillus cereus*'-group is also known for its controversial taxonomy[1,7,22], it is probable that the application of RSA contributes to the elucidation of the vague taxonomic situation of these bacilli. Apart from collection strains, food isolates were included that were attributed to the '*Bacillus cereus*'-group by phenotype (classical tests, API, FAME), 16S rDNA sequencing or specific PCR. Amongst them, there are strains that are genomically highly related according to rep-PCR fingerprinting[4,5] and hence it is aimed to obtain a better evaluation on the reliability of the proposed Raman method. The clusters obtained based on RSA and those obtained by FAFLP analysis were compared visually; the original similarity matrices were compared by a non-parametric statistical approach (Kendall's Tau).

Materials and Methods

Strains. In this study, 76 strains were selected based on their allocation to the '*B. cereus*'-group. The designations of both the reference strains (LMG-strains) and the food isolates used (R-strains) are shown in Table 1, together with their origins and corresponding groupings as obtained by different methods. Reference strains were derived from the BCCM™ / LMG (Belgian Coordinated Collections of Micro-organisms, LMG Bacteria collection). Pure cultures were stored at –80°C in Microbank™ vessels containing 15% v/v glycerol (cryoprotectant) until use.

Raman spectroscopic analysis (RSA). For Raman measurements, strains were grown on Brain Heart infusion (BHI) at 37 °C (± 1°C) for exactly 7 hours to reduce discrepancies in the Raman spectra, due to colony heterogeneity[2]. Preparation of BHI culture plates, growth of strains, sampling and sample preparation was performed as described previously[12]. Confluent grown cells were harvested from the third quadrant with a disposable, plastic inoculation loop and manually spotted on a CaF_2 substrate (polished, UV-grade, Crystran Ltd., Poole, United Kingdom). The spotted biomass was allowed to dry at room temperature for 20 min in a desiccator containing drying beads. The CaF_2 substrate with the dried smears was placed directly under the microscope of a Kaiser System Hololab 5000R modular Raman micro spectrometer (f/1.8) (KOSI, Ecully, France). The microscope was fitted with a 100x objective (PL Fluotar L, 100x/0.75, W.D. 4.7 mm, Leica). Samples were excited using 45 - 50 mW of 785 nm laser light from a diode laser (Toptica Photonics AG, Martinsried/Munich, Germany). The scattered light is guided to the spectrograph by means of a confocal, 15 µm aperture collection fiber. A back illuminated deep depletion CCD camera (Andor, Belfast, Northern Ireland) was used for the detection of the scattered light. Raman signal was collected in the spectral interval of 150 cm^{-1} – 3100 cm^{-1} with a spectral resolution of 4 cm^{-1}. To compensate for potential sample heterogeneity, Raman spectra were collected at 10 randomly chosen locations within each spot, using a signal collection time of 60 s for each spectrum. Calibration of the Raman set-up and preprocessing of the collected spectra was performed as previously described[12,13].

To eliminate the potential interference of unpredictable day-to-day variations, planning of the experiments was fully randomized and at least three replicate experiments were recorded for each strain on different days and different medium batches.

Table 1: Overview of the strains used, their allocation and origin. Groupings obtained by Raman spectroscopy, Fatty Acid Analysis (FAME) and FAFLP.

* Identification of the BCCM/LMG Bacteria Collection strains are given as in the catalogue; identification of research strains (R-numbers) are based on 1) classical tests, API and FA, 2) 16S rDNA sequencing, 3) gyrB and cspA specific PCR, 4) rep-PCR, and 5) literature (Stenfors & Granum, 2001)[30]

† RAMAN spectra were grouped according to the Canberra metric and the Ward linkage algorithm. Groups (A-D) and subgroups (indices) were delineated at -40 and -20 (Ward scale), respectively; FAME patterns were clustered using Euclidean distance with the UPGMA coefficient – clusters showing more than 97.5 % internal similarity are named V to Z and subgroups showing more than 98.5 % similarity are indicated by an additional numeral; for FAFLP patterns a band based (Dice) cluster analysis (Ward) was used – clusters delineated above 30 % were indicated as 1 to 5 and subgroups delineated above 50% have an additional small character (a – b). Clusters that are not consistent in all three clusterings are highlighted in gray.

‡ R-22308 is not identical but highly similar to R-22292, R-22293 (Pearson correlation > 80 %).

For strains with similar rep-PCR profiles, only those strains were analyzed with FAME and FAFLP as indicated between brackets.

Strain	**Identification***	**Origin**	**RAMAN†**	**FAME†**	**FAFLP†**
R-22295	*B. cereus*-group[1]	pasteurized milk (Belgium)	C_1	X_1	1
R-22312	*B. cereus*-group[1]	rinsing water pasta bowl in fatal food borne outbreak (Belgium)	C_2	X_1	1
R-23229	*B. cereus*-group[1]	/	C_2	X_1	1
R-23230	*B. cereus*-group[1]	/	C_2	X_4	1
R-23231	*B. cereus*-group[1]	/	C_2	X_4	1
R-23232	*B. cereus*-group[1]	/	C_2	X_4	1
R-23233	*B. cereus*-group[1]	/	C_2	X_4	1
LMG 17612	*B. cereus*	faeces	C_2	W	-
LMG 6923^T	*B. cereus*	/	D_1	X_3	2

LMG 7138^T	*B. thuringiensis*	mediterranean flour moth	D_1	X_3	2
R-13620	*B. cereus*-group[2]	gelatin (porcine skin, Belgium)	D_1	X_3	2
R-22294	*B. cereus*-group[1]	raw milk (Belgium)	D_1	X_3	2
R-22317	*B. cereus*-group[1]	fatal food poisoning in elderly home (Belgium)	D_1	X_3	2
LMG 14742	*B. cereus*	contamination of pharmaceutique preparation	D_1	X_3	-
R-13500, R-13554	*B. cereus*-group[2]	gelatin (porcine skin, Belgium)	D_1	X_1 (R-13500)	2 (R-13500)
R-22313	*B. cereus*-group[1]	boiled pasta in fatal food borne family outbreak (Belgium)	D_1	X_5	2
LMG 17602	*B. cereus*	brain abscess, enterotoxigenic	D_1	/	/
LMG 18993^T	*B. pseudomycoides*	soil	C_1	/	3b
R-22296	*B. cereus*-group[1]	raw milk (Belgium)	C_1	/	3a
R-22298	*B. cereus*[4]	raw milk (Belgium)	C_1	Y	3a
R-22303	*B. pseudomycoides*[4]	pasteurized milk (Belgium)	C_1	Y	3b
R-22304	*B. pseudomycoides*[4]	raw milk (Belgium)	C_1	Y	3b
R-22305	*B. cereus*-group[1]	raw milk (Belgium)	C_1	Z	3a
R-20450, R-20452, R-20453, R-20454, R-20455, R-20456, R-20457, R-20458, R-20459, R-20460, R-20461, R-20462, R-20464, R-20465, R-20466, R-20469	*B. cereus*-group[2]	gelatin (porcine skin, USA)	A	X_1 (R-20460, R-20462)	4a (R-20460, R-20462)
R-22297	*B. cereus*[3]	raw milk (Belgium)	A	X_1	4a
R-22309 = LMG 22733	*B. cereus*-group[1]	pasta salad in fatal food borne family outbreak (Belgium)	A	X_1	4a
R-22310 = LMG 22729	*B. cereus*-group[1]	pasta salad in fatal food borne family outbreak (Belgium)	A	X_1	4a

R-22311 = LMG 22728	*B. cereus*-group[1]	vomit of deceased girl in fatal food borne family outbreak (Belgium)	A	X_1	4a
LMG 8943	*B. cereus*	soil of subsurface gold deposit	B_2	/	/
LMG 12334	*B. cereus*	vomit of monkey with implicated rice associated illness	B_2	X_3	/
LMG 12335	*B. cereus*	meat loaf causing a typical diarrhea type of food poisoning	B_2	X_3	/
R-13605	*B. cereus*-group[2]	gelatin (porcine skin, Belgium)	B_2	X_6	4a
R-22318 = LMG 17613	*B. cereus*-group[1]	mastitic milk (UK)	B_2	X_2	4b
R-10947, R-13559, R-13573, R-13574	*B. cereus*-group[2]	gelatin (porcine skin, Belgium)	B_1 and B_2	X_1 (R-10947, R-13574)	4b (R-10947, R-13574)
R-12866		water in gelatin factory (Belgium)	B_1	X_1	4b
R-20100, R-20144	*B. cereus*-group[2]	gelatin (bovine bones, France)	B_1	X_3 (R-20144)	4a (R-20144)
R-22300	*B. cereus*[3]	pasteurized milk (Belgium)	B_1	X_1	4b
R-13609, R-13610, R-13612, R-13613, R-13614, R-13615, R-13616	*B. cereus*-group[2]	gelatin (porcine skin, Belgium)	B_3	X_3 (R-13614)	2 (R-13614)
LMG 7128^T	*B. mycoides*	soil	D_2	V_1	5
LMG 18989^T	*B. weihenstephanensis*	pasteurized milk	D_2	V_1	5
R-22291	*B. mycoides/ weihenstephanensis*[3]	pasteurized organic milk (Belgium)	D_2	V_1	5
R-22314	*B. weihenstephanensis*	WS 2485 (Technische	D_2	V_1	5

		Universität München)			
R-22315	*B. weihenstephanensis*	milk (Norway)	D_2	V_1	5
R-22302	*B. cereus*-group[1]	raw milk (Belgium)	D_2	V_1	3a
R-22292, R-22293‡	*B. mycoides/ weihenstephanensis*[3]	raw milk (Belgium)	D_2	V_2	5
R-22308‡	*B. cereus*-group[1]	raw milk (Belgium)	D_2	/	5
R-22301	*B. weihenstephanensis*[4]	raw milk (Belgium)	D_2	V_3	5
R-22319 = 453-92	*B. weihenstephanensis*[5]	milk (Norway)	D_2	V_3	5
R-22299	*B. mycoides/ weihenstephanensis*[3]	pasteurized milk (Belgium)	D_2	V_3	5
LMG 17607	*B. cereus*	enterotoxemia in pig, produces lethal toxin	D_2	/	/

Statistical analysis of Raman data. The dataset was autoscaled and data reduction was performed using principal component analysis (PCA)[33] performed with the PLS toolbox (Eigenvector Research Inc., Manson, WA, USA) for Matlab software (Mathworks Inc., Natick, MA, USA). Those principal components (PC's) accounting for 99.9 % of the variation in the data set were retained. An *F*-test was used to individually select the most significant ($\alpha < 0.05$) PC's allowing strain discrimination. The selected PC's served as input for Linear Discriminant Analysis (LDA)[33] LDA was used only to calculate the canonical variates (CV's) with *a priori* knowledge of which spectra are replicate experiments, minimizing the variance within these replicate experiments, while maximizing the variation between different replicate experiments[15]. Hence, there is aimed at extracting that information, specific for each replicate experiment reducing variations in the Raman spectra as might be introduced by reculturing strains on different medium batches. These variations can interfere with the groupings as obtained by cluster analysis. In the following, the CV's thus obtained will be referred to as '*patterns*'.

Finally, for each strain all patterns were averaged and the similarity matrix of these average patterns was calculated using the Canberra metric. A dendrogram was constructed using Ward's algorithm. Construction of the dendrogram, calculation of similarity matrices and calculation of Kendall's tau coefficients were performed using the BioNumerics software version 4.01 (Applied Maths, Belgium).

Fatty Acid Methyl Ester (FAME) analysis. Analysis of the fatty acid contents of the cell walls was performed as previously described[11], starting from cells grown on tryptone soya agar (TSA) for 24 h.

Rep-PCR. DNA was extracted from the bacterial strains using a slight modification of the method of Pitcher[26], as described by Heyndrickx *et al.*[9]. Rep-PCR genomic fingerprinting was performed with the $(GTG)_5$ primer[34] applying the PCR conditions described by Rademaker *et al.*[28]. Electrophoresis and pattern analysis were further performed as described by Heyrman *et al.*[10]. Cluster analysis of rep-PCR patterns was done by analyzing the profiles with BioNumerics version 4.01 using the Pearson correlation coefficient and the UPGMA algorithm.

Fluorescent Amplified Fragment Length Polymorphism (FAFLP). FAFLP template preparation and PCR reactions were performed as described by Thompson *et al.*[31]. The sequencing products were analyzed by capillary electrophoresis as previously described by Storms *et al.*[31]. Cluster analysis of FAFLP patterns was done by analyzing the profiles with BioNumerics version 4.01 using the Dice coefficient and the Ward algorithm.

RESULTS AND DISCUSSION

In this work, the exploration of Raman spectroscopy as a taxonomic tool has been approached twofold. First, using cluster analysis, the groupings as obtained based on RSA as well as on FAFLP analysis are compared. Since the resulting clusters are affected by the selected linkage algorithm (e.g. UPGMA, Ward, etc.), these should be regarded as 'secondary data' while the original similarity matrices constitute 'primary data'[27]. Therefore, besides the comparison of delineated clusters, a correlation study is performed on the similarity matrices obtained via RSA, FAFLP, rep-PCR and FAME.

Raman spectrometric analysis. Figure 1 shows the dendrogram as generated using the Ward-linkage algorithm on the similarity matrices (Canberra metric). Four major clusters (A-D) delineated at -40 (Ward scale) are further subdivided in sub-clusters at -20 (Ward scale) thus resulting in eight groups (A, B_1, B_2, B_3, C_1, C_2, D_1 and D_2) that are in composition generally in good agreement with those obtained by FAFLP (Figure 2). For FAFLP, clusters delineated at 30 % similarity are indicated as 1-5. Subgroups were delineated at 50 % similarity (Dice coefficient). In addition, strains that showed high similarities, based on rep-PCR fingerprinting, generally clustered together in the Raman analysis (not shown).

In what follows, the Raman clusters are discussed.

Raman cluster A is somewhat more separated from the other Raman clusters B-D and encompasses the enterotoxin producing strains *B. cereus* LMG 22733, *B. cereus* LMG 22729 and *B. cereus* LMG 22728, that were isolated from pasta salad; R-22297 isolated from milk (Belgium) and a series of strains with the same rep-PCR fingerprints isolated from a gelatin producing factory in the USA. These gelatin isolates do not group with other '*B. cereus* group' strains isolated from gelatin producing factories in Europe. All strains from Raman cluster A group together in FAFLP sub-cluster 4a and based on their Fatty Acid (FAME) profiles in X_1. Here, it can already be noted that FAME-analysis does not differentiate between FAFLP groups 1, 2 and 4, in contrast to Raman spectroscopic analysis (C_1, D_1, A and B, Table 1).

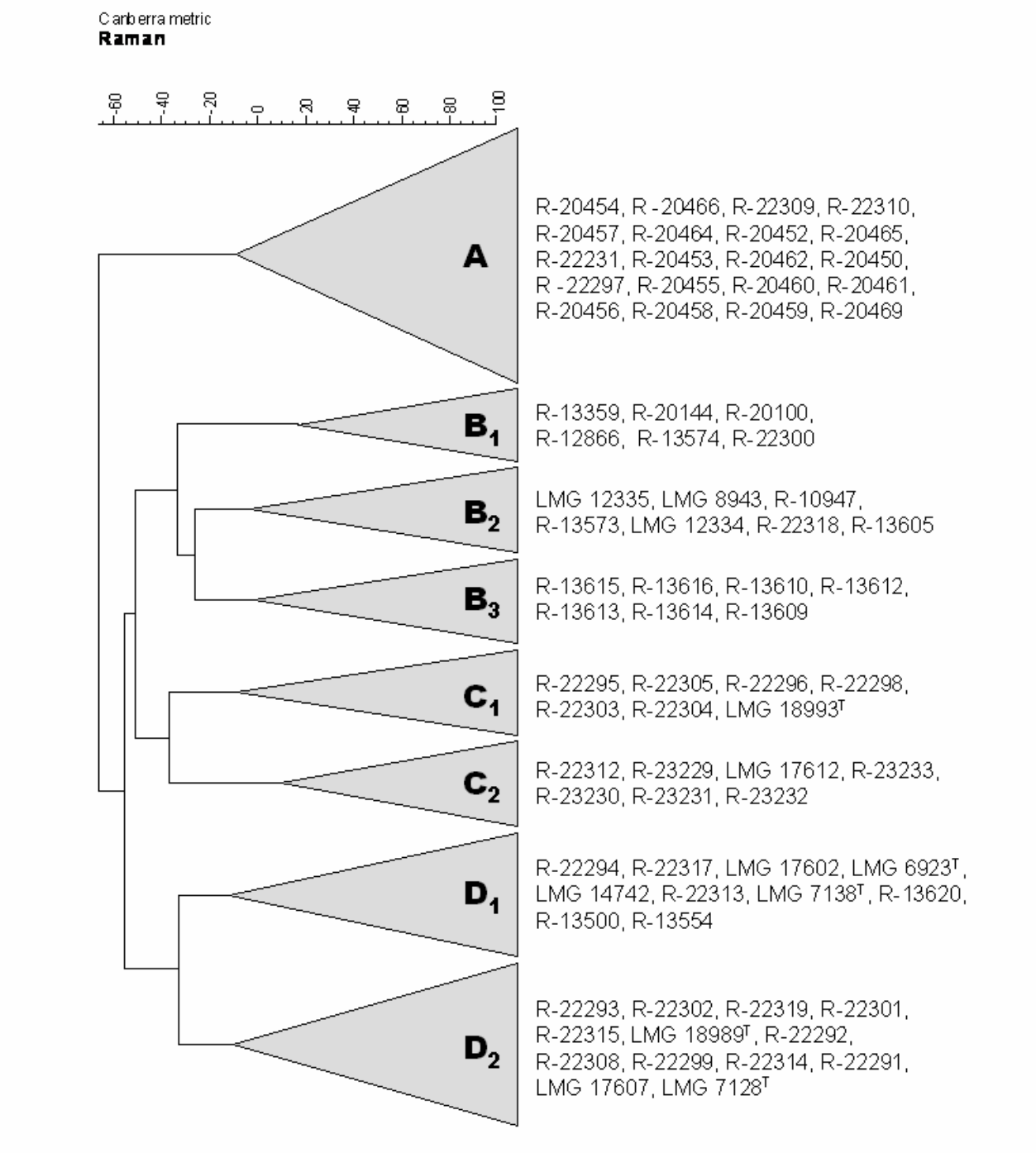


Figure 1: Dendrogram of the obtained Raman patterns using Ward's algorithm (see text for details). Major clusters (A –D) were delineated at -40 (Ward scale); sub-clusters at – 20. Sub-clusters are marked with an index (1-3).

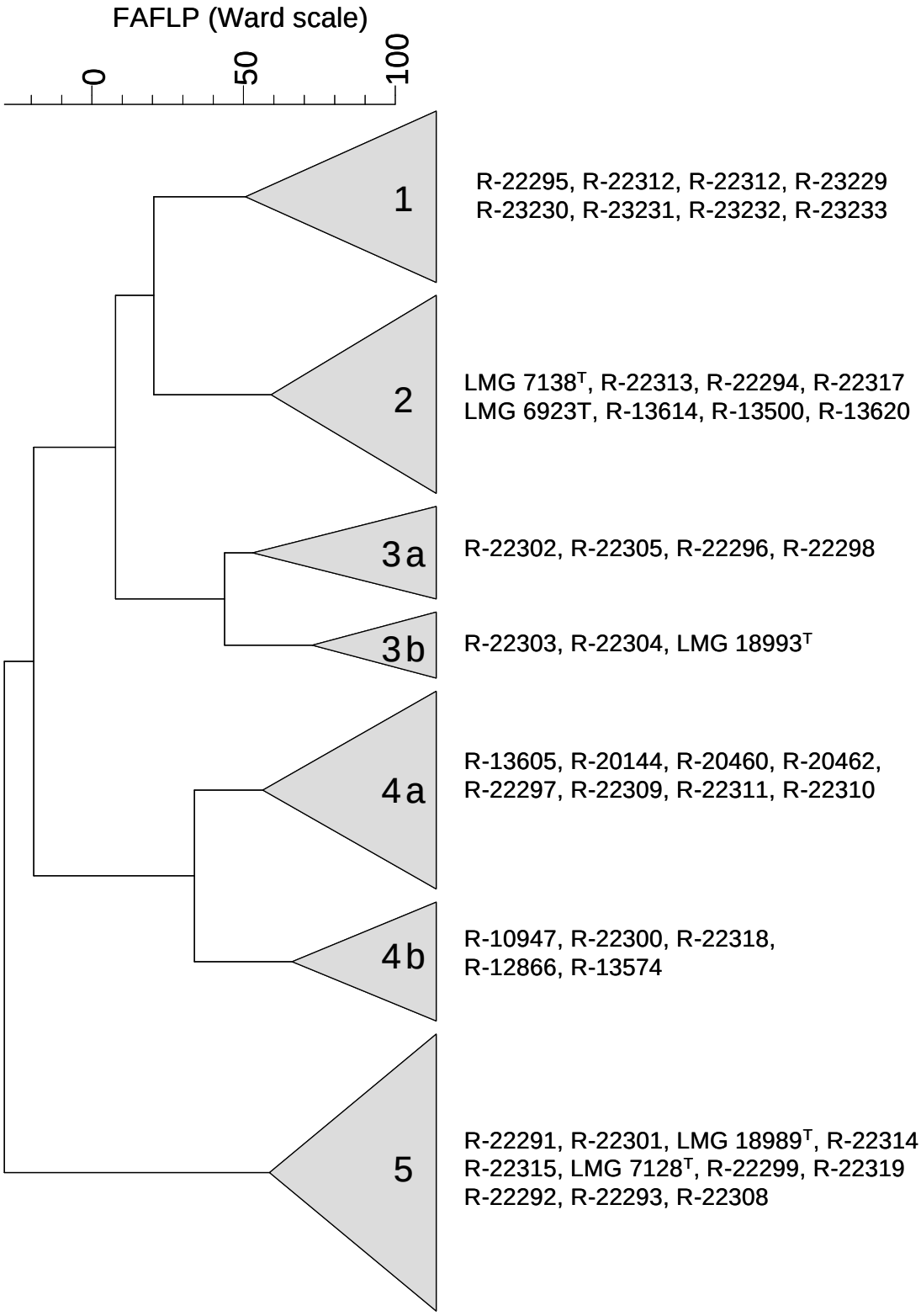


Figure 2: Dendrogram of the obtained FAFLP profiles using Ward's algorithm (see text for details). Major clusters are indicated as 1 – 5 and delineated at 30 % (Ward scale); sub-clusters are delineated at 50 %. And have an additional small character (a and b).

In general strains with the same rep-PCR profiles group together in the delineated Raman sub-clusters (Table 1). However, a group of five Belgian gelatin isolates is distributed over two Raman sub-clusters B_1 (R-13559, R-13574 and R-12866) and B_2 (R-10947 and R-13573). The other strains attributed to sub-clusters B_1 and B_2 show heterogeneous FAME-profiles and appear in four different FAME-clusters (X_1, X_2, X_3 and X_6) and two different FAFLP-clusters (4a and 4b).

Raman sub-cluster B_2 includes *B. cereus* LMG 8943, *B. cereus* LMG 12334, *B. cereus* LMG 12335 and *B. cereus* LMG 17613 (Figure 1), together with three isolates (R-10947, R-13573, and R-13605) from gelatin producing factories (Belgium/France).
Sub-cluster B_3 encloses seven isolates (R-13609, R-13610, R-13612 to R-13616) from gelatin (Belgium) that have similar rep-PCR profiles and the constitution of this sub-cluster is in good agreement with those obtained via FAFLP (cluster 2) and FAME (sub-cluster X_3; Table 1). However, based on their Raman fingerprints, these strains form a sub-cluster clearly separated from the other strains assigned to FAFLP cluster 2.

A third large delineation concerns Raman cluster C and its two sub-clusters (C_1 and C_2). Sub-cluster C_1 includes *B. pseudomycoides* LMG 18993^T, four isolates from raw milk (R-22296, *B. cereus* R-22298, *B. pseudomycoides* R-22304, R-22305) and two isolates from pasteurized milk (*B. pseudomycoides* R-22303 and R-22295). The delineation of this Raman sub-cluster is supported both by the FAFLP-analysis (cluster 3) and by the fatty acid composition of these strains (FAME-cluster Y) and is an indication for the application of RSA in taxonomic studies. In FAFLP, cluster 3 is subdivided in two sub-clusters. From the identification it can be concluded that cluster 3b contains strains of *B. pseudomycoides*, while cluster 3a contains *B. cereus* strains. This distinction cannot be made from the RSA cluster analysis.

It is noticed that strain R-22305 clusters separately, based on the FAME profiles, while its assignment to Raman sub-cluster C_1 is in correspondence with its position in FAFLP-group 3.
Raman sub-cluster C_2 encompasses *B. cereus* LMG 17612 and together with six isolates. All strains assigned to FAFLP cluster 1, cluster in Raman sub-cluster C_2, with exception of R-22295 that groups in sub-cluster C_1, again demonstrating the good agreement between RSA and FAFLP.

Raman cluster D is the fourth main cluster, delineated at -40 (Ward scale) and divided in sub-clusters D_1 and D_2. Raman sub-cluster D_1 encloses *B. cereus* LMG 6923^T, *B. cereus* LMG 14742, *B. cereus* LMG 17602 and *B. thuringiensis* LMG 7138^T together with six isolates from diverse origin (gelatin, raw milk, boiled pasta in food borne family outbreak). The constitution of this Raman sub-cluster is in good agreement with FAFLP cluster 2. Based on their FAME profiles, these strains group mainly together in FAME sub-cluster X_3. However, as stated before, FAME-analysis is not able to discriminate between FAFLP groups 1, 2 and 4 (Table 1). Isolates R-13500 and R-13554 are, in concordance with rep-PCR analysis (Table 1), placed in the same sub-cluster by RSA.

Raman sub-cluster D_2 includes *B. weihenstephanensis* LMG 18989^T *B. weihenstephanensis* R-22314, *B. weihenstephanensis* R-22315, *B. weihenstephanensis* R-22319, *B. mycoides* LMG 7128^T, *B. cereus* LMG 17607 and six additional isolates attributed to *B. mycoides* / *B. weihenstephanensis* by various techniques. The grouping of *B. cereus* LMG 17607 in an exclusive *B. mycoides* / *B. weihenstephanensis* cluster is based only on RSA. This rather unexpected Raman result is however not in contradiction with real-time PCR data (De Clerck, E., PhD Thesis) that illustrate a weak positive reaction of this strain with a *B. cereus* Taqman probe. Further research is needed for the finalization of the identification of this strain that most probably belong to the *B. mycoides* / *B. weihenstephanensis* subgroup. The majority of these isolates originate from milk (Belgium / Norway). Again, strains R-22292 and R-22293, considered as identical based on rep-PCR analysis; group together in the same Raman cluster.

The observation that *B. mycoides* LMG 7128^T and *B. weihenstephanensis* LMG 18989^T cluster together in RSA is in good agreement with the clustering obtained by both FAFLP and FAME analysis and with the known very close relationship between both taxa (16S rDNA sequences of the strains with accession numbers AB021191 (*B. mycoides*^{T}) and AB02199 (*B. wheihenstephanensis*^{T}) show 100 % similarity). Strain R-22302 is most closely related to the type strain of *B. pseudomycoides* according to FAFLP (Table 1); in Raman analysis this strain falls in the direct vicinity of *B. mycoides* / *B. weihenstephanensis* (cluster D_2 – Table 1). Its correct identification needs further characterization.

The delineation of the Raman clusters at -20 (Ward scale) is thus probably significant as can be demonstrated repeatedly by FAME, FAFLP and rep-PCR analysis. Hence, these results are in favor of appointing a taxonomic significance to the delineated Raman data.

In agreement with the clusters obtained via FAME and FAFLP, milk-isolates, emetic strains (isolated from pasta (Belgium)), as well as gelatin isolates are spread over diverse Raman clusters. However, the strains isolated from a gelatin-producing factory in the USA (Raman cluster A), are clearly separated from the ones isolated in Belgium or France (Raman clusters B_{1-3} and D_1).
The strains of FAFLP cluster 4 are spread in the Raman clustering over A, B_1 and B_2. Also in the FAME analysis these strains exhibit heterogeneous profiles. In this particular case, the heterogeneity of the cell wall composition is reflected in the overall Raman spectra.

The differences between the Raman clusters (A–D) obtained by statistical analysis, are illustrated by the Raman spectra of selected representative strains. If possible LMG-collection strains were preferred as representative. The most important spectral differences between the obtained Raman clusters (Figure 3) are highlighted.

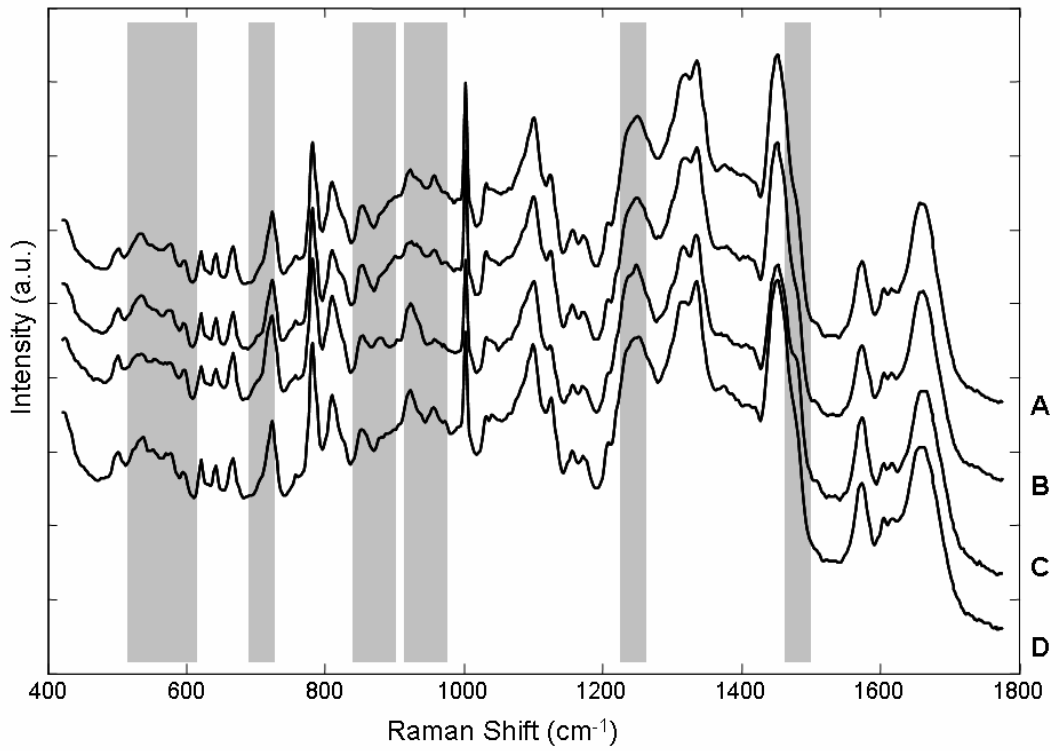


Figure 3: Raman spectra of selected representative strains of the four major Raman clusters (A – D) as defined in figure 1. If possible spectra of LMG-collection strains are shown. A, R-22310 (Raman cluster A); B, *B. cereus* LMG 8943 (Raman cluster B); C, *B. pseudomycoides* LMG 18993^T (Raman cluster C); D, *B. cereus* LMG 6923^T (Raman cluster D). Remarkable differences are highlighted. Spectra are shifted vertically for clarity. (a.u.) arbitrary units.

Correlation between similarity matrices using Kendall's Tau coefficient. So far, the exploration of RSA as a potential tool for bacterial taxonomy was based on the comparison of the clustering obtained by numerical analysis of FAME and FAFLP data. However, clusters are affected by the linkage algorithm (e.g. UPGMA, Ward) used and thus can be considered as 'secondary data'[27]. To objectify the congruence in the numerical interpretation of the data from various methods, a consensus study on the primary data (i.e. the similarity matrices) was performed using the non-parametric Kendall's Tau coefficient.

Table 2: Overview of Kendall's Tau coefficients ($\alpha < 0.0001$) obtained by comparison of the similarity matrices for the different techniques. Fatty acid methyl ester (FAME) analysis, Raman spectroscopic analysis (RAMAN), Fluorescent Amplified Length polymorphism (FAFLP) analysis and rep-PCR analysis using the $(GTG)_5$ primer ($(GTG)_5$).

		FAFLP	RAMAN	$(GTG)_5$	FAME
Kendall's Tau coefficients					
	FAFLP	**100**			
	RAMAN	32.5	**100**		
	$(GTG)_5$	19.0	16.3	**100**	
	FAME	28.7	19.4	24.7	**100**

For the comparison of different fingerprinting techniques, the Kendall's Tau coefficient is preferred to the Pearson correlation coefficient because (1) this coefficient is robust towards outliers, (2) deals with eventual non-linear correlations (this is often the case for the relationship between similarity values for two different approaches) and (3) no statistical assumptions need to be fulfilled since it concerns a non-parametric method.

An overview of the Kendall's Tau correlation coefficients obtained between the different techniques is given in Table 2. These results offer a more objective interpretation of the correlation between Raman patterns at the one hand and phenotypic (FAME) and genomic (FAFLP) data at the other, since no bias or interference is introduced by a linkage algorithm. The observed congruence of RSA with FAFLP was found to be highly significant ($\alpha < 0.0001$). Of the techniques compared, Raman and FAFLP show the highest Kendall's tau coefficient (32.5). The comparison of Kendall tau coefficients of some techniques that are commonly applied in taxonomical studies (FAME, rep-PCR and FAFLP) learns that none of them have a Kendall's tau coefficient that exceeds the one obtained by comparing Raman and FAFLP.

Regression analysis. Since significant correlation ($\alpha < 0.0001$) was observed between the Raman spectroscopic pattern similarities and the similarities obtained by FAFLP, it was attempted to describe their relationship. A second degree polynomial was fitted to the data, subject to the constraints of being monotonically increasing and passing trough the (100%, 100%) position as described[27] and implemented in the BioNumerics 4.01 software. The similarity values of Raman patterns (Canberra metric) were plotted against the Dice coefficient similarities of FAFLP fingerprints (Figure 4). The regression illustrates a good concordance between both techniques, which is confirmed by the determination coefficient of the obtained curve ($r = 78.8$). However, since the value of r is too low to be of analytical importance, it is concluded that Raman spectroscopy should not be considered as a substitute for FAFLP, but rather as an interesting precursor to FAFLP or an additional tool in polyphasic taxonomy.

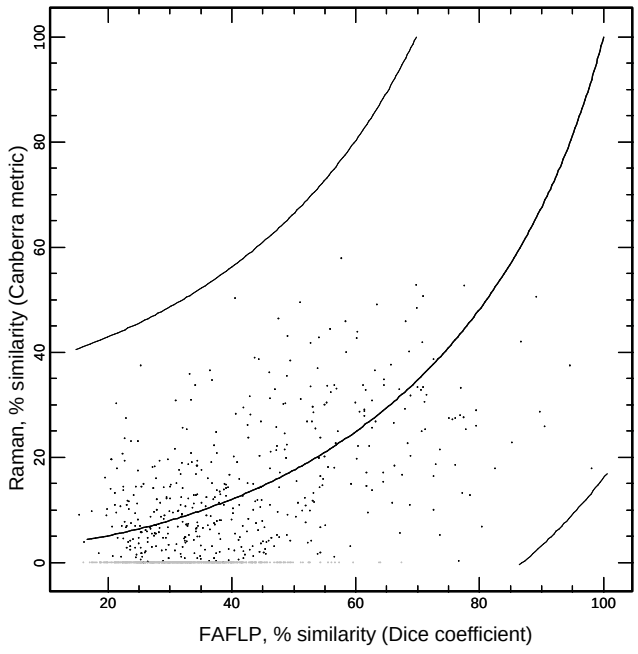


Figure 4: The similarity values of Raman patterns (Canberra metric) were plotted against the Dice coefficient similarities between FAFLP fingerprints. Confidence intervals are as indicated. Determination coefficient (r) = 78.8.

Conclusion

Raman spectrometric analysis (RSA) yields a clustering of strains belonging to the '*Bacillus cereus* group' that are in good agreement with those obtained from FAFLP and FAME data. Despite some inconsistencies of clustering of individual strains, well-separated Raman clusters and sub-clusters were obtained, delineated at – 40 and -20 (Ward Scale), respectively.

These conclusions are supported by other studies, where vibrational techniques (FT-IR and Raman) are combined with genomic fingerprinting techniques. Ehling *et al.*[6] found also similar groupings based on FT-IR spectroscopy, exoprotein profiling and RAPD within *B. cereus* and Maquelin *et al.*[19] showed that the groupings obtained by Raman spectroscopy are highly similar to the one obtained by AFLP within *Acinetobacter* (*Acinetobacter*-complex). In none of these studies, a complete identical topography was obtained when dendrograms of data from two different techniques were compared. This was also not the case in the present study.

However, when the primal similarity matrices of Raman spectroscopy and FAFLP were compared, a highly significant ($\alpha<0.0001$) but rather low correlation between both techniques was found. Regression analysis confirmed that Raman spectroscopy is not a substitute for FAFLP. However, this should not be regarded as a restriction. Indeed, (1) in general similar groupings as in FAFLP are obtained by Raman spectroscopy, being a less labor intensive method and (2) the composition of FAFLP and Raman clusters are not identical because additional discriminative information is revealed by polyphasic studies.

Hence, Raman spectroscopy allows retrieving reliable and manageable groups prior to or in combination with polyphasic approaches herewith reducing the need for more laborious techniques (e.g. FAFLP). A highly significant ($\alpha < 0.0001$) correlation between FAFLP and Raman spectroscopy gives further evidence that the discriminatory power of vibrational techniques as Raman spectroscopy lays even at the subspecies level as was suggested previously[12].

Acknowledgements

The authors are grateful to the Research Foundation - Flanders (FWO-Vlaanderen) for financial support (Grants G.0156.02 and G.0343.03) of this research project. D.H. greatly acknowledges the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for his doctoral grant. J.H. and P.V. wish to thank the FWO-Vlaanderen for their postdoctoral fellowships.

Reference List

1. **Chang, Y. H., Y. H. Shangkuan, H. C. Lin, and H. W. Liu**. 2003. PCR assay of the groEL gene for detection and differentiation of *Bacillus cereus* group cells. Appl Environ Microbiol **69**:4502-4510.

2. **Choo-Smith, L. P., K. Maquelin, T. van Vreeswijk, H. A. Bruining, G. J. Puppels, N. A. Ngo Thi, C. Kirschner, D. Naumann, D. Ami, A. M. Villa, F. Orsini, S. M. Doglia, H. Lamfarraj, G. D. Sockalingum, M. Manfait, P. Allouch, and H. P. Endtz**. 2001. Investigating microbial (micro)colony heterogeneity by vibrational spectroscopy. Appl Environ Microbiol **67**:1461-1469.

3. **Curk, M.C., F. Peladan, and J.C. Hubert.** 1994. Fourier-Transform Infrared (FTIR) Spectroscopy for identifying *Lactobacillus* species. FEMS Microbiol Lett **123**:241-248.

4. **De Clerck, E. and P. De Vos**. 2002. Study of the bacterial load in a gelatine production process focused on *Bacillus* and related endosporeforming genera. Syst Appl Microbiol **25**:611-617.

5. **De Clerck, E., T. Vanhoutte, T. Hebb, J. Geerinck, J. Devos, and P. De Vos.** 2004. Isolation, characterization, and identification of bacterial contaminants in semifinal gelatin extracts. Appl Environ Microbiol **70**:3664-3672.

6. **Ehling-Schulz, M., B. Svensson, M. H. Guinebretiere, T. Lindback, M. Andersson, A. Schulz, M. Fricker, A. Christiansson, P. E. Granum, E. Martlbauer, C. Nguyen-The, M. Salkinoja-Salonen, and S. Scherer**. 2005. Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. Microbiology **151**:183-197.

7. **Helgason, E., N. J. Tourasse, R. Meisal, D. A. Caugant, and A. B. Kolsto**. 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl Environ Microbiol **70**:191-201.

8. **Helm, D., H. Labischinski, G. Schallehn, and D. Naumann**. 1991. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. J Gen Microbiol **137**:69-79.

9. **Heyndrickx, M., L. Vauterin, . P. Vandamme, K. Kersters, and P. De Vos.** 1996. Applicability of combined amplified ribosomal DNA restriction analysis (ARDRA) patterns in bacterial phylogeny and taxonomy. J Microbiol Methods **26**:247-259.

10. **Heyrman, J., N. A. Logan, H. J. Busse, A. Balcaen, L. Lebbe, M. Rodriguez-Diaz, J. Swings, and P. De Vos**. 2003. *Virgibacillus carmonensis* sp. nov., *Virgibacillus necropolis* sp. nov. and *Virgibacillus picturae* sp. nov., three novel species isolated from deteriorated mural paintings, transfer of the species of the genus *salibacillus* to *Virgibacillus*, as *Virgibacillus marismortui* comb. nov. and *Virgibacillus salexigens* comb. nov., and emended description of the genus *Virgibacillus.* Int J Syst Evol Microbiol **53**:501-511.

11. **Heyrman, J., J. Mergaert, R. Denys, and J. Swings.** 1999. The use of fatty acid methyl ester analysis (FAME) for the identification of heterotrophic bacteria present on three mural paintings showing severe damage by microorganisms. FEMS Microbiol Lett: **181**:55-62.

12. **Hutsebaut, D., J. Vandroemme,. J. Heyrman, P. Dawyndt, P. Vandenabeele, L. Moens,. and. P. De Vos.** 2005. Raman microspectroscopy as an identification tool within the phylogenetically homogeneous '*Bacillus subtilis*'-group. Syst Appl Microbiol (submitted) .

13. **Hutsebaut, D., P. Vandenabeele, and L. Moens.** 2005. Evaluation of an Accurate Calibration and Spectral Standardization Procedure for Raman Spectroscopy. The Analyst **130**: 1204 -1214.

14. **Ibelings, M. S., K. Maquelin, H. P. Endtz, H. A. Bruining, and G. J. Puppels**. 2005. Rapid identification of *Candida* spp. in peritonitis patients by Raman spectroscopy. Clin Microbiol Infect **11**:353-358.

15. **Jarvis, R. M. and R. Goodacre**. 2004. Discrimination of bacteria using surface-enhanced Raman spectroscopy. Anal Chem **76**:40-47.

16. **Kirschner, C., K. Maquelin, P. Pina, N. A. Ngo Thi, L. P. Choo-Smith, G. D. Sockalingum, C. Sandt, D. Ami, F. Orsini, S. M. Doglia, P. Allouch, M. Mainfait, G. J. Puppels, and D. Naumann**. 2001. Classification and identification of enterococci: a comparative phenotypic, genotypic, and vibrational spectroscopic study. J Clin Microbiol **39**:1763-1770.

17. **Lemmer, K., D. Naumann, B. Raddatz, and K. Tintelnot**. 2004. Molecular typing of *Cryptococcus neoformans* by PCR fingerprinting, in comparison with serotyping and Fourier transform infrared-spectroscopy-based phenotyping. Med Mycol **42**:135-147.

18. **Maquelin, K., L. P. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H. A. Bruining, and G. J. Puppels**. 2000. Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. Anal Chem **72**:12-19.

19. **Maquelin, K., L. Dijkshoorn, T. J. van der Reijden, and G. J. Puppels**. 2005. Rapid epidemiological analysis of *Acinetobacter* strains by Raman spectroscopy. J Microbiol Methods: .in press

20. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. A. Ngo-Thi, T. van Vreeswijk, M. Stammler, H. P. Endtz, H. A. Bruining, D. Naumann, and G. J. Puppels**. 2003. Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. J Clin Microbiol **41**:324-329.

21. **Maquelin, K., C. Kirschner, L. P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, and G. J. Puppels**. 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J Microbiol Methods **51**:255-271.

22. **Nakamura, L. K.** 1998. *Bacillus pseudomycoides* sp. nov. Int J Syst Bacteriol **48**:1031-1035.

23. **Naumann, D., D. Helm, and H. Labischinski**. 1991. Microbiological characterizations by FT-IR spectroscopy. Nature **351**:81-82.

24. **Oberreuter, H., J. Charzinski, and S. Scherer**. 2002. Intraspecific diversity of *Brevibacterium linens*, *Corynebacterium glutamicum* and *Rhodococcus erythropolis* based on partial 16S rDNA sequence analysis and Fourier-transform infrared (FT-IR) spectroscopy. Microbiology **148**:1523-1532.

25. **Oust, A., T. Moretro, C. Kirschner, J. A. Narvhus, and A. Kohler**. 2004. FT-IR spectroscopy for identification of closely related *Lactobacilli*. J Microbiol Methods **59**:149-162.

26. **Pitcher, D. G., N. A. Saunders, and R. J. Owen.** 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. **8** :151-156.

27. **Rademaker, J. L., B. Hoste, F. J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, and F. J. de Bruijn**. 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. Int J Syst Evol Microbiol **50**:665-677.

28. **Rademaker, J. L. W., and F.J. de Bruijn.** 1997. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis. p. 151-171. *In : DNA Markers: Protocols, Applications and Overviews.* G. Gaetano-Anollés and P.M. Gresshoff (Eds.), New York: Wiley

29. **Sandt, C., G. D. Sockalingum, D. Aubert, H. Lepan, C. Lepouse, M. Jaussaud, A. Leon, J. M. Pinon, M. Manfait, and D. Toubas**. 2003. Use of Fourier-transform infrared spectroscopy for typing of *Candida albicans* strains isolated in intensive care units. J Clin Microbiol **41**:954-959.

30. **Stenfors, L. P. and P. E. Granum**. 2001. Psychrotolerant species from the *Bacillus cereus* group are not necessarily *Bacillus weihenstephanensis*. FEMS Microbiol Lett **197**:223-228.

31. **Storms, V., M. Baele, R. Coopman, A. Willems, T. de Baere, F. Haesebrouck, G. Verschraegen, M. Gillis, and M. Vaneechoutte**. 2002. Study of the intra- and interlaboratory reproducibility of partial single base C-sequencing of the 16S rRNA gene and its applicability for the identification of members of the genus *Streptococcus*. Syst Appl Microbiol **25**:52-59.

32. **Thompson, F. L., B. Hoste, K. Vandemeulebroecke, and J. Swings**. 2001. Genomic diversity amongst *Vibrio* isolates from different sources determined by fluorescent amplified fragment length polymorphism. Syst Appl Microbiol **24**:520-538.

33. **Vandeginste, B. G. M., D. L. Massart, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers –Verbeke.** 1998. *In*: B.G.M. Vandeginste, S.C. Rutan (Eds.), Handbook of Chemometrics and Qualimetrics: Part B, Elsevier Science, Amsterdam,

34. **Versalovic, J., M. Schneider, F. J De Bruijn, and J. R. Lupski,** 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol Cell Biol **5**:25-40.

35. **Wenning, M., H. Seiler, and S. Scherer**. 2002. Fourier-transform infrared microspectroscopy, a novel and rapid tool for identification of yeasts. Appl Environ Microbiol **68**:4717-4721.

Discussion

Chapter VI

VI. Discussion

Generally, Raman spectroscopy distinguishes itself from other techniques currently used by it is ease-of-use, its low cost, high speed of analysis and its broad informational content. In particular, it is the ability to provide a complete chemical '*picture*' of micro-organisms that is, at the same time, the weakness and strength of Raman spectroscopy. Diverse taxonomic relevant information is achieved by one single technique, while else multiple techniques are required to analyze proteins (SDS-PAGE), fatty acids (FAME), nucleic acids (16S rDNA sequencing, rep-PCR, DNA-DNA hybridizations, etc.) and other chemotaxonomic markers. As a consequence, Raman spectroscopy is pushed forward as a powerful and interesting tool for microbial analysis because of the following reasons:

1. An important reduction in the total number of analyses, cost of analysis and overall time of accurate analysis is achieved.

2. Raman spectroscopy does not require time-consuming sample preparation or (hazardous) chemical reagents (e.g. ethidium bromide, organic solvents).

3. Apart from the initial investment, Raman spectroscopy is a relatively cheap technique for routine analysis since no reagents, except from culture media, are needed. Therefore, Raman spectroscopy can be considered as superior even to other spectroscopic techniques as FT-IR spectroscopy and MALDI-TOF mass spectroscopy.

4. By combining multiple discriminating features, an accurate identification may be achieved even of highly related species, that are hard (or impossible) to discriminate by conventional techniques.

However, the obtained Raman profiles are, exactly because of their high informational content, very complex and special attention should be paid to numerical analysis. Because taxonomic non-relevant molecules (e.g. carotenoid structures) and variations in growth conditions (including medium composition) affect the Raman profiles, more advanced statistical techniques are needed to separate useful information from invalid information. These so called 'supervised techniques' are mathematical tools allowing correct identification based on *a priori* knowledge of the strains' identity. The latter is, from a microbiological point of view, not always available.

Moreover, in order to be generally applicable, the defined groups need to be representative for the species included, requiring a large number of strains that represent biological, ecological as well as genomic diversity. Supervised approaches (e.g. LDA) and thus reliable and representative Raman databases ask for a profound, polyphasic characterization of the included strains.

In theory, the problem of unreliable *a priori* knowledge is easily tackled by using unsupervised techniques as cluster analysis or by matching profiles towards a reference database based on a similarity measure. These approaches are commonly used in taxonomic studies. However, for Raman spectroscopy some problems might arise.

First, Raman profiles depend on the presence of specific molecules even if they are not of taxonomic importance and are sensitive to variations in growth conditions (e.g. medium composition). The latter is a typical drawback of phenotypic techniques and explains partly the success of genotypic (molecular) techniques as rep-PCR, 16S-rDNA sequencing and FAFLP and hampers the delineation of clusters of taxonomic value.

Second, even if such clusters were obtained, a similarity cut-off value should be defined at which different species can be discriminated. This approach is well-known for e.g. DNA-DNA hybridization studies and 16S rDNA sequence analysis but has never been used for Raman analysis so far. These cut-off levels probably need to be determined per taxon or group of taxa and might not hold in general. Species delineations from Raman clusterings need thus to be correlated with DNA-DNA hybridization studies for final taxonomic conclusions.

Both factors hamper the breakthrough of Raman spectroscopy as a taxonomic tool but does not affect its application in specific identification scheme's within well delineated taxa.

Taken together, Raman spectroscopy can be considered as a powerful tool for the fast and accurate identification of microorganisms, even when they are phylogenetically highly related, provided that (1) an extensive database is available, (2) that species are represented by multiple strains, polyphasically characterized and covering the biological, ecological and genotypic heterogeneity of the species.

As an application in taxonomic studies and/or in situations were no reliable *a priori* knowledge is available, Raman spectroscopy can be considered as a valuable, additional tool of which the results should always be interpreted in combination with other, preferably molecular, techniques as FAFLP, rep-PCR, DNA-DNA hybridization etc. These other techniques are often time consuming and more expensive. Raman spectroscopy can therefore be considered as an interesting first approach to obtain a manageable grouping of large sets of strains prior to more extensive studies of a reduced number of representatives with more powerful (molecular) techniques.

Summary & Future prospects

Chapter VII

VII. Summary and Future Prospects

Summary

Raman spectroscopy is a fast and easy-to-use analytical technique and has recently been introduced as new microbial identification tool. The technique is based on the inelastic scattering of light during which molecules are excited to a virtual energy level, and upon relaxation the excess of energy is released as radiation. When using (Stokes) Raman spectroscopy the molecules end in an excited vibrational state and the scattered radiation has a longer wavelength than the incident laser beam. The resulting Raman spectrum is highly specific and serves as a fingerprint, unique for the specific molecule analyzed. In the case of micro-organisms, the spectrum reflects the complete molecular microbial set-up, that is determined partly by stable taxonomic markers (e.g. fatty-acids, proteins, poly-amines, etc.), partly by non-species-descriptive molecules (e.g. carotenoid structures) and partly by growth parameters (e.g. medium composition, incubation temperature, growth time). The recent publications on the identification of micro-organisms with Raman spectroscopy taking into account, this thesis evaluates the applicability of this technique for the genus *Bacillus*. Special attention was paid to the standardization of the approach and the selection of well-chosen species (model systems) in order to obtain a realistic view on the possibilities of Raman spectroscopy as a future taxonomic and identification tool. In the first introductory chapter a general overview is given of the most important (chemo)taxonomic markers with their significance for modern bacterial taxonomy. Furthermore, a concise theoretical background on Raman spectroscopy is given. The introduction is concluded by outlining the current possibilities of vibrational spectroscopy (FT-IR and Raman) for microbiological applications. Subsequent chapters describe the different steps to evaluate Raman spectroscopy as a microbial identification tool. Attention is paid to the instrumental calibration, the standardization of the spectra, the proper selection of growth conditions, the profound evaluation of the identification accuracy of the proposed method and the exploration of its taxonomic potential.

When Raman spectra are highly similar, as is the case for closely related bacteria, instrumental stability becomes a main issue. The possibilities of Raman spectroscopy cannot be limited by artefacts such as spectral shifts or changes in transmission efficiencies. Hence, a protocol for both Raman wavenumber and intensity calibration of the Raman set-up is required. This is described in Chapter II. The principles outlined in this chapter resulted in a calibration module written in Matlab, that is applied in a standard way in the laboratory of Analytical Chemistry for Raman spectroscopic applications in (micro)biology (e.g. study of biochemical pathways and fungi). To correct for spectral shifts, a method is described by which Raman spectra are standardized prior to multivariate analysis.

Hence, the long-term stability of the spectral database is guaranteed. However, before this database can be extended to Raman fingerprints of micro-organisms, a clear understanding of the reproducibility and of the taxonomic resolution is needed. As stated before, Raman spectroscopy reveals a *complete* molecular fingerprint of micro-organisms. Therefore, it can be expected that the reproducibility of these fingerprints and their discrimination ability can be affected by growth conditions. The effect of growth conditions on the achievable taxonomic resolution of Raman spectroscopy is discussed in Chapter III by means of three phenotypically different species (*B. cereus*, *B. licheniformis* and *B. pumilus*).

A wide variety of growth conditions was applied and several complications came forward:

1. Since (little) biomass is needed, Raman spectroscopic analysis, as any phenotypic technique, is restricted to strains that are cultivable. Hence, growth conditions (e.g. medium, incubation temperature and time) should be selected to be as widely applicable as possible. As Raman spectra are affected by growth conditions, the latter must be well-chosen and standardized to achieve the best taxonomic resolution.

2. Extracellular slime (e.g. extracellular polysaccharides, capsules) hinder the collection of Raman spectra. This phenomenon can be avoided by the selection of proper growth parameters.

3. At specific growth conditions, some strains might stick to the growth medium, making it impossible to harvest and analyze them using a standardized protocol. This problem can be solved by the selection of proper growth parameters.

In addition, the results indicated that identification at the species level (by linear discriminant analysis - LDA) was less sensitive to growth conditions, while groupings of strains (relying *only* on cluster analysis) were strongly affected by growth parameters as medium, incubation time and temperature). This supports their importance to achieve the best taxonomic resolution. However, speed and comfort are imposing factors to determine the incubation time and the selection of a proper medium facilitates harvesting of bacterial colonies. After comparison of multiple growth conditions, Brain Heart Infusion (BHI) was selected as the standard culture medium in combination with 7 hours incubation at 37°C.

After standardization of the calibration protocol and growth conditions, Chapter IV reports on the validation of the developed methodology by a first model system. This model system consists of eight, highly related *Bacillus* species that are assigned to the homogenous '*Bacillus subtilis*'- group or that are closely related to it (*B. pumilus, B. licheniformis* and *B. sonorensis)* represented by 69 strains. Based on small but reproducible spectral differences, species identification could be established by a supervised LDA-based approach. Despite the high relatedness of the species, promising identification accuracies were obtained, both by internal (> 90%) and external (blind) validation (> 78%). Raman spectroscopy is concluded to be a general identification tool as indicated by the correct identification of strains that were previously not represented in the spectral library. Difficulties and contradictions were only observed within taxa that are most probably not clearly separated.

So far, strains were successfully identified at the species level, based on a set of strains that were allocated to a restricted group of well-characterized, highly related species. The next step concerns a situation where no reliable *a priori* knowledge on the strain identity is available. Therefore, in Chapter V, the potential of Raman spectroscopy as a taxonomic tool is explored within the phylogenetically homogenous '*Bacillus cereus*'-group. Here, it is not aimed to clarify the complex taxonomic situation of this group, yet to investigate the taxonomic potential of Raman spectroscopy. In practice, it is investigated whether the clusters that are delineated by Raman spectroscopy correspond to those obtained by other techniques used in taxonomy (e.g. FAME, FAFLP). It is found that the groupings based on the similarity matrix obtained by Raman spectroscopy show considerable correspondence to the ones obtained based on FAFLP-similarities. Isolates that were identified as identical based on rep-PCR analysis, clustered in the same group based on their Raman patterns, indicating for the application of Raman spectroscopy in taxonomic studies. However, since the topology of a dendrogram is highly affected by the used linkage algorithm (e.g. UPGMA, Ward), an

additional consensus study was performed to calculate the correlation between the original similarity matrices. A highly significant correlation was found between Raman spectroscopy and FAFLP. However, the correlation found was too low to conclude that Raman spectroscopy is able to replace FAFLP as a stand-alone typing technique.

In conclusion, Raman spectroscopy is an easy-to-use vibrational technique that requires a minimum of sample preparation and limited consumables (the culture medium) and that allows a fast (< 8 h, including growth) and accurate identification of bacteria within the frame of a strictly standardized protocol. This thesis has focused on a group of phylogenetically highly related and phenotypically highly similar species ('*B. subtilis*' –group) to evaluate the identification ability of Raman spectroscopy. However, for most applications (food, soil and water isolates; plant pathogens; clinical relevant organisms, etc.) such highly related species are seldom encountered. As such, it is credible that this technique can easily be used for identification purposes in these fields.
Furthermore, on condition of a dedicated optimization, nor additional reagents, nor highly skilled experts are needed to operate the Raman equipment. Although the instrumentation is not automated, per day, 30 – 40 samples can be analyzed, including growth, inoculation and sampling. The non-automated nature of Raman spectroscopy is one of the most important drawbacks, hampering the breakthrough of Raman spectroscopy in microbial laboratories. However, taking the high potential of Raman spectroscopy into account, this should be considered only as a matter of time and financial support.

"What the mind can conceive, the mind can achieve"

Clement Stone

Future prospects

In the following, a brief outline is given of some ideas, suggestions and critical remarks which result from the many questions that are left open after this PhD. Moreover, they support my faith in this technique for future applications in bacteriology and/or microbiology. Although this is not the most scientific part of this work, these remarks can contribute to the improvement and the further development of this technique to become a powerful microbial tool in microbiology.

1. The Raman effect is, based on its theoretical background, an inherently weak effect. Since closely related species differ by minor spectral details (cf. Chapter 4), superior Raman spectra are of primary importance for an accurate identification. Hence, the total Raman scattering intensity should be optimized, mainly by increasing the laser power at the sample and by achieving a better overall collection efficiency. This can result in a different Raman set-up than the one currently used. A Raman set-up should be fully optimized for a particular application and there is no such thing as a 'general' Raman set-up for analyzing 'just about anything'.

2. Alternative growth conditions can ensure a higher reproducibility, while biochemical differences between closely related species might become better expressed. These growth conditions should be practical in use (see Chapter 3) and as generally applicable as possible, allowing extension to other genera. Possibly, these growth conditions will cause strains to be analyzed during their stationary phase. However, a variety of growth conditions will be necessary to deal with, for example, thermopiles and psychrophiles.

3. A renewed approach on sampling (that is performed hitherto by a disposable plastic loop) can result in a higher reproducibility between successive measurements of the same strain. Alternatives should be devised and validated. If possible, the advantage of minimal sample preparation should be maintained. However, the informational content should always be the first concern.

4. It could be valuable to consider the use of polarization measurements to investigate whether or not the additional structural information results in a better discrimination and a finer taxonomic resolution.

5. A higher reproducibility of the (overall?) Raman fingerprints should ideally exclude supervised identification approaches (e.g. LDA) for bacterial identification. This would tackle the problem of non-descriptive Raman bands (e.g. carotenoid structures), the effect of misidentified strains on the training of supervised identification approaches and the possible misidentification of species that are not represented in the database. Once achieved, Raman spectroscopy could be used for taxonomic studies.

6. The current spectral library should be extended as such that the biological and ecological diversity of each species is represented. The inclusion of related genera (e.g. *Paenibacillus*, *Brevibacillus*, *Virgibacillus*, etc.) as well as selected representatives of Gram negative genera will allow the further evaluation of the technique and its application in various microbiological studies.

7. Raman spectroscopy should be the first choice for a fast and accurate identification.

In spite of these suggestions, there is no doubt that the method, as applied today in our laboratory, has proven to be sufficiently discriminative and reproducible. The possibility to discriminate highly related species together with a significant correlation with FAFLP and the comparable groupings obtained by both Raman spectroscopy and FAFLP encourages the further development of this technique for microbial applications. These should be further explored in various microbial studies, in Gram positive as well as Gram negative genera, aiming at identification, discrimination, screening, or at the collection of fundamental insights into biochemical pathways.

"There are no shortcuts to any place worth going"

Beverly Sills

Samenvatting

Raman spectroscopie is een snelle en eenvoudig te gebruiken analytische techniek, die recent werd geïntroduceerd als een nieuwe techniek voor de identificatie van bacteriën. De werking ervan steunt op de niet-elastische verstrooiing van licht. Tijdens dit proces worden moleculen geëxciteerd tot een virtueel energieniveau, waarna ze opnieuw terugvallen tot een energetisch meer stabiele toestand. Tijdens deze relaxatie wordt het verschil aan energie vrijgesteld door emissie van licht. In het geval van (Stokes) Raman strooiing, vallen de moleculen terug tot een vibrationeel geëxciteerd energieniveau, waarbij het verstrooide, geëmitteerde licht, een langere golflengte heeft (en dus een lagere energie) dan de oorspronkelijke lichtbundel. Het Raman-spectrum dat hieruit voortvloeit is zeer specifiek en dient als een *vingerafdruk* die uniek is voor de moleculen die werden geëxciteerd.

In het geval van micro-organismen geeft dit spectrum dus de volledige moleculaire samenstelling weer, die ten dele wordt bepaald door taxonomische merkers (vb. vetzuren, eiwitten, polyamines, enz.), ten dele door moleculen die niet speciespecifiek zijn (vb. carotenoïde structuren) en ten dele door groeiparameters (vb. samenstelling van het medium, temperatuur, tijd). Rekening houdend met de recente publicaties rond de identificatie van micro-organismen met Raman spectroscopie, evalueert deze scriptie de toepasbaarheid van deze techniek binnen het genus *Bacillus.* Om een realistisch beeld te verkrijgen van de mogelijkheden van Raman spectroscopie, als een toekomstige taxonomische en identificatie techniek, ging speciale aandacht uit naar de standaardisatie van de benadering en de keuze van goed gekozen species (model systemen) voor de evaluatie. In de eerste inleidende hoofdstukken wordt een algemeen overzicht gegeven van de belangrijkste (chemo)taxonomische merkers en hun betekenis voor de moderne bacteriële taxonomie. Daarnaast wordt een bondige, theoretische achtergrond gegeven van Raman spectroscopie. De inleiding wordt afgesloten met een overzicht van de huidige mogelijkheden van vibrationele spectroscopie (FT-IR and Raman) voor microbiologische toepassingen. In de volgende hoofdstukken worden de verschillende stappen ter evaluatie van Raman spectroscopie als een microbiologisch instrument beschreven. Er wordt aandacht besteed aan de kalibratie van de spectrometer, de standaardisatie van de spectra, de aangepaste selectie van de groeiomstandigheden, de doorgedreven evaluatie van de uiteindelijke identificatie en aan de verkenning van het taxonomisch potentieel van de voorgestelde techniek.

Wanneer Raman spectra sterk gelijkend zijn, zoals in het geval van nauw verwante bacteriën, wordt de instrumentele stabiliteit een belangrijk punt. Het is onaanvaardbaar dat de mogelijkheden van Raman spectroscopie zouden gelimiteerd worden door artefacten, zoals spectrale verschuivingen of wijzigingen in de transmissie efficiëntie van de gebruikte set-up. Om dit te voorkomen is een protocol noodzakelijk voor de kalibratie van zowel de Raman-Shift as als van de intensiteitsas. Dit wordt beschreven in hoofdstuk 2. De principes die in dit hoofdstuk worden besproken, resulteerden in een kalibratie module geschreven in Matlab die vandaag standaard wordt gebruikt in het laboratorium voor Analytische Scheikunde voor toepassingen van Raman spectroscopie in de (micro)biologie (o.a. de studie van biochemische reacties in de cel en fungi). Om voor spectrale verschuivingen te corrigeren, wordt een methode beschreven waardoor de Raman spectra worden gestandaardiseerd voorafgaand aan de multivariate analyse. Hierdoor wordt de lange termijn stabiliteit van de spectrale databank gegarandeerd.

Vooraleer deze databank kan worden uitgebreid met spectrale fingerprints van micro-organismen, is een duidelijk inzicht nodig in de reproduceerbaarheid en de taxonomische resolutie van deze techniek. Zoals reeds werd vermeld, bekomt men via Raman spectroscopie een *volledige* moleculaire 'vingerafdruk'. Bijgevolg kan verwacht worden dat de reproduceerbaarheid van deze 'vingerafdrukken' en hun discriminerend vermogen bij microorganismen beïnvloed worden door de groeiomstandigheden. In hoofdstuk III wordt het effect van deze groeiomstandigheden onderzocht voor drie, fenotypisch verschillende species *B. cereus*, *B. licheniformis* en *B. pumilus*. Een brede waaier aan groeiomstandigheden werd getest, en leidde tot diverse vaststellingen:

1. Aangezien er (weinig) biomassa nodig is, is Raman spectroscopie in het geval van bacteriën beperkt tot stammen die kunnen gekweekt worden. Bijgevolg moeten de groeiomstandigheden zo worden gekozen, dat deze voor een zo breed mogelijke waaier van microorganismen toepasbaar zijn. Doordat Raman spectra beïnvloed worden door de groeiomstandigheden, moeten deze laatste zorgvuldig gekozen en gestandaardiseerd worden om de beste taxonomische resolutie te bereiken.

2. Extra-cellulair slijm (vb. Extracellulaire polysachariden, capsules) verstoort het opnemen van Raman spectra. Dit fenomeen kan gelukkig worden vermeden door een juiste selectie van de groeiparameters.

3. Bepaalde groeiomstandigheden zorgen ervoor dat sommige stammen moeilijk (of niet) kunnen geoogst worden. Dit bemoeilijkt het opstellen van een gestandaardiseerd protocol.

Daarenboven tonen de resultaten aan dat de identificatie op species niveau (d.m.v. lineaire discriminant analyse – LDA) in relatief beperkte mate afhankelijk is van de groeiomstandigheden, terwijl de groepering van stammen (*enkel* gebaseerd op clusteranalyse) sterk beïnvloed wordt door de groeiparameters (zoals het geselecteerde medium, de incubatietijd en –temperatuur). Dit ondersteunt het belang van de keuze van de groeiparameters voor het bekomen van de hoogste taxonomische resolutie. De analysesnelheid en het gebruiksgemak zijn echter bepalende factoren in de keuze van de incubatietijd en een gepast medium maakt het oogsten van bacteriële kolonies algemeen mogelijk. Na een vergelijking van diverse groeiomstandigheden, werd Brain Heart Infusion (BHI) geselecteerd als het standaard kweekmedium in combinatie met een incubatie van 7 uur bij 37°C.

Na de standaardisatie van het kalibratieprotocol en de groeiomstandigheden, rapporteert hoofdstuk IV over de validatie van de ontwikkelde methodologie d.m.v. een eerste modelsysteem. Dit modelsysteem bestaat uit acht nauw verwante *Bacillus* species die vertegenwoordigd zijn door 69 stammen en die behoren tot de homogene '*Bacillus subtilis*'- groep of hier nauw aan verwant zijn (*B. pumilus, B. licheniformis* en *B. sonorensis)*. Gebaseerd op kleine, maar reproduceerbare, spectrale verschillen was het mogelijk de verschillende species te identificeren a.d.h.v. een gesuperviseerde, LDA-gebaseerde, benadering. Ondanks de sterke verwantschap van de bestudeerde species werden beloftevolle identificatie resultaten behaald, zowel door interne (> 90%) als door externe (blinde) validatie (> 78%). Er werd besloten dat Raman spectroscopie een algemeen toepasbare identificatie techniek is doordat stammen, die voorheen niet waren ingesloten in de bibliotheek, toch correct konden worden geïdentificeerd. Moeilijkheden en tegenstrijdigheden kwamen enkel voor binnen taxa die heel waarschijnlijk niet duidelijk afgelijnd zijn.

Tot dusver werden stammen succesvol geïdentificeerd op het species niveau, gebaseerd op een set van stammen die voorheen werden toegeschreven aan een beperkte groep van nauw verwante, goed gekarakteriseerde species. De volgende stap betreft een situatie waarin geen betrouwbare *a priori* kennis over de identiteit van stammen beschikbaar is.

Daarom wordt in hoofdstuk V het potentieel van Raman spectroscopie binnen de taxonomie verkend a.d.h.v. de fylogenetisch homogene '*Bacillus cereus*'-groep. Het is niet de bedoeling om de complexe taxonomische situatie van deze groep te verduidelijken, maar wel om het taxonomisch potentieel van Raman spectroscopie te onderzoeken.

In de praktijk werd onderzocht of de clusters die konden afgelijnd worden d.m.v. Raman spectroscopie overeenkwamen met taxonomische relevante groepen. Op basis van de similariteitsmatrix, werd vastgesteld dat de gevonden groepen, verkregen door Raman spectroscopie, een aanzienlijke overeenkomst hebben met deze gebaseerd op de FAFLP-similariteiten. Bovendien werden isolaten, die als identiek werden beschouwd op basis van rep-PCR analyse, in dezelfde Raman cluster teruggevonden. Omdat de topologie van een dendrogram sterk wordt beïnvloed door het gebruikte clusteralgoritme (vb. UPGMA, Ward) werd een bijkomende consensus studie uitgevoerd om de correlatie tussen de originele similariteitsmatrices te berekenen. Een sterk significante correlatie werd vastgesteld tussen Raman spectroscopie en FAFLP. De gevonden correlatie was echter te klein om te concluderen dat Raman spectroscopie FAFLP volledig kan vervangen als een op zichzelf staande typeringstechniek.

Als besluit kan men stellen dat Raman spectroscopie een gemakkelijk te gebruiken vibrationele techniek is die een minimum aan monster voorbereiding vraagt. Bovendien laat deze techniek een snelle (< 8u, inclusief groei) en accurate identificatie van bacteriën toe binnen de context van een strikt gestandaardiseerd protocol.

Deze scriptie heeft zich geconcentreerd op een groep van fylogenetisch nauw verwante en fenotypisch sterk gelijkende species (e.g. '*B. subtilis*' –groep) om de identificatie mogelijkheden van Raman spectroscopie te evalueren. Binnen de meeste toepassingen (vb. voeding, grond en water isolaten; plant pathogenen; klinisch relevante organismen, etc.) komen zulke nauw verwante species echter zelden voor en er kan dus verwacht worden dat deze techniek gemakkelijk kan uitgebreid worden naar dergelijke toepassingen. Daarenboven, op voorwaarde van een verdoorgedreven optimalisering, zijn er noch additionele reagentia, noch hooggeschoolde experts vereist voor de bediening van de instrumentatie. Hoewel de instrumentatie niet is geautomatiseerd, kunnen er per dag 30-40 stalen worden geanalyseerd inclusief groei, enten en staalvoorbereiding. Het niet geautomatiseerd zijn van Raman spectroscopie is zonder twijfel één van de belangrijkste aspecten die de doorbraak ervan in microbiologische laboratoria verhindert en/of beperkt. Rekening houdend met de brede toepassingsmogelijkheden van Raman spectroscopie, is dit allicht slechts een kwestie van tijd en financiën.

Dankwoord

Ik weet niet of je het gemerkt hebt, maar ik heb me enorm geamuseerd de voorbije vier jaar. Weet je niet waarover ik het heb? Blader dan eens terug naar het begin van deze scriptie… pagina 2… staat daar niet in zeven grote rode letters:

"Content"

Je mag het gerust weten, ik ben "heel content"! *Content* omdat ik het geluk heb gehad om me vier jaar te kunnen uitleven op een uitermate boeiend en uitdagend onderzoeksonderwerp dat, op de koop toe, een leuk einddoel had. Bovendien blijkt het nog écht te werken ook! Wat had je nog meer gewild?

U vraagt zich ongetwijfeld af hoe het allemaal begon. Als licentiaatstudent was ik gebeten door het onderzoek. Vol enthousiasme was ik in de bibliotheek artikels aan het zoeken toen plots een heel sympathieke man, Luc Moens, me aansprak en vroeg "Ik zie je hier zo vaak de laatste tijd, zou jij niet beter hier blijven". Dat was het prille begin van mijn vier jaar op het INW. Luc, bedankt om in mij te geloven en me de voorbije vier jaar de kans te geven om deze scriptie tot een goed einde te brengen. We hebben elkaar niet heel veel gezien, maar die keren dat we elkaar zagen hadden steeds een enorm positief effect op wat zou volgen. Heel erg bedankt voor de leuke herinneringen die ik aan alles heb overgehouden: ons etentje in "The house of Elliot", je advies, je visie op bepaalde topics, etc. Bedankt! Terug naar de bib. Daar stond ik dan. Had ik eindelijk de knoop doorgehakt om 'te gaan doctoreren' op de analytische scheikunde, wist ik natuurlijk niet waar te beginnen. Dit is waar Peter in het plaatje past. Ik herinner me nog goed dat je naar me toe kwam met zeker een halve meter artikels (in de hoogte wel te verstaan, in de breedte zouden dit er maar drie zijn) over Raman spectroscopie en biologische moleculen. Dat heeft me volledig overhaald en vormde eveneens de officiële start van onze 4-jarige samenwerking. Bedankt voor alle hulp!

Naast Luc en Peter, wil ik in het bijzonder twee personen bedanken. Laat ik je voorstellen aan de eerste: Paul De Vos. Paul, het is gek, maar ik kan me niet herinneren hoe onze eerste ontmoeting er aan toe ging. In elk geval herinner ik me maar al te goed wat er nadien kwam. Je leerde me stap voor stap dat de microbiologie helemaal niet zo exact is als ik dacht. Species lopen in elkaar over, kunnen hernoemd worden, zijn zeer divers, fylogenetisch homogeen, fenotypisch heterogeen… Kortom: een soepje. Het was duidelijk dat dit iets volledig nieuw voor mij zou worden. En wat die Raman spectrometrie betreft, daar wou je wel eens van zien wat de mogelijkheden waren, want een *Escherichia* en een *Bacillus* onderscheiden, dat kon jij met een microscoop.

"Zal ik je eens wat stammetjes geven, om mee te experimenteren", zei je. Voor mij geen probleem want een 'bacterie' was (toen nog) een 'bacterie' en als ze een verschillende naam hadden dan zou ik die wel uit elkaar halen. Al gauw moest ik die gedachte echter grondig herzien. Die bacteriën waren fenotypisch nauwelijks te onderscheiden en ik moet toegeven dat ik herhaaldelijk heb gezegd dat, als dit ooit lukt, het wel zou zijn door de exquise selectie van stammen. Nu, los daarvan, moet ik erkennen dat ik steeds genoot van onze wetenschappelijke praatjes. Bovendien vond ik het fantastisch hoe je keer op keer de tijd nam om een praatje te maken, om artikels grondig door te nemen, om een kritische kijk te geven op het onderzoek, om me te motiveren en te steunen, om advies te geven op diverse vlakken. Paul, dit is zeker geen understatement, zonder jou was dit nooit gelukt! Heel erg bedankt hiervoor!

Een tweede 'super'man zonder wie dit nooit zou zijn gelukt is Kees. Onze eerste ontmoeting ging als volgt: "Haai, ik ben Kees Maquelin!" U raadt het al, Kees is een Nederlander ! En wat voor één! In die maand dat ik de kans kreeg om bij jou op het lab te komen heb ik echt het licht gezien. Je kan dit zelfs vrij letterlijk nemen. Want laserlicht, optica, uitlijnen en het begrip 'fotontellen' (jullie gebruikten eigenlijk een ander woord, maar ik wou het hier netjes houden) dat is zowat de essentie van Raman spectroscopie. Vooral het 'fotontellen'. Zonder dit verblijf was me dit nooit, maar dan ook nooit gelukt. Hiervoor wil ik ook Gerwin bedanken. Bedankt dat ik toen op je lab de kneepjes van het vak mocht komen leren en voor de kritische kijk op mijn onderzoek. En eigenlijk bedankt aan iedereen ginds in Rotterdam want die maand was echt onvergetelijk ! Maar ik wijk weer eens af. Terug naar Kees. Kees, graag wil ik je ook bedanken voor de vele tips die je me gaf om dit werk tot een goed einde te brengen. Subtiele hints. Jouw visie gaf me vaak een boost aan nieuwe ideeën. Naast de boeiende wetenschap en onze vage plannen om een Raman in elk microbiologisch laboratorium te plaatsen, zijn we ondertussen ook beste maatjes geworden. Bedankt man!

Naast Luc, Peter, Paul en Kees zijn er nog een hele hoop collega's te bedanken van het labo Analytische scheikunde en Microbiologie. Enkele verdienen echter wat extra aandacht. Neem nu Peter Dawyndt bijvoorbeeld. Hij wijdde me in, in de wereld van BioNumerics en was de perfecte interface tussen het onomlijnde van de microbiologie en het exacte, het afgelijnde van de chemie. De discussies die we samen gehad hebben, hebben zeer zeker het onderzoek gestuurd naar wat het nu is. Na elke discussie kreeg ik steeds opnieuw nieuwe ideeën door jou geïnspireerd! Bedankt om altijd – maar dan ook altijd - klaar te staan om me te helpen! Uit hetzelfde nest wil ik ook graag Jeroen extra bedanken. Jeroen, onze samenwerking is pas in de laatste maanden écht van start gegaan maar met wat voor een

enorme intensiteit! Niet alleen hielp je me met de selectie van de stammen en het schetsen van het microbiologische kader rond dit onderzoek, je bijdrage in de laatste maand was van cruciaal belang om dit werk tijdig af te krijgen. Zonder jou was dit écht niet gelukt. Bedankt! En uiteraard wil ik ook Liesbeth bedanken! Bedankt om me in te wijden in de praktische handelingen die eigen zijn voor een microbioloog en mij (en de thesisstudenten) de weg te wijzen in het labo, te enten, platen te maken, etc… Je nam ons een groot pak werk uit handen.

En uiteraard…, wanneer ik het heb over cruciale bijdragen kan ik moeilijk zwijgen over Joachim, Leen en Wim. Het was me een plezier om met jullie als thesis studenten te werken! Jullie zorgden voor de eersteklas data die aan de basis liggen van dit onderzoek. Maar meer dan dit, hebben onze gesprekken me enorm geïnspireerd om bepaalde instrumentele aanpassingen door te voeren of mathematische algoritmes aan te passen. Los daarvan – en eigenlijk het allerbelangrijkste – is dat jullie me een reeks fantastische herinneringen hebben gegeven.

U merkt het, dit werk heb ik zeker niet alleen gedaan en nu ik er zo over denk, zijn er massa's mensen die ertoe hebben bijgedragen dat dit tot een goed einde is gebracht. En één iemand speelt hier toch wel een belangrijke rol in, des te meer omdat ik soms nogal een sloddervos ben die moeilijk om kan met papierwerk. Chantal, bij deze heel erg bedankt om het papierwerk steeds veel beter bij te houden dan ik het zou kunnen en vooral voor de zeer welkome verstrooiing die onze babbeltjes altijd brachten. Ik heb het je al vaker gezegd, en bij deze blijft het niet meer tussen ons, maar jij bent de beste secretaresse van de hele wereld! In de categorie 'verstrooiing brengen' gaan uiteraard Joke, Kris, Niko en Marleen met de eer gaan lopen. Joke, Kris en Marleen, bedankt voor de vele onwetenschappelijke praatjes en uiteraard – hoe kan het ook anders – voor de wetenschappelijke praatjes. Jullie opmerkingen en vragen hielden me steeds scherp en hebben er toe bijgedragen dat alles loopt zoals het vandaag loopt. Kris, bedankt voor de ultieme test-kit te zijn voor de calibratie-software. Net als ik dacht dat alles goed ging wist jij weer een bug te vinden. En ja, mijn programmeertaal zal uiteraard nooit zo poëtisch zijn als de jouwe, maar samen hebben we het toch maar mooi voor elkaar gekregen! Dit laatste brengt me bij mijn bureaumaatje: Niko. Wat had ik een geluk. Je zorgde dat de spectrometer steeds puik stond gecalibreerd, dat de koffie gezet was en – ik weet nog steeds niet hoe je dat doet – om de dingen klaar te hebben voor ik ze je gevraagd had. Zonder twijfel het toppunt van pro-activiteit!

En als ik het heb over activititeit, zei het pro- of radio-, dan kan ik moeilijk om Frans heen. De laatste maanden sprong je elke week eens binnen om me een nieuw verhaal te vertellen, het ene al hilarischer dan het andere (ik kan hier nu moeilijk een voorbeeld geven, want had je belooft dat het tussen ons zou blijven). Verder ook enorm bedankt om me al die nieuwe Jazz-namen te leren kennen, dankzij jou is mijn CD-rek te klein. Maar ik zet het je betaald: als ik je tegenkom op één van de vele jazz-concerten in het verschiet, trakteer ik je op een trappist. Beloofd!

Antoine en Hermine, als steun achter de schermen waren jullie onontbeerlijk. De vele etentjes, barbecues en bijhorende, heerlijke wijnen waren een enorme deugd na het vele werk. Ik kijk al uit wat het deze zomer gaat geven, om maar te zwijgen van het najaar. Stephen, jij was mijn vaste donderdagmiddag afspraak waar ik telkens weer naar uitkeek. Zeer deugddoend om de hersenpan te verluchten door gewoon eens het verstand op nul te zetten of wat bij te praten over het komende feestje, het feestje van voorbije week of over sport.

En zoals het hoort in dankwoorden staan de vriendinnen achteraan. Verschiet niet, ik heb er drie. Bij de eerste, Stéphanie, sta ik in het krijt wegens het samen koffie gaan drinken op zonnige terrasjes en vooral om die heerlijke gebakjes die je steeds voor me maakte. De tweede is zonder twijfel de oudste en beste vriendin die ik ooit zal hebben. Bebonne, bedankt om ons elke woensdag te voorzien in veel plezier, vol-au-vent, fricandontjes en "*tomatches*". Het is plezant om een Bebonne als jij te hebben!
En uiteraard, de derde, de kers op de taart: mijn liefste schat Daniëlle. Jij was er steeds om naar mij te luisteren, me te steunen, tips te geven etc… Uiteraard, zo zijn vriendinnen nu éénmaal in dankwoorden. Maar jij bent echt ongelofelijk en al van de eerste dag dat ik je leerde kennen heb ik me enorm geamuseerd met je, elke dag en elke seconde. En nu dat ik aan de laatste zin van mijn dankwoord bezig ben, kijk ik al uit naar wat ons nog allemaal te wachten staat.

Bedankt !

Short explanation of the mathematical methods used for numerical analysis

Appendix

Canberra metric

This similarity measure is usually meant for non-negative variables only. The Canberra metric makes a summation of a series of ratios between corresponding planar values. This therefore accounts for the distance between two points but also their relation to the 'origin'. The similarity of two objects *x* and *y*, each described by *d* variables is defined as:

$$Canberra(x, y) = \sum_{i=1}^{d} \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

The output ranges from 0 to the number of variables (*d*) used.

Dice Coefficient

The dice coefficient is a correlation coefficient for discrete events and is often used as a similarity measure when banding patterns (e.g. gel electrophoresis, FAFLP) are involved. The correlation between two banding patterns *x* and *y* is defined as:

$$Dice(x, y) = \frac{2 \times N_{common}}{N_{common} + N_{total}}$$

Where N_{common} is the number of bands that both patterns have in common and N_{total} is the total number of band positions (see Figure 1 for details). Figure 1 shows 2 patterns (1 and 2), each containing six bands. Five bands are at the same position, while each pattern contains one discriminating band at a different position.

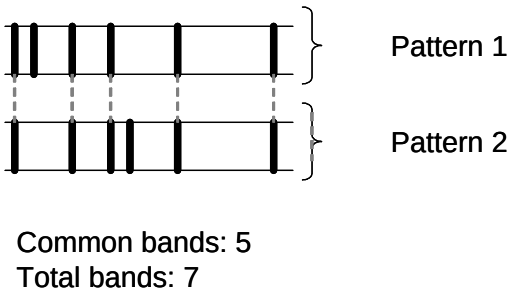


Figure 1: illustration of two different banding patterns. The dice coefficient in this example equals 0.83.

Euclidean distance:

The Euclidian distance can be considered to be the shortest distance between two points, and is basically the same as Pythagorus' equation when considered in 2 dimensions. The Euclidean distance between objects *x* and *y* is defined as:

$$Euclidean(x, y) = \sqrt{\sum_{j=1}^{m} (x_j - y_j)^2}$$

Where *m* is the number of variables. Figure 2 illustrates the Euclidean distance for two objects *x* and *y*, each described by two variables (X_1 and X_2)

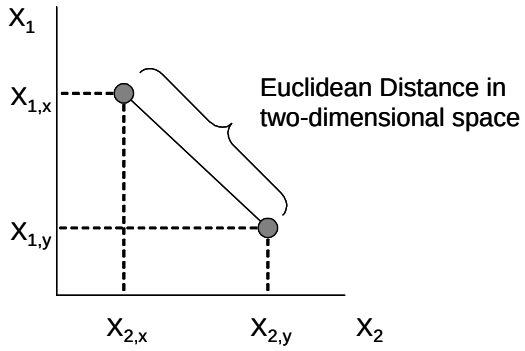


Figure 2: illustration of the Euclidean distance in the case where objects are described by two variables.

Hierarchical Cluster analysis (HCA):

Hierarchical cluster analysis is comprised of agglomerative methods and divisive methods that find clusters of observations within a data set. Cluster analysis is used to classify objects (e.g. Raman spectra of strains), characterized by a set of variables (e.g. wavenumbers, LDA-scores,…), into groups based on their similarity. Cluster analysis constitutes two main steps: (1) calculation of a similarity matrix based on a similarity measure (e.g. Euclidean distance, Dice coefficient, Canberra metric) and (2) the formation of groups based on a linkage algorithm (e.g. UPGMA, Ward). The obtained groupings are visualized by a dendrogram.

Linear discriminant Analysis (LDA)

Linear discriminant analysis is a supervised method to discriminate between two or more groups of samples. The method is based on the minimization of variance within the same group (e.g. replicate measurements of one particular strain) and the maximization of variance between different groups (e.g. replicate measurements of different strains), hence aiming at an optimal discrimination. Since groups need to be defined prior to the definition of the groups, LDA needs *a priori* knowledge on the identity of the groups (e.g. to which species do these strains belong?). Figure 3 illustrates how two classes A and B, characterized by the variables X_1 and X_2, can be discriminated using the discrimination function Z.

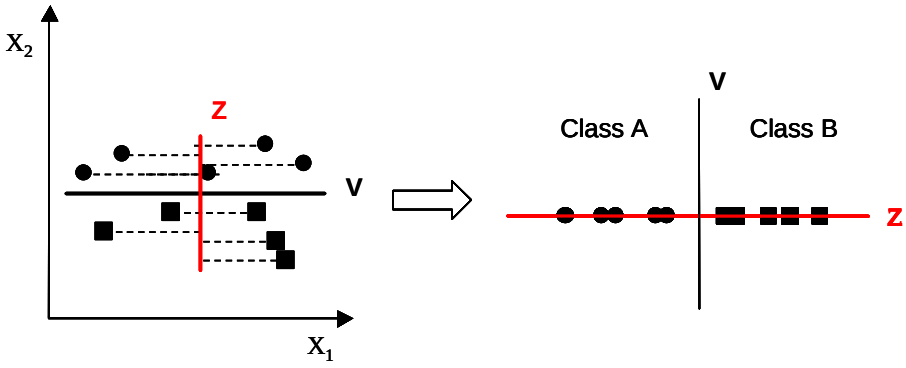


Figure 3: The principle of LDA. Class A and B can be discriminated by the discriminant function Z. Note that Z does not represent the largest variation in the data set. The largest variation is represented by V.

In this example, the discrimination function Z is of the form:

$$D = a_0 + a_1 X_1 + a_2 X_2$$

Where D is the discriminant score and a_0-a_2 are constants. Since unknown objects (strains) are always assigned to one of the classes enclosed during training, it happens that objects, belonging to other classes than enclosed during training, are erroneously identified. The latter is a typical shortcoming of LDA.

Principal component analysis (PCA)

Principal component analysis is a well known method for the reduction of the dimensionality of the data, in other words to reduce the number of variables. Suppose a data matrix contains 1000 items, each described by 701 variables. In this case a new set of orthogonal variables (Principal components – PC's) can be calculated by means of PCA.

Since the first new variables (e.g. the first 50) explain the majority (e.g. > 99%) of the variance present in the dataset, the dimensionality of the original data matrix can be reduced from 1000 to 50, wile maintaining 99% of the information.

The scores of each PC can be represented by a column vector, the loadings by a row vector. When applied on a data matrix X (*i* rows and *j* columns) PCA results in *n* PC's (Figure 4). The scores represent the contribution of each PC to the indiviual spectra, the loadings the contribution of each variable to the multiple PC's.

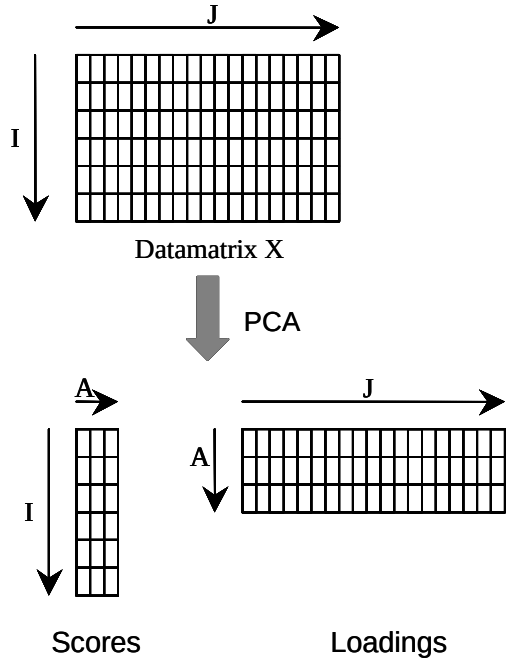


Figure 4: The original data matrix X is reduced in dimensionality by means of PCA. As a result a new set of variables (PC's) is obtained, containing the different scores and a set of loadings describing the relation between the original variables and the PC's.

However, the calculated PC's does not all contribute in a significant way to the total variance present in the data set. The significance of the a^{th} PC is represented by its eigenvalue (EV_a). The eigenvalue is calculated as the sum of squares of the scores (s_{ia}) for that particular PC:

$$EV_a = \sum_{i=1}^{I} s_{ia}^2$$

The eigenvalue (informational content) decreases for each successive PC. The cumulative eigenvalue represents that part of the variation of the data set that is explained by the first few retained PC's and is calculated as:

$$C\%_{EV} = \sum_{a=1}^{A} \%EV_a$$

By selecting a cut-off value of e.g. 95% for the cumulative eigenvalue, a large number of PC's can be rejected, with a loss of information of only 5%. This five percent is often explained by 'noise'. The score matrix of the retained PC's is the final solutionof PCA-datareductie (Fig. 5).

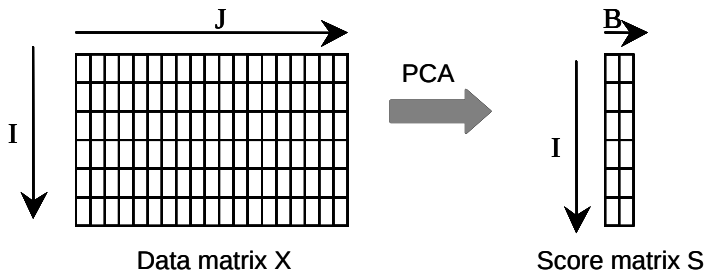


Figure 5: Representation of the original Data matrix X (1000 x 701) before PCA (left) and the resulting score datamatrix (1000 x 50) after datareduction with PCA (right)

Where *B* is the number of PC's explaining 95% of the total variance in the data set, *I* the number of objects (e.g.1000 Raman spectra) in the data set and *J* the number of original variables present in a Raman spectrum (e.g. 701)

UPGMA

Unweighted Pair Group Method with Arithmetic averages (UPGMA) also known as the average linkage algorithm. In UPGMA the distances of a new object (*u*) to all the objects of a particular cluster are averaged (Figure 6). The object is finally grouped with that cluster resulting in the lowest average distance (or highest similarity).

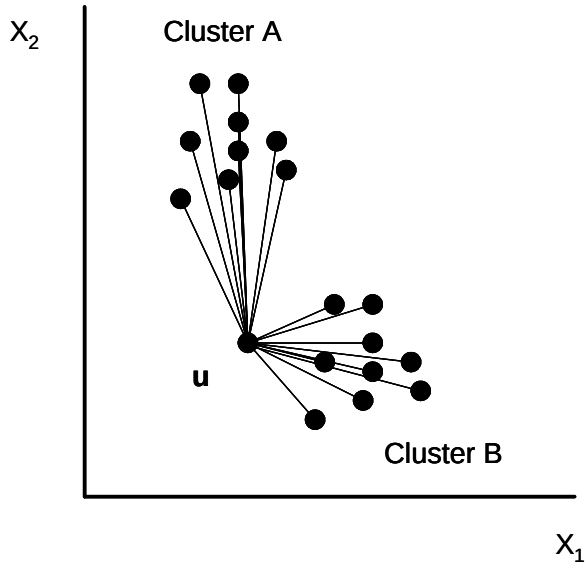


Figure 6: the principle of UPGMA. The distance between a new object *u* and all the objects of each cluster (A and B) is calculated and averaged. The object will join that cluster resulting in the lowest average distance, in this case cluster B

Ward's Algorithm

Ward's linkage is distinct from all the other linkage methods because it uses an analysis of variance approach to evaluate the distances between clusters. It is based on a heterogeneity criterion which is defined as the sum of the squared distances of each member of the cluster to the centroïd of the cluster. Elements or clusters are joined with as criterion that the sum of heterogeneities of all clusters should increase as little as possible.