Assessment of complex microbial assemblages: description of their diversity and characterisation of individual members

Von der Fakultät für Chemie und Physik

der TU Bergakademie Freiberg

angenommene

Habilitationsschrift

zur Erlangung des akademischen Grades

eines habilitierten Doktors der Naturwissenschaften

Dr. rer. nat. habil.

von Dr. Martin Mühling

geboren am 09. Dezember 1969 in Pleinfeld (Bayern)

eingereicht am 01. Juli 2016

Gutachter: Prof. Dr. Michael Schlömann (TU Bergakademie Freiberg)
Prof. Dr. Klaus Jürgens (Leibniz-Institut für Ostseeforschung, Warnemünde)
Prof. Dr. Hans-Peter Grossart (Leibniz-Institut für Gewässerökologie und Binnenfischerei, Stechlin)

Tag der Verleihung: 14. Februar 2017

ACKNOWLEDGEMENTS

I would like to thank Professor Michael Schlömann for giving me the opportunity to undertake research that contributed to this thesis. I am also grateful to him for examining the results described in this thesis including those obtained from research that I have undertaken outside Freiberg. In this context, I am also indebted to my former line manager Dr Ian Joint (Plymouth Marine Laboratory) and to Prof. Nicholas Mann (Warwick University). I also appreciate the time and efforts of the two additional examiners that evaluate this thesis.

It is obvious, that the research presented in this thesis was not the product of exclusively my own hands' work. Supporting colleagues are co-authors on the various publications or are mentioned in appropriate ways in the acknowledgements. It should be noted in this context that the sometimes long list of co-authors does not imply 'inability' to carry out experimental work or to write manuscripts myself! In contrast, these lists of co-authors reflect a culture of honest acknowledgements of sometimes even small contributions of colleagues. *Vice versa*, regarding those publications that list me as co-author I either undertook experimental work or contributed to the design of the project (including the grant application and the planning of the experiments), the writing of parts of the manuscript or data analyses.

Furthermore, the reader of this thesis should not only notice the list of PhD students in the appendix, but also the names of research technicians and graduate and undergraduate students that have been supervised by me and who contributed to my research. Of those I would like to name two PhD students and a MSc student who contributed in particular to the research that I was engaged in: Dr Ellie Jameson, Dr Sophie Ullrich and MSc André Pollender. Additionally, the genomic and metagenomic sequence data analysed over the last years at Freiberg have been produced, assembled and annotated at the Goettingen Genomics Lab (G2L; Goettingen University). Most of the work has been undertaken by Dr Anja Poehlein. I am particularly grateful to Anja for her contributions.

I am also indebted to my colleagues at the TU Bergakademie Freiberg. I would like to give special thanks to Mrs Beate Erler, who actively supported me in two ways: firstly by supervising much of the laboratory based work of my PhD students, graduates and undergraduates and, secondly, by providing unparalleled technical and organisatorial help in the teaching of lab practicals. I am also indebted to Mr Janosch Gröning and several other colleagues for their help in the laboratory. Particular thanks also go to Monique Leibelt for her professional support regarding administrative tasks and to Juliane Heim for corrections to this thesis.

to Clara, Moritz and Kristin

1. Microbial ecology

According to Caumette *et al.* (2015) the term ecology is derived from the Greek words "oikos" (the house and its operation) and "logos" (the word, knowledge or discourse) and can, therefore, be defined as the scientific field engaged in the "knowledge of the laws governing the house". This, in extension, results in the simple conclusion that microbial ecology represents the study of the relationship between microorganisms, their co-occurring biota and the prevailing environmental conditions (Caumette *et al.* 2015).

The term microbial ecology has been in use since the early 1960s (Caumette *et al.* 2015) and microbial ecologists have made astonishing discoveries since. Microbial life at extremes such as in the hydrothermal vents (see Dubilier *et al.* 2008 and references therein) or the abundance of picophytoplankton (Waterbury *et al.* 1979; Chisholm *et al.* 1988) in the deep and surface waters of the oceans, respectively, are only a few of many highlights. Nevertheless, a microbial ecologist who, after leaving the field early in their career, now intends to return would hardly recognise again their former scientific field. The main reason for this hypothesis is to be found in the advances made to the methodologies employed in the field. Most of these were developed for biomedical research and were subsequently hijacked, sometimes followed by minor modifications, by microbial ecologists.

The Author presents in this thesis scientific findings which, although spanning only a fraction of the era of research into microbial ecology, have been obtained using various modern tools of the trade. These studies were undertaken by the Author during his employment as postdoctoral scientist at Warwick University (UK), as member of staff at Plymouth Marine Laboratory (UK) and as scientist at the TU Bergakademie Freiberg. Although the scientific issues and the environmental habitats investigated by the Author changed due to funding constraints or due to change of work place (i.e. from the marine to the mining environment) the research shared, by and large, a common aim: to further the existing understanding of microbial communities. The methodological approach chosen to achieve this aim employed both isolation followed by the characterisation of microorganisms and culture independent techniques. Both of these strategies utilised again a variety of methods, but techniques in molecular biology represent a common theme. In particular, the polymerase chain reaction (PCR) formed the work horse for much of the research since it has been routinely used for the amplification of a marker gene for strain identification or analysis of the microbial diversity.

To achieve this, the amplicons were either directly sequenced by the Sanger approach or analysed *via* the application of genetic fingerprint techniques or through Sanger sequencing of individual amplicons cloned into a heterologous host. However, the Author did not remain at idle while with these 'classical' approaches for the analysis of microbial communities, but utilised the advances made in the development of nucleotide sequence analysis. In particular, the highly parallelised sequencing techniques (e.g. 454 pyrosequencing, Illumina sequencing) offered the chance to obtain both high genetic resolution of the microbial diversity present in a sample and identification of many individuals through sequence comparison with appropriate sequence repositories. Moreover, these next generation sequencing (NGS) techniques also provided a cost-effective opportunity to extent the characterisation of microbial strains to non-clonal cultures and to even complex microbial assemblages (metagenomics).

The work involving the high throughput sequencing techniques has been undertaken in collaboration with Dr Jack Gilbert (PML, lateron at Argonne National Laboratory, USA) and, since at Freiberg, with Dr Anja Poehlein (Goettingen University). These colleagues are thanked for their support with sequence data handling and analyses.

2. The extended 'classical' approach to the study of micro-organisms: isolation followed by phenotypic and genomic characterisation

The number of species of prokaryotic microorganisms (bacteria and archaea) formally described in the *International Journal of Systematic Bacteriology* or, later on, in the *International Journal of Systematic and Evolutionary Microbiology* amounts to 10,360 (01 August 2013; http://www.bacterio.net/-number.html) and is, therefore, very similar to that of recognised bird species (10,596; IOC World Bird Names (v 3.3). F. Gill, D. Donsker, 2013, status of 16 April 2013). However, it is widely accepted that this does not reflect reality. The discrepancy between the numbers of formally described and estimated species is largely explained by the lack of representative isolates for most of those species: a pure culture is a prerequisite for formal species descriptions. In this context, a pure culture is defined as being

clonal (i.e. derived from a single cell or, in the cases of filamentous microorganisms, derived from one filament) and axenic (i.e. presence of only one genetic strain or species).

To appreciate this view the reader is reminded of the "great plate count anomaly" (Staley and Konopka 1985) which illustrates this important issue in environmental microbiology; that is, the limited success of the isolation of microorganisms from their natural environment. In brief, it summarises the fact that the cell number in a sample determined through counting a e.g. DAPI-stained aliquot by epifluorescence microscopy is orders of magnitude larger than the number of colony forming units (cfu) obtained on conventional selective media using, for instance, standard plating techniques (e.g. Fig. 1). In general, it is concluded that only a small proportion, thought to be within the range of 0.001–1% of the total microbial assemblage, is brought into culture by such an approach (Kogure *et al.* 1979; Staley and Konopka 1985; Amann *et al.* 1995). Moreover, those microbes that are among the small fraction of so far isolated strains do (often) not reflect a representative subset of the various taxonomic and, therefore, presumably also functional groups present in the natural microbial assemblage (Joint *et al.* 2010).

Fig. 1: The Author as 'traditional' microbiologist using plating techniques.

Methods used for isolation and cultivation have been reviewed by many, including the Author (Joint *et al.* 2010). Issues that have been discussed herein address parameters of a presumed complex network of interactions between various microbial species. These interactions are thought as being potentially relevant for cellular growth, but are likely interrupted during the standard isolation procedures widely used to obtain pure cultures for laboratory studies.

However, pure cultures are a prerequisite for physiological and biochemical tests to elucidate the life style of a microorganism, for taxonomic classification and detailed analyses of its

cellular responses to environmental changes and, in many instances, the biotechnological exploitation of relevant enzymes and compounds. Here, contamination with a different microbial strain will influence the observed response since biochemical interactions between the strains resemble an additional parameter. Moreover, variation in the ratio of the various strains within the mixed culture does not permit precisely controlled experiments and, thus, results in not reproducible datasets. Although genome analysis of microbial strains in mixed cultures is possible (see section 2.4.2), mixed cultures are still a problem in investigations that utilise modern techniques in gene expression (i.e. transcriptomics) which are now widely used for the characterisation of prokaryotes (hence, extended 'classical' approach). Therefore, novel approaches for the isolation of microorganisms are required to improve the likelihood of bringing novel groups of microbes successfully into culture.

2.1 Novel methods for the isolation of microorganisms

The marine realm

Efforts for the isolation of microorganisms for biodiscovery have long focused on heterotrophic bacteria from the terrestrial environment, primarily due to the relative ease with which these organisms can be grown and manipulated. Among these, a particular emphasis was on members of the phylum *Actinobacteria* due to their well-known production of high levels of secondary metabolites, in particular antibiotics, and the (additional) enzymes that are required to synthesise those compounds. However, over the last decade several lines of evidence have emerged that marine bacteria harbour a wide range of enzymes with novel substrate specificities and even novel enzymatic activities (Wagner-Döbler *et al.* 2002), though no large-scale systematic analysis to test this hypothesis had been undertaken.
This together with the need for and rewards of new and improved enzymes for biocatalysis in, for example, organic synthesis has led to the resurgence of the interest in the isolation and culture of marine bacteria as a valuable strategy to access the potential wealth of marine microorganisms for biotechnology. As part of the EU-funded consortium project MIRACLE (Microbial marine communities diversity: from culture to function) a culture collection has been assembled containing approx. 900 microbial strains using a diversity of experimental procedures for the isolation. Due to potential conflicts with future applications for protection of intellectual property (IP) related to those isolates and their activities, only a very limited

fraction of the information gathered from this programme of research has so far been published (Joint *et al.* 2010; Allen *et al.* 2012; Willetts *et al.* 2012; Mühling *et al.* 2013), and none of the novel isolates underwent official species description. An overview of the strains obtained can be found in Joint *et al.* (2010) and information on functional characteristics of a subset of approx. 400 of those by Mühling *et al.* (2013). The screening effort for enzymatic functions relevant for biotechnological applications also revealed a novel type of Bayer-Villiger monooxygenase in a strain (PML 168) belonging to the species *Stenotrophomonas maltophilia* (Willetts *et al.* 2012). The Bayer-Villiger monooxygenase activity was further cooperated by the analysis of the genome sequence of the same strain (Allen *et al.* 2012; see also section 2.4.1).

Many of these isolates have been obtained by various techniques of isolation (Joint *et al.* 2010) including a novel *state-of-the-art* approach which has great potential for the isolation of novel microbes, though – as is shown below – also seems to result in a bias towards particular microbial strains. This technique, which had been developed by Zengler and colleagues (2002), attempts to overcome three of the possibly most relevant limitations to cultivation (see Joint *et al.* 2010): procedures used for the isolation are often *i)* not only unsuitable for maintaining molecular interactions between individuals, but *ii)* also permit the fast growing strains to outcompete the slow growing strains which, in turn, results in an imbalance of cell-to-cell communications, and *iii)* often provide neither the correct type nor the appropriate concentrations of substrate and nutrients required by the microbes for growth. In this context, it must be considered that marine microorganisms have evolved to cope with the oligotrophic conditions prevailing in many natural environments. Therefore, the (relatively) high concentrations of substrate and nutrients used in most isolation processes may be toxic to those microorganisms.

With the support of a grant from the Natural Environment Research Council (NERC) of the UK the Author was able to establish at PML the method of Zengler *et al.* (2002). Briefly, the method employs the CellSys 100™ Microdrop Maker technology developed by One Cell Systems, Inc. (USA). This uses microdroplet encapsulation in an agarose matrix to isolate and analyse individual cells. Gel microdrops are prepared by mixing molten agarose with seawater containing natural microbial assemblages and dispersing the mix into an excess of a non-aqueous phase (oil) to form an emulsion. The emulsion is cooled quickly causing the molten agarose drops (some containing cells) to gel. By altering the speed of stirring the emulsion it

is possible to change the size of the microdrops. The Author has developed a protocol that results in microdrops of 40 to 60 µm diameter. Encapsulation of only one cell per microdrop (Fig. 2) is achieved by the use of a 1:10 ratio of microbial cells to drops formed (e.g. by adding approx. 1 million cells to 500 µL of low melting agarose which, applying the particular protocol, results in ca. 10 million microdrops). Following an appropriate incubation period in a flow through system (Fig. 3) the gelled agarose microdrops are physically distinct and can be manipulated by electrostatic sorting of individual drops using a flow cytometer (Fig. 4).

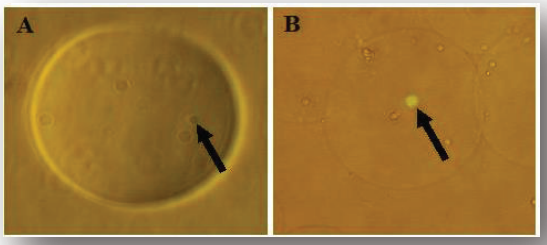


Fig. 2: Agarose microdrops containing a single encapsulated cell (A) and a micro-colony (B) after eight weeks of incubation in seawater (flow-through system). Cells are stained with *SYBR Green I* DNA stain and visualised by fluorescence microscopy.

Using this approach it is possible to routinely achieve five to ten percent of the microdrops containing a single cell (Fig. 2A). Thus, the microdrop formation method allows massively parallel encapsulation of microbial cells: estimates suggest that up to one million cells are individually encapsulated per experiment.

Fig. 3: Flow through culturing system setup permitting continuous supply of fresh medium, here the supernatant from a natural complex phytoplankton assemblage. The supernatant is supplied at a low flow rate *via* a peristaltic pump to the suspension of millions of agarose droplets within the bioreactors (XK16/20 chromatography column, Pharmacia).

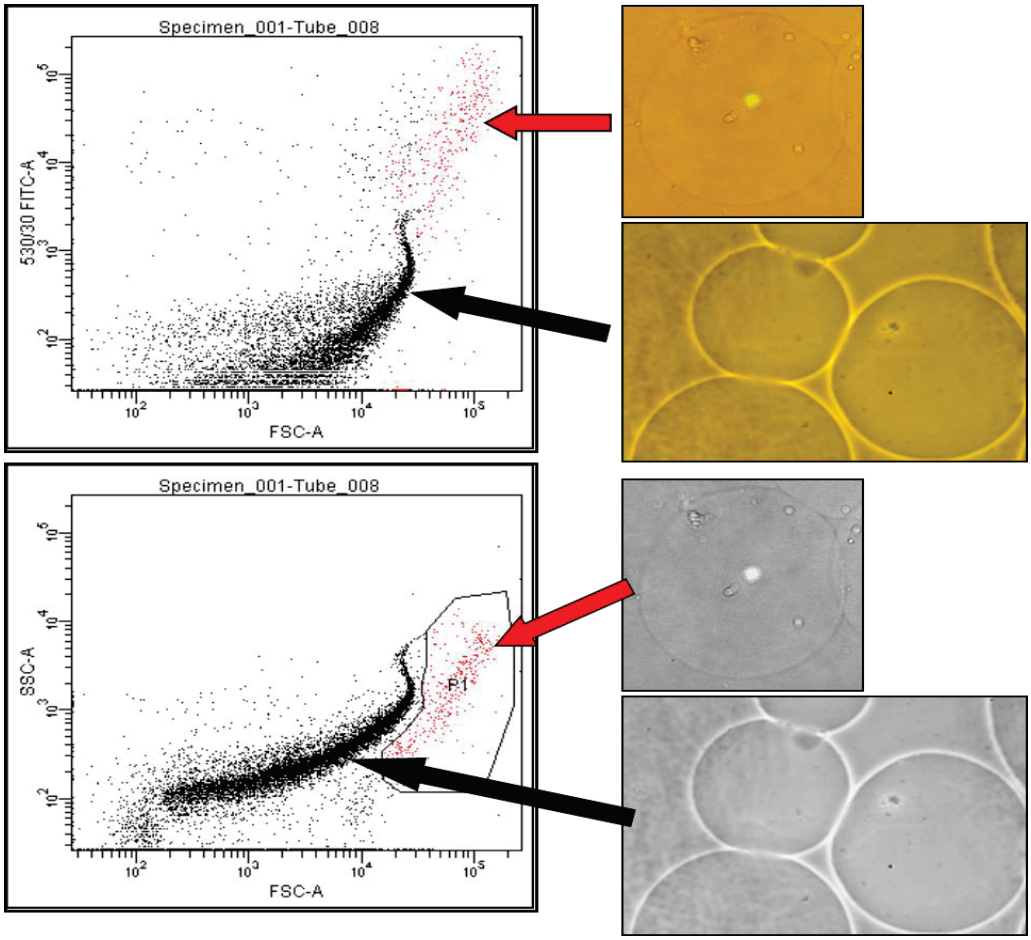


Fig. 4: Flow cytometric sorting of microdrops containing micro-cultures derived from (presumably) individual cells. The presence of those micro-colonies within the droplets results in (A) altered fluorescence upon staining with a fluorescent DNA stain (here *SYBR Green I*). However, these droplets can also be distinguished from those without cultures by their altered scattering properties (B), thus allowing to avoid DNA stains which may impact cell viability. The cells within these droplets can then be liberated by directly sorting the droplets on solidified media (e.g. agarose plates).

Encapsulation into agarose provides a mechanism to manipulate individual bacterial cells and fulfils two further functions. Firstly, physical separation of an individual bacterial cell from other cells provides a matrix for the development of all of the components of a mixed assemblage, yet quarantines the developing clones by keeping individual species physically apart. Agarose is porous and nutrients, metabolites and signal molecules can diffuse into the microdrop containing a growing colony (Fig. 2) derived from an individual cell (i.e. a clonal culture). Similarly, secondary metabolites and QS molecules can diffuse out, being available

to nearby bacteria, also quarantined within their own microdrops. The physical constraint of the gel matrix does not prevent bacteria from replicating and growing, and any faster growing bacteria do not take over the culture since they are constrained within the agarose matrix. Secondly, nutrients are supplied at ambient concentrations so avoiding perturbation by high nutrient concentrations. It has been suggested that exposure to much higher nutrient concentrations may be one reason why so few bacteria from natural environments have been successfully brought into laboratory culture. By containing the agarose microdrops in a flow-through system, bacteria are exposed to a continuous supply of nutrients at the ambient nutrient concentration.

The method has been used to isolate not only the most abundant, but also scarce bacteria within an assemblage. This success is entirely due to the physical separation and the containment within the gel microdroplets which permits the simultaneous and relatively non-competitive growth of both slow- and fast-growing microorganisms in media, thereby preventing over-growth by fast-growing microorganisms – the "microbial weeds" (Eilers *et al.* 2000). It is increasingly common to find examples of co-culture leading to the growth of novel bacteria. Kaeberlein *et al.* (2002) showed that clonal cultures of novel bacteria were obtained only in a minority of cases. Most isolates required co-culture with other bacteria. Encapsulation allows the incubation of a diversity of bacteria within the same medium while still being maintained in a clonal state within the gel microdroplet. Apart from its relevance for the isolation of microbial strains recalcitrant to laboratory culture, this approach also provides the potential to test hypotheses relating to metabolic consortia. Clearly, the contribution of species richness and composition to bacterial activities (Bell *et al.* 2005) plays a role in, for example, the degradation of xenobiotics (see e.g. Slater and Bull 1982): bacteria with different metabolic capabilities are capable of concerted actions. By containing the agarose microdrops in a flow-through system, bacteria can be exposed to a continuous supply of substrate (e.g. xenobiotics or specific breakdown products of the xenobiotics) at the ambient or any experimental nutrient concentration. Such lines of research are, however, not part of this thesis.

The potential of this approach for the isolation of difficult to culture microbes shall be briefly demonstrated on the basis of one isolate that was subsequently lost and publishing details on that is therefore not affected by any sanction due to potential IP issues. Isolate 705e was obtained by incubating microdrops with encapsulated microbes for 35 days in a growth

column with continous supply of aged sea water (0.45-µm filtered sea water, stored for several weeks in the dark with no additional nutrients added). Microdrops were then isolated using flow cytometry and spotted directly on solidified (agar) nutrient plates (Marine Broth with aged seawater: Joint *et al.* 2010). The resulting colony was identified by directly sequencing a PCR amplicon of the 16S rRNA gene fragment obtained with primers 9bfm and 1513uR (Mühling *et al.* 2008). Based on colony morphology and the quality of the obtained sequence (Table 1) the isolate was judged to be clonal. Sequence comparison of its 16S rRNA gene revealed that it belonged to a then unknown lineage of *Planctomycetes* sharing only 82% sequence similarity (over 84% of the 1404-bp sequence submitted to NCBI for the BLAST search) to *Rhodopirellula* sp. strain SM48 (Winkelmann and Harder 2009; accession no. FJ624354.1).

Table 1: Sequence of the 1404-bp fragment of the 16S rRNA gene of isolate 705e (FASTA format). The fragment was amplified by PCR using primer pair 9bfm/1513uR (Mühling *et al.* 2008). The amplicon was directly submitted to Sanger sequencing. The sequence has not been submitted to any sequence repository.

```
>Isolate_705_16S_rRNA_gene_fragment
ATTGAACGCTGGCGGCATGGCTAAAACATGCAAGTCGAACGATCCCTTCGGGGAGAGTGGCGAAAGGGCGAGGAA
TAGAATCGAATGTACCCCAAGGTGGGGGATAGCTTCTGGAAACGGAAGGTAATACCCCATGTGCTCTACGGAGGA
AAGGTTTACCGCCTTGGGAGCAGCGATTCTCCTATCAGGTAGTTGGTGTGGTAACGGCTCACCAAGCCGAAGACG
GGTAGCGGGTGTGAGAGCATGACCCGCCGCATCGGGACTGAGACACTGCCCGGACTCCTACGGGAGGCTGCAGTA
ACGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCAATGCCGCGTGTGGGATCAAGCATCTTCGATGTGTAAAC
CACTGTCAGGGTCTAGGAATACTGACCAGACCCAGAGGAAGGGCCGGCTAATTCAGTGCCAGCAGCCGCGGTAAT
ACTGAAGGCCCGAGCGTTAATCGGAATCACTGGGCTTAAAGCGTACGCAGGCGGACTTGTAGGTATTTTGTGAAA
TCCCACGGCTTAACCGTGGAACTGCAGAGTAAACCACAAGTCTTGAGACATGTAGGGGTCAGTGGAACGATGTGT
GGAGCGGTGAAATGCGTAGATATACATCGGAACGCCAATGGCGAAgGCAGCTGACTGGGCATGTTCTGACGCTCA
GGTACGAAAGCGTGGGTAGCGAACGGGATTAGATACCCCGGTAGTCCACGCCGTAAACGATGCACACTAGGTCGC
GGGAACTCTCATGTTCTCGCGGCCGAAGGTAAACTGGTAAGTGTGCCGCCTGGGGAGTACGGTCGCAAGGCTAAA
ACTCAAAGGAATTGACAGGGGCTCACACAAGCGGTGGAGCATGTTGCTTAATTCGAGGCAACGCGAAGAACCTTA
CCTGGGCTTGACATGTATGGATTAGCTCTTGGAAACAAGAGTGACGCCCTTGGGTGGAACATACACAGGTGCTGC
ATGGCTGTCGTCAGCTCGTGCTGTGAAGTGTCGGGTTAAGTCCCTTAACGAGCGCAACCCCTATCGTTAGTTACT
AACGCGTTATGGCGAGGACTCTAGCGAGACTGCCGGTGTCAAACCGGAGGAAGGTGGGGATGACGTCAAGTCCTC
ATGGCCTTTATGCCCAGGGATGCAAACGTGCTACAATGGTACGGACAAAGCGATGCGATACCGCGAGGTGGAGCA
AATCGCAAAAACCGTGCCCCAGTTCGGACAGCAGGCTGAAATTCGCCTGCTTGAAGCCGGAATCGCTAGTAATCG
CGTATCAGCTACGACGCGGTGAATACGTTCCTGAGCCTTGTACACACCGCCCGTCACGTCATGGAAGCTGCTAGT
GCCCGAAGTCACCTCGTTTCAGGGGTGCCTACGGCAAGAGTGGTGACTGGGACG
```

There are two issues that result from this observation. Firstly, given the low sequence similarity to any known isolate, it is thrilling to address the question regarding the taxonomic level of this novel putative lineage within the phylum *Planctomycetes*. A phylogenetic analysis indicates that planctomycete strain 705e may be sufficiently different to the currently known genera within the single known familiy (*Planctomycetaceae*) of the *Planctomycetes* to justify possibly the introduction of a novel familiy (Fig. 5).

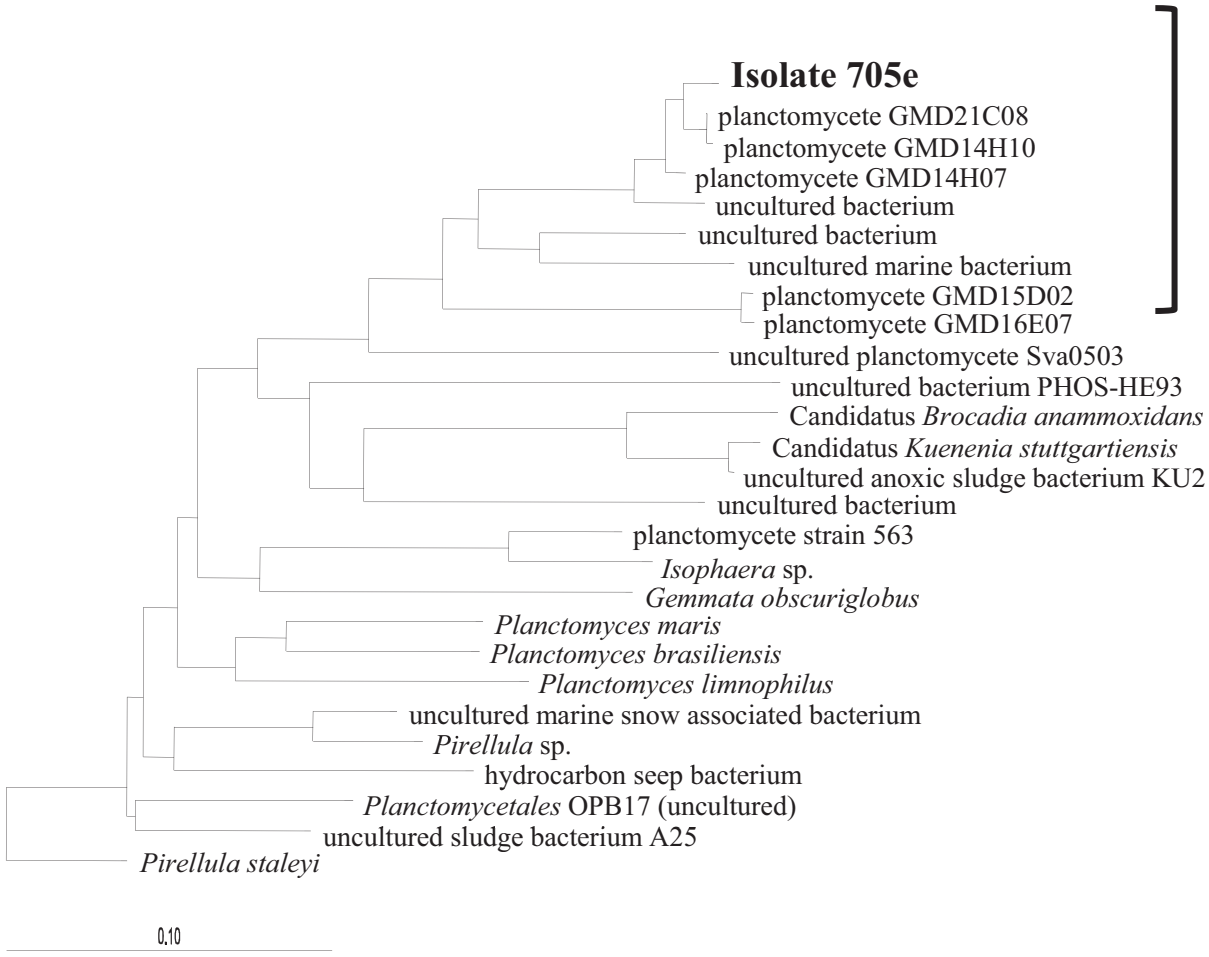


Fig. 5: Neighbor-Joining phylogenetic tree, with Jukes-Cantor corrections based on a nucleotide alignment of 1300 bp fragments of the 16S rRNA gene fragments of isolate 705 and representatives of the *Planctomecetales*. The representative sequences are mainly those that matched most closely that of isolate 705 (98% sequence similarity to Planctomycete GMD21C08). The closed bracket indicates the clade of *Planctomycetales* for which isolate 705 represents the first cultured strains. Sequence alignment and phylogenetic analysis was performed using the ARB program (http://www.mikro.biologie.tu-muenchen.de/pub/ARB/).

The second interesting observation reveals a potential caveat of this approach for microbial isolation: the hitherto only other representative isolates of this novel putative group of *Planctomycetes* were obtained by the same approach (Zengler *et al.* 2002). The 16S rRNA gene of isolate GMD14H07 (accession no.: AY162124.1) shares 98% sequence similarity (over 94% of the 1404-bp sequence submitted to NCBI for comparison by the BLAST approach) with the 16S rRNA gene of strain 705e. (The GMD produced by Zengler *et al.* (2002) were directly added to the PCR for amplification of 16S rRNA gene fragments for identification. Those isolates therefore do no longer exist.) It is important in this context to note that the samples analysed by Zengler *et al.* (2002) were collected from the Sargasso Sea (31°50’ N 64°10’ W and 32°05’ N 64°30’ W) at a depth of 3 m. In contrast to that, the sample used for the isolation effort by the Author derived from a sample collected on 25. August 2005 from surface waters of the English Channel at Station L4 (50º15' N, 04º13' W). Analyses of the microbial diversity within those environments, though at different time points, did not indicate a particular abundance of this novel group of *Planctomycetes* (Gilbert *et al.* 2012). This observation therefore seems to underline the notion that this co-culture approach affords a specific selective pressure on the microorganisms in the growth columns. Although this selection has to be judged fortuitous and beneficial in this particular study since it led to the isolation of a representative species of a novel lineage of *Planctomycetes*, it may, inversely, prevent the growth and culture of other novel groups of microorganisms.

2.2 Detailed and robust phylogenetic classification of microbial strains

Multi locus sequence analysis in microbial ecology

Another issue that microbiologists involved in the isolation of microbes often come across concerns the fact that isolation efforts usually result also in a number of strains that are – based on 16S rRNA sequence similarity – very similar to each other. This raises the question regarding the robust taxonomic and phylogenetic classification of the isolates. A solution to this challenge will be provided in the following.

A recent article in *Nature* (Editorial 2006) called for clinical microbiologists to “ctach up” with their colleagues working in basic research in microbiology, and to adopt their “crafty ways to hunt down microbes” – that is, to use the various methods in molecular biology that

are known as genetic fingerprinting techniques or metagenomics. However, while the Author – representing environmental microbiologists – agrees with this statement, he also believes that the opposite is true just as well! Several leading marine microbial ecologists called at a recent Royal Society discussion meeting on "*Species and speciation in micro-organisms*" (see Spratt *et al.* 2006; see also Gevers *et al.* 2005) for the application of multi-locus sequence analysis (MLSA) to be adopted to the characterisation of marine bacteria. MLSA is a technique that has been used for a long time by clinical microbiologists to characterise isolates of pathogens and to assess their origin and spread (Maiden *et al.* 1998). However, no studies had followed that adopted MLSA despite its potential to resolve some of the most pertinent questions relating to the biogeography of microorganisms.

For example, geographical isolation is considered to be the primary cause of speciation (e.g. Darwin, 1859). However, whether this concept – originally developed for macroscopic plants and animals – is also valid for microscopic species is still under vivid debate today. The origin of this debate goes back to the infamous principle that 'everything is everywhere, but, the environment selects' which was put forward by the Dutch microbiologist Bass Becking (1934). This hypothesis is based on the thoughts of 'cosmopolitan' microbes, a concept originally formulated by another Dutch microbiologist, Martinus Willem Beijerinck (1913). Bass Becking's hypothesis is of utmost relevance for discussions on the origin of microbial species, in particular in terms of the question as to whether allopatric speciation, the generally accepted process of evolution of species (Coyne and Orr, 2004), also represents the underlining mechanism for the marine microorganisms since dispersion rates in the oceans are much higher (in the order of months to years) than rates of microbial evolution.

To evaluate the usefulness of MLSA for the purpose of detailed biogeographic analysis of microorganisms a study was undertaken by the Author that aimed to apply MLSA to a range of isolates of a ubiquitous free-living marine bacterium, *Alteromonas macleodii* (Ivars-Martínez *et al.* 2008). A previous study using the ITS region as molecular marker had already demonstrated that the isolates of *A. macleodii* cluster into two phylogenetic clades: one composed of isolates from the sea surface and one in which the deep sea isolates group. Ivars-Martínez and colleagues (2008) then applied MLSA on a larger set of isolates of *A. macleodii* from diverse geographic origin and seawater depth, and used a new approach for the multi-variate statistical comparison of the phylogenetic distances of the strains for each of the nine marker genes. This study showed that, although phylogenetic analyses of each of the nine

different marker genes led to different phylogenies, analysis of the combined sequence data of the individual genes (i.e. in form of a concatemer) resulted in the clustering of the strains into the surface and deep sea clades, as well as a sub-clade formed by the five isolates from the Black Sea (Fig. 6).

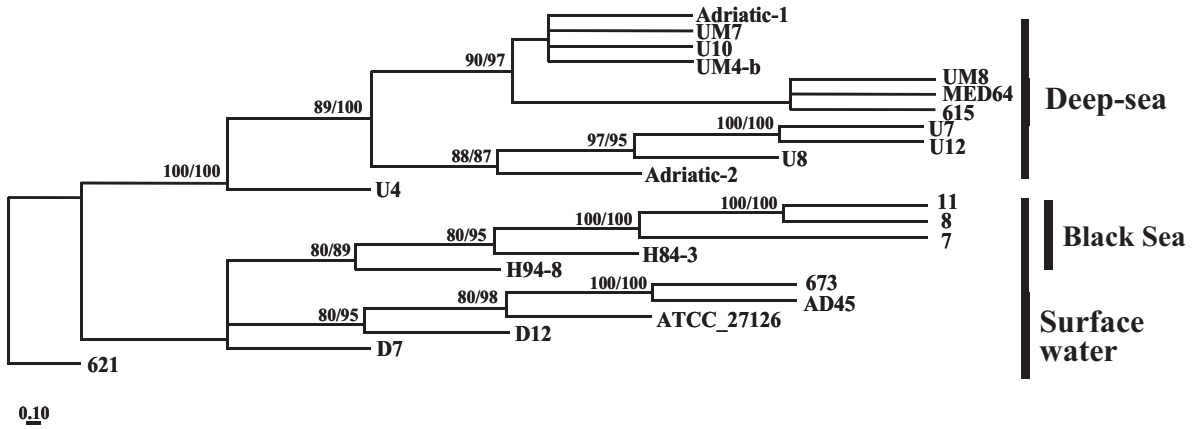


Fig. 6: Phylogenetic analysis of the concatenated nucleotide sequences (3932 – 4104 bp) from the 23 isolates of *A. macleodii*. The confidence of branch points was determined by two separate analyses (maximum parsimony – MP, neighbor-joining – NJ), with multifurcations indicating branch points that were collapsed using a strict consensus rule until supported in both analyses. The maximum parsimony tree was chosen to depict the phylogenic relationship in this 'second-stage consensus' tree. Values of 100 bootstrap replicates are indicated in brackets (calculated using the NJ / MP method) at branching points, but those < 80 are omitted. Bars to the right of the tree indicate the two clades based on the origin of strain isolation. (From Ivars-Martínez *et al.* 2008.)

This study therefore demonstrated for the first time for a large set of marine bacterial isolates that MLSA can overcome the problems of incongruencies in phylogenetic analyses caused by horizontal gene transfer (Rokas *et al.* 2003). In contrast to eukaryotic cells (e.g. yeasts: Rokas *et al.* 2003) where ca. 20 genes were estimated to be necessary for overcoming the impact of potential HGT events among the marker genes used for MLSA, it appears that the nine markers used for the classification of *A. macleodii* strains were sufficient for a robust identification. Whether this is a species specific feature or representative for the generally smaller microbial genomes is, however, unanswered.

Moreover, multi-variate statistical analyses of the substrate usage of the various isolates of *A. macleodii* revealed that the strains, by and large, grouped according to their depth of isolation (Fig. 7). So, the genetic differentiation as detected by molecular markers is also expressed in the phenotype of the strains, though both molecular markers and phenotypic features are functionally unrelated. This approach, therefore, has not only the potential to achieve intraspecific identification to the level of genotypes but, by comparison with relevant environmental data, also to reveal ecotypes.

In summary, it is argued that these findings strengthen the case that MLSA for the classification of microorganisms and, together with recent advancement in high-throughput single cell isolation and whole-genome amplification techniques (Stepanauskas and Sieracki 2007; Rodrigue *et al.* 2009), forms the basis for a new era in in the study of the biogeography of aquatic microbes with the aim to establish databases based on MLSA results for relevant environmental bacteria similar to those developed for bacterial pathogens (Ivars-Martínez *et al.* 2008).

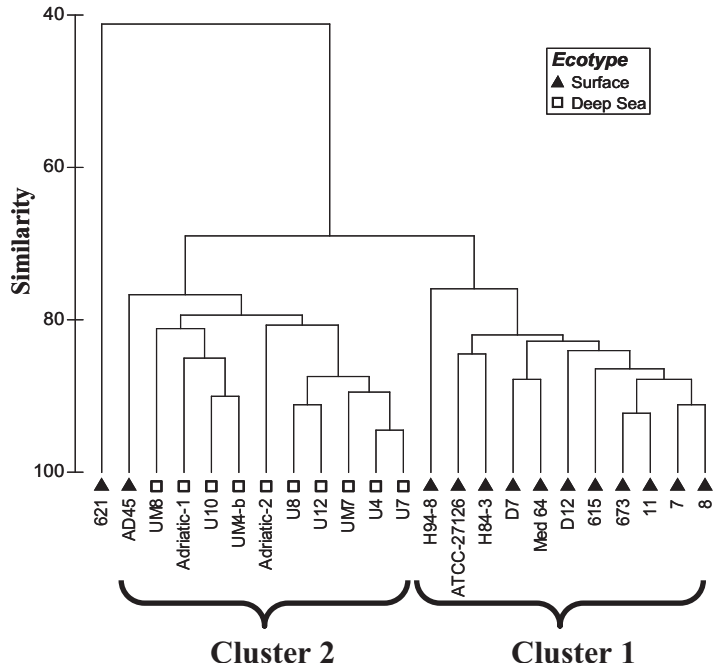


Fig. 7: Cluster analysis of substrate utilization using the 'Biolog Eco Plate'. A binary matrix was produced with 0 and 1 indicating no usage of a particular substrate or utilization, respectively. The data were analyzed using the 'simple matching co-efficient' (Sokal and Michener 1958). Closed triangles and open squares indicate surface water and deep sea isolates, respectively. (From Ivars-Martínez *et al.* 2008.)

Second-stage consensus trees as a simple and pragmatic technique to filter a tree for robust phylogenetic information

Despite the simultaneous use of multiple marker genes for the calculation of phylogenetic trees drawing, based on the branching within those, conclusions on the precise phylogenetic grouping of, for instance, a set of novel isolates has, however, to be profoundly justified and, therefore, requires further assessment of the reliable analyses. This is in particular the case if the branching in the phylogenetic trees is not congruent with that in published information: a situation that has also been observed by the Author (Mühling *et al.* 2008; see section 3.1.1).

While mathematicians still disagree on the algorithms to obtain the (most) correct tree resembling best the 'true' phylogenetic relationship of strains (i.e. sequences), the typical user of the phylogenetic analysis tools – being, by and large, an 'amateur-mathematician' – remains left alone with this problem. The way of solving this problem is presumably often directed by facts other than those purely based on mathematical reasoning; this possibly even includes the selection for the tree that best meets expectations. There are, however, relatively easy-to-use techniques available that help to obtain robust trees. For example, boostrapping – that is calculating a consensus tree of the results of numerous repeats of the analysis – provides a measure for trustworthiness of the branching to the various phylogenetic clusters. The branching can also be improved by filtering those positions in the alignment that show very high variability. Highly variable positions may lead to erroneous conclusions since the high frequency of mutations in a sequence may have led to reversal to the original nucleotide, though this strain has undergone a different evolution in comparison to the one without a single mutational event, but the same nucleotide at this position. Another approach available to every 'amateur' phylogeneticist employs a very pragmatic tactic to obtain a tree that reflects more likely the true phylogenetic relationship of strains since it involves calculating trees using various mainstream algorithms (i.e. neighbor-joining, maximum likelihood, maximum parsimony). One of those trees is then chosen as the representative for all, and those branching points within this tree that are not supported by each of the other methods, are then collapsed using a strict consensus rule until the branching is supported in all three analyses. Since such a tree forms the consensus of multiple (in that case of three) consensus trees it should be referred to as 'second stage consensus tree', a term introduced in the literature by the Author (Ivar-Martinéz *et al.* 2008). Although some may lament the resolution that is lost by this method, second-stage consensus trees are, however, more likely to only depict true branching and, thus, true phylogenies. To provide but one example, the Author

refers to the recognition that, based on sequence analysis of a 16S rRNA gene fragment library, a sample collected from a mesocosm filled with coastal seawater (i.e. mesotrophic in nature) harboured sequences that cluster within phylogenetic clades of the oligotrophic marine *Gammaproteobacteria* (OMG) (Mühling *et al.* 2008). The introduction of the OMG group was based on the fact that all available isolates had been found to be able to only grow in low nutrient (i.e. oligotrophic) media (Cho and Giovannoni, 2004). The robust phylogenetic analysis using the 'second stage consensus tree' approach (Mühling *et al.* 2008), however, provides robust evidence that the OMG group is genetically and, hence, metabolically more diverse than previously thought (Cho and Giovannoni, 2004) and appears to also include representative strains that are adapted to mesotrophic conditions.

2.3 Screening isolates for relevant enzymatic activities

Although it is laborious and time consuming to produce a culture collection of (novel) microbial isolates, once available it offers a variety of opportunities for subsequent research. Apart from the various facets in basic research it also includes screening of this microbial treasure chest filled with enzymes and other molecules for applications in biotechnology. Such an approach might be thought of as yesterday's methodology, now overtaken by the development and application of molecular techniques in the field of microbiology. That is, over the last 25 years it has become possible to access the huge genetic pool present in the various enironmental samples by culture independent means, mainly through PCR based amplification of relevant genes or cloning of genomic fragments in (high capacity) cloning vectors followed by sequence analysis or functional screening. However, despite an array of fancy techniques to handle the vast pool of microbial genes, it has so far proven difficult to exploit those genes for biotechnological applications *via* culture independent approaches. A major draw back remains the need for heterologous hosts with the appropriate promoters, regulators and supplies of relevant cofactors etc. for each of the cloned genes. The potential obstacles resulting from such requirements provides another lifeline to microbial isolation for biodiscovery (Mühling *et al.* 2013).

In this context it should be noted that most of the effort to isolate bacteria for biodiscovery has focused on heterotrophic bacteria from the terrestrial environment. Particular emphasis has been on members of the phylum *Actinobacteria* (Bull and Stach, 2007) and a large number of

enzymes of current industrial importance developed from terrestrial *Actinobacteria* (Bull *et al.* 2000). These bacteria are excellent candidates because they have large genomes with an extensive biochemical complement and are well established industrial producers of high levels of secondary metabolites.

However, over the last decade, as more metagenomic data have become available, it has been suggested that marine bacteria, other than *Actinobacteria*, possess a wide range of enzymes with novel substrate specificities and novel enzymatic activities (Wagner-Döbler *et al.* 2002) that should make them attractive candidates for biodiscovery. So far, few large-scale systematic analyses have been undertaken to assess their general suitability.

Therefore, 374 of the approx. 900 marine strains isolated by the Author (see section 2.1) were screened for 34 different enzymatic activities (Mühling *et al.* 2013). The selected activities used existing enzymatic assays but also targeted activities of particular biotechnological relevance; that is, there was perceived industrial demand, as well as a good level of scientific understanding, with a significant probability of progress leading to biotechnological advance.

The resulting data were then analysed with the focus on two particular aspects using multivariate statistical methods. Firstly, it seemed intriguing to assess whether members of specific (though broad) phylogenetic groups of bacteria are predominantly likely to harbour particular sets of relevant enzymatic activities. This may appear somewhat unlikely since phylogeny does not usually correlate to metabolic function. The fact that there are some notable exceptions to this general observation (e.g. the clades of bacteria and archaea responsible for ammonia and nitrite oxidation) however renders testing this hypothesis worthwhile. Therefore, all of the isolates were identified to the taxonomic level of genus based on sequence comparison of fragments of their 16S rRNA genes which then allows to apply multivariate statistical methods to reveal potential links between the phylogenetic identity of an isolate and its enzymatic activities. However, no significant correlation was detected between phylogeny and function at none of the various taxonomic levels. However, there was evidence of co-occurrence of some enzyme activity in the same isolate. That is, marine *Proteobacteria* had a higher complement of enzymes with biodiscovery potential than *Actinobacteria*; this contrasts with the terrestrial environment where the *Actinobacteria* phylum is a proven source of bioactive compounds. So, on the basis of this screening study, it would appear that marine *Alphaproteobacteria* and *Gammaproteobacteria* are just as likely to be high-priority candidates for biodiscovery research as *Actinobacteria*. The only enzyme

activities that occurred in a much higher proportion of the *Actinobacteria* isolates than other taxa were Baeyer-Villiger monooxygenase (BVMO) and lactone hydrolase activities (Fig. 8) – and then only in a small proportion (< 20%) of the *Actinobacteria* isolates.

Since, as mentioned above, enzyme screening studies have largely focused on terrestrial microorganisms, a second analysis of the data aimed at addressing the fundamental question as to whether bacteria from the terrestrial and marine environment are intrinsically different in terms of their enzymatic capabilities, a fact that would shed new insights into the elemental forces that shape microbial evolution and diversity. On a practical note, such information would also be welcome by those interested in biodiscovery since it would reveal the environment that poses the greater potential for relevant enzymes.

A particularly astonishing finding from this analysis was that some enzymes which are generally regarded as 'specialist' enzymes (that is they are not commonly found and are not required by bacteria for general cell functions), were by far more abundant among the marine cultures than would be expected on the basis of knowledge from terrestrial bacteria (Fig. 8). In particular laccase (in 355 of the 374 strains) and peroxidase (360 strains) activities correlated with the occurrence of other 'core' enzymes. Their biochemical and physiological role within the isolates remains, however, enigmatic. Nevertheless, an important aspect influencing the distribution of the two activities is the specific marine environment from which the isolates originated. All strains that did not show either both peroxidase and laccase activity (14 strains) or laccase activity (five strains) were derived from planktonic samples; both activities were, however, present in all of the isolates from the rocky shore environment. So, a function of these enzymes might be found in the context of biofilm growth or exposure to solar irradiance and short-term dessication events (i.e. at low tide). Additionally, it also indicates that biodiscovery programmes might be more effective if they targeted biofilm rather than planktonic bacteria.

A more detailed outline on the outcomes from this screening program and the assessment of the individual activities for biotechnological applications are restricted due to IP issues – with one exception: the presence of a BVMO activity.

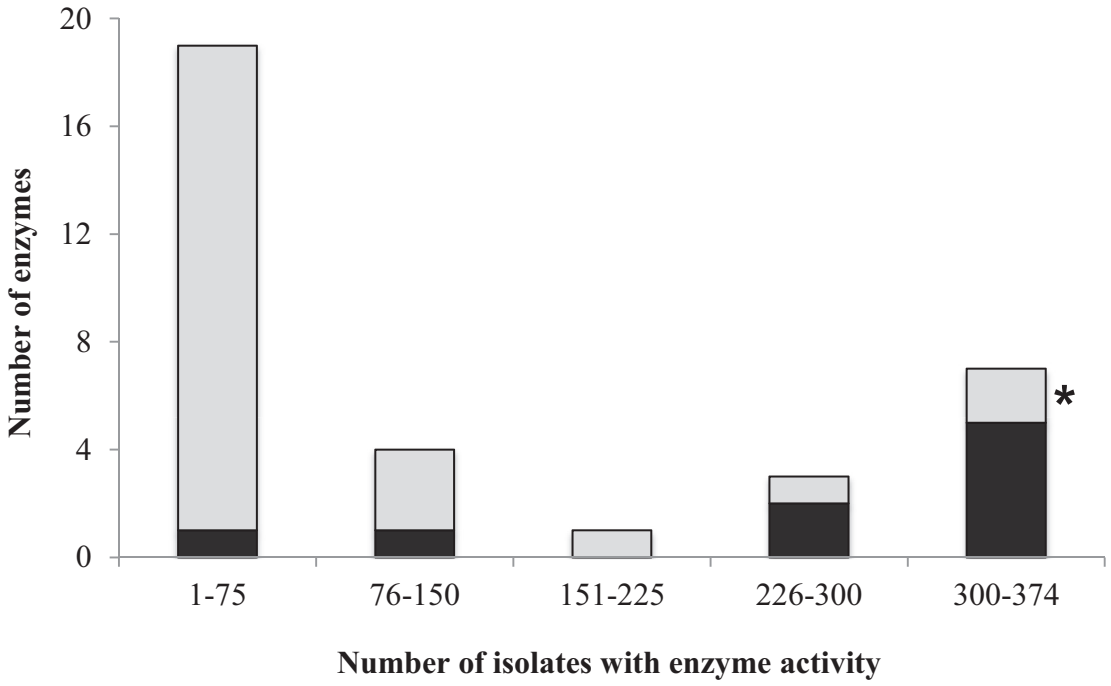


Fig. 8: Frequency of the occurrence of the 34 enzyme activities tested among the 374 isolates screened in the study. Black and grey bars indicate 'core' and 'specialist' enzymes, respectively. The asterik indicates the laccase and peroxidase activities that are generally regarded as 'specialist' activities. Based on Mühling *et al.* (2013).

As mentioned above, BVMO activities were the only enzymatic activities that were found at a higher rate among the *Actinobacteria* than in either of the *Alphaproteobacteria* or *Gammaproteobacteria*. Apart from the 374 strains included in the above multivariate statistical analysis a further set of 74 isolates were additionally screened for BVMO and lactone hydrolase (LH) activity using solid-phase screening (Willetts *et al.* 2012). Of this set of 448 strains 16 isolates were BVMO-positive with 12 of these 16 isolates containing an additional complementary LH activity. From this initial screen, isolate PML168 (BVMO+; LH+), a representative of *Stenotrophomonas maltophila* based on its partial 16S rRNA sequence, was chosen as the most promising candidate for additional characterization which was based on a genomic approach (Allen *et al.* 2012; section 2.4.1).

2.4 Genome analysis of microorganisms

Whole genome sequence analysis has gone a long way since the publication of the first microbial genome sequence 20 years ago (Fleischmann *et al.* 1995), mainly accelerated by the development of the low cost and high sequence output (in comparison to Sanger) next generation sequencing (NGS) approaches. However, the increasing availability and user-friendliness of the bioinformatic analysis tools must also be listed as an important parameter that has been contributing to the advances in microbial genomics. Therefore, genome sequence analysis provides a relatively fast approach to not only obtain detailed insights in the genetic pool and evolutionary heritage of a microorganism, but also for a more robust phylogenetic assignment than possible with individual genes (see section 2.2). However, it must be noted in this context that the amount of sequencing effort and time for subsequent bioinformatic sequence analysis depends on the scientific question in hand.

2.4.1 A novel type of Baeyer–Villiger monooxygenase activity from a marine gammaproteobacterium

For example, and to continue with the last theme, the Author proposed genome analysis for the identification and characterisation of the BVMO activity of bacterial isolate PML168. A more 'traditional' approach to obtain the sequence information (e.g. preparation of a genomic library and screening by Southern blot hybridisation for BVMOs using a heterologous probe) would have proved to be considerably more expensive and laborious than the genomic approach using, as was the case then, the Illumina sequencing platform.

Therefore, a draft genome sequence of strain PML168 was obtained and, though consisting of 562 contigs, confirmed both isolate PML168 as a strain of *Stenotrophomonas maltophila* by phylogenetic analysis of the 16S rRNA gene and the presence of a BVMO (Allen *et al.* 2012). BVMOs can be categorised into two types with type 1 representing obligate FAD plus NADPH dependent members of the Class B flavoprotein superfamily whereas type 2 BVMOs are obligate FMN- plus NADH-dependent members of the Class C flavoprotein superfamily (van Berkel *et al.* 2006) and consist of a loosely bound trimeric assemblage ($\alpha_2\beta$) of two different polypeptide types (Willetts *et al.* 2012). The single CDS in the genome of strain PML168 encoding the BVMO implies the presence of two GxGxxG motifs, indicative of two

Rossman-fold motifs involved in the binding of the AMP moieties of FAD and NAD(P)H (Vallon 2000) and the signal sequence FxGxxxHxxx(W/Y) for Class B flavoproteins, a feature that clearly distinguishes Class B flavoproteins from other flavin-dependent proteins (van Berkel *et al.* 2006).

However, while the BVMO of *S. maltophila* strain PML168 was shown to perform 'standard' BVMO-catalysed reactions it also contains features that are rather atypical for Class B type 1 BVMO flavoproteins, suggesting that this enzyme from strain PML168 represents a novel putative type 3 BVMO (Willetts *et al.* 2012). The main defining feature for the new type 3 group of Class B BVMOs is connected to the fact that the gene product consists of 357 amino acids (38.4 kDa) which is below the average size (>510 amino acids) known for type 1 BVMOs, and the related preference of NADH as the reduced nicotinamide nucleotide. The smaller size is due to the truncated C-terminal domain which is confirmed by the genome derived nucleic acid sequence of the encoding gene. This, in turn, results in the absence of amino acids critical for accomodating the 2'-phosphate group of the adenosine ribose that distinguishes NADPH from NADH (Kamerbeek *et al.* 2004) and, hence, support NADPH-dependency that is typical of almost all Class B flavoproteins (van Berkel *et al.* 2006). This novel type of BVMOs may therefore be of interest for biotechnological applications as it combines the higher operational stability of type 1 enzymes with the NADH dependence of type 2 enzymes. Finally, although this study reports for the first time the natural existence of the novel type 3 of Class B flavoproteins bioinformatic screening of metagenomic and metatranscriptomic datasets revealed that this group of enzymes is widely distributed in the marine environment. (Willetts *et al.* 2012).

2.4.2 Genome analysis of acid mine drainage related acidophilic bacteria

In the case described above genome analysis was performed to obtain relevant genetic information on those activities that were detected by a screening programme using biochemical assays. Apart from using genome sequencing as supporting tool to obtain relevant information on specific genes and operons it can also be used to reconstruct the metabolic capability of strains whose physiological and biochemical characterisation proves to be difficult due to the lack of pure cultures or due to poor and 'unreliable' growth in laboratory cultures. Such a scenario is commonly experienced by environmental

microbiologists, in particular by those working on microorganisms adapted to extremely oligotrophic conditions or chemolithoautotrophic isolates from extreme environments.

Research along these lines of genome analyses were originally planned on cultures obtained *via* isolation efforts from the marine environment; in fact, the first genome sequencing project of the Author was that of the marine gammaproteobacterium *S. maltophila* PML168 (see above). However, due to the Author's geographical move from a marine research base (Plymouth Marine Laboratory – PML) to the TU Bergakademie Freiberg, the choice of target organisms had to be adapted to the new employer's profile, that is geomicrobiology. Nevertheless, the scientific aims remained the same, that is the reconstruction of the metabolic potential of novel microorganisms.

One aspect in geomicrobiology that has been addressed by the Author concerns acidophilic bacteria in acid mine drainage (AMD) water. The main reason for the formation of AMD is connected to the fact that the geological formation in the East German mining district (Lusatia, Saxony) is rich in pyrite (FeS_2) which is unstable in water in the presence of dissolved oxygen (or ferric iron: Fe^{3+}). This together with the fact that mining activities result in massively increased exposure of pyrite to water and dissolved oxygen form the basis for the large scale oxidation of pyrite *via* equation (1), thus producing acidity and a high sulfate and iron load of mine drainage water.

$$FeS_2 + 3.5\ O_2 + H_2O \rightarrow Fe^{2+} + 2\ SO_4^{2-} + 2\ H^+ \qquad (1)$$

The ferrous iron formed *via* this reaction is unstable at neutral pH in the presence of oxygen; that is, ferrous iron is oxidised to ferric iron which, in turn, is a potent oxydant of pyrite, thus resulting in further acidification of mine drainage water (equation 2).

$$FeS_2 + 14\ Fe^{3+} + 8\ H_2O \rightarrow 15\ Fe^{2+} + 2\ SO_4^{2-} + 16\ H^+ \qquad (2)$$

Again, the newly formed ferrous iron is then abiotically oxidised to ferric iron as long as the pH of the mine drainage water is around 4 and above, but slows down at lower pH due to the increasing stability of ferrous iron under decreasing pH conditions (e.g. < 4) even in the presence of dissolved oxygen. This rate limiting step of AMD formation has become an environmental niche for microorganisms that oxidise ferrous iron to ferric iron (equation 3) in order to obtain energy (ATP) and to recycle reduction equivalents (NAD(P)H) for metabolic processes.

$$4\ Fe^{2+} + O_2 + 4\ H^+ \rightarrow 4\ Fe^{3+} + 2\ H_2O \quad (3)$$

Although the energy yield from this redox reaction is relatively low (~30 kJ mol^{-1} at pH 2: Hedrich *et al.* 2011a) it still satisfies the energy requirement for growth within the extreme environment where selection pressure is high and competition low. Ferric iron also precipitates in form of iron rich hydoxides which again results in the liberation of further protons (equation 4).

$$4\ Fe^{2+} + O_2 + 10\ H_2O \rightarrow 4\ Fe(OH)_3 + 8\ H^+ \quad (4)$$

The mineral formed during the precipitation process depends on the pH, with schwertmannite ($Fe_8O_8(OH)_{5.5}(SO_4)_{1.25}$) as the predominant mineral at a pH at around 3 (Bigham *et al.* 1990). Schwertmannite is of interest since its structure and high surface area results in high absorption capacity for ions (e.g. arsenic) and its rusty colour provides a pigment that can be added to paints (Hedrich *et al.* 2011b; Janneck *et al.* 2010).

Interestingly, microorganisms that provide an alternative path to the pH induced restraint in abiotic pyrite oxidation, the acidophilic iron oxidiser, can also be used to remediate the environmental problem they cause. This approach is employed within the pilot plant Tzschelln (open-pit lignite mine Nochten within the east German mining district) for the remediation of AMD. This pilot plant comprises a 10-qm^3 tank with arrays of artificial carrier material as surface to improve mineral formation and precipitation. Sustaining a hydraulic retention time for the AMD (pH 4.5 – 4.8) of approximately 8 h maintains pH 3 within the oxidation tank of the pilot plant, which results in the formation of schwertmannite.

Previous research programmes at the Institute of Biological Sciences of TUBAF have revealed that the microbial assemblage within this pilot plant is dominated by two groups of *Betaproteobacteria*: one which, based on their 16S rRNA sequence, is closely related to the known but neutrophilic genus *Gallionella* (hence referred to as *Gallionella*-like bacteria); and the second group represents a novel putative genus ("*Ferrovum*") and, possibly, novel putative order ("*Ferrovales*") of *Betaproteobacteria* (Heinzel *et al.* 2009a,b; Johnson *et al.* 2014). Using a newly developed medium that simulates the inorganic components of the AMD within the pilot plant (Tischler *et al.* 2013) it was possible to bring various strains of "*Ferrovum*" into laboratory cultures. Unfortunately however, all attempts to obtain pure (clonal and axenic) cultures of these strains have so far been unsuccessful despite a concerted programme of research involving a variety of approaches, including the use of purified

agarose (Schober 2011) and of alternative solidifying agents (Pissoke 2013). Interpretation of results from physiological and biochemical experiments are, therefore, problematic due to the unknown contribution of the contaminant and the varying ratio of contaminant to "*Ferrovum*" cells. The knowledge available on "*Ferrovum*" was therefore limited to that of the designated type strain, "*F. myxofaciens*" P3G, the hitherto only strain available in pure culture (Johnson *et al.* 2014). Culturing of the *Gallionella*-like acidophilic (or acidotolerant) strains in the laboratory environment has so far been even entirely unsuccessful (see section 3.3.2).

Since "*Ferrovum*" and *Gallionella*-like strains within the pilot plant Tzschelln are likely to play a fundamental role in the performance of the biotechnological process, it was aimed to obtain insights into their metabolic potential and their strategy to adapt to the acidic environment by circumventing culture associated problems through application of genomic and metagenomic approaches.

Genome analysis of "*Ferrovum*" sp. JA12 and its contaminant *Acidiphilium* sp. JA12-A1

The problem of the contamination of the "*Ferrovum*" cultures with *Acidiphilium* sp. was addressed in a two-tiered approach. Firstly, using improved culturing techniques with tightly controlled ferrous iron concentrations in the medium (kept between 5 mM and 15mM) it was possible to obtain "*Ferrovum*" cultures that, based on TRFLP analysis, consisted of up to ≥ 99% of "*Ferrovum*" cells (B. Erler, unpubl. results). The second part of the approach to resolving the problem was based on the much simpler isolation of *Acidiphilium* sp. JA12-A1 from the mixed culture using a heterotrophic medium (SJH medium: Johnson and McGinness 1991; Johnson and Hallberg 2007) unsuitable for growth of the chemolithotrophic "*Ferrovum*" (or other chemolithoautotrophic iron oxidising) strains. The successful isolation, in turn, permitted the independent analysis of the genome of *Acidiphilium* sp. strain JA12-A1 and, thus, the subsequent identification and removal of *Acidiphilium* derived sequence reads from the raw sequence data obtained for the mixed culture. This approach, albeit more costly and time consuming, was thought to be more reliable than building the identification of *Acidiphilum* sequence reads merely on bioinformatic interpretations (e.g. the higher GC content of *Acidiphilium*; differences in codon usage; etc.).

Sequencing and automated assembly and annotation of the sequence reads was performed at the G2L (Goettingen University: Dr. Anja Poehlein) by a hybrid approach combining both

pyrosequence analysis and sequencing using an Illumina platform (except in the case of "*Ferrovum*" strain Z-31 which was sequenced only by the Illumina approach).

Apart from providing support for the assembly of "*Ferrovum*" derived sequence reads obtained from the mixed culture the genome sequencing of *Acidiphilium* JA12-A1 was also persued in the context of the tenacious association between strains of the heterotrophic *Acidiphilium* and chemolithotrophic iron oxidising bacteria that has long been known. That is, a second aim was to reveal metabolic properties that are fundamental for the seemingly synthrophic interaction between *Acidiphilium* sp. JA12-A1 and the co-occurring chemolithoautotrophic iron oxidizer (Ullrich *et al.* 2015).

The genome sequence of *Acidiphilium* sp. strain JA12-A1 consists of 4.18 Mbp on 297 contigs and harbors 4015 protein-coding genes and 50 RNA genes. A phylogenetic analysis of the relationship of strain JA12-A1 to other *Acidiphilium* isolates indicates its position within the subgroup that contains the type strain of the genus, *A. cryptum* Lhet2 (Fig. 9). Moreover, the molecular and functional organization of the *Acidiphilium* sp. JA12-A1 draft genome proved to be similar to those of the close relatives *A. cryptum* JF-5, *A. multivorum* AIU301 and *Acidiphilium* sp. PM DSM 24941, thus extending the close relationship between these strains as detected by 16S rRNA based phylogenetic analysis (Fig. 9). This and the highly similar metabolic potential encoded in the individual genomes supports the idea that other *Acidiphilium* strains play a similar role in AMD communities to that reconstructed from the genome of *Acidiphilium* sp. strain JA12-A1 (Ullrich *et al.* 2015).

Of particular interest in this respect are metabolic properties of *Acidiphilium* sp. JA12-A1 that are fundamental for the synthrophic interaction with the co-occurring chemolithoautotrophic iron oxidiser: in this case strain(s) of "*Ferrovum*". This syntrophic relationship is probably most clearly expressed by the potential interspecies carbon transfer that may follow a similar path to that described and experimentally confirmed for *Acidiphilium* sp. and the acidophilic iron oxidiser *Acidithiobacillus* (Kermer *et al.* 2012). This potential carbon transfer involves *Acidiphilium* sp. JA12-A1 excreting polysaccharide hydrolyzing enzymes, such as β-glucosidases or endoglucanases, to break down cell envelope polysaccharides from decaying cells and from the co-occurring iron oxidizer that is related to "*F. myxofaciens*" P3G (Johnson *et al.* 2014). Monosaccharides originating from polysaccharide hydrolysis or from lysed cells are taken up by *Acidiphilium* sp. JA12-A1 *via* specific uptake systems to produce bacterial biomass. Alternatively, the monosaccharides or parts thereof are oxidized to gain energy for

the cellular metabolism. Under aerobic conditions the electron donor is completely oxidized to carbon dioxide which is the preferred carbon source for the autotrophic iron oxidizer. However, the iron oxidizer may not only profit from the local increase of the carbon dioxide availability but also from the removal of organic compounds by *Acidiphilium* sp. JA12-A1, since chemolithoautotrophic iron oxidizers have long been known to be sensitive to organic compounds (Touvinen and Kelly 1973). The sum of these potential interactions may account for the tenacious association of both organisms in the mixed culture and provide an explanation for the difficulties encountered when attempting to obtain pure cultures of the iron oxidizing bacteria.

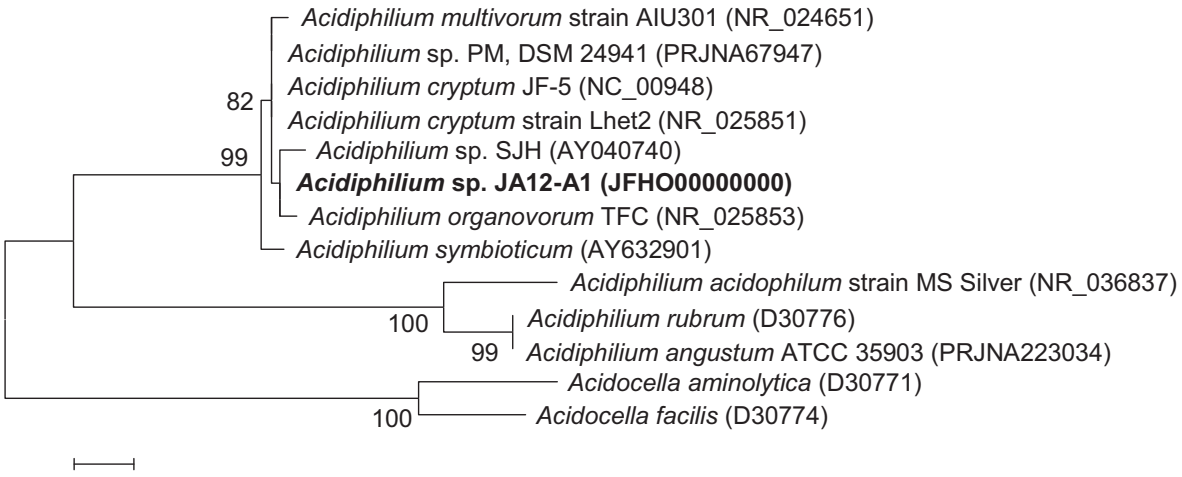


Fig. 9: Dendrogram of representative strains of the alphaproteobacterial genus *Acidiphilium* – based on their partial 16S rRNA gene sequences. The dendrogram was calculated with MEGA5 (Tamura *et al.* 2011) using the Maximum Likelihood method based on the Jukes-Cantor model (Jukes and Cantor 1969). The analyzed sequences were aligned by CLUSTALW (Larkin *et al.* 2007). The clustering of the sequences was tested by the bootstrap approach with 1000 repeats. The length of the tree branches was scaled according to the number of substitutions per site (see size bar). *Acidocella aminolytica* (D300771) and *Acidocella facilis* (D30774) were used as outgroup. Whole genome sequences are only available for *Acidiphilium* cryptum JF-5, *Acidiphilium* multivorum AIU301, *Acidiphilium* sp. PM DSM 24941 and *Acidiphilium angustum* ATCC 35903 (GOLD project IDs: Gc00559, Gc01862, Gi09776, Gi0051610; accession numbers: NC_009484, NC_015186; PRJNA67947, PRJNA223034). Taken from Ullrich *et al.* (2015).

Comparative genome analysis of four "*Ferrovum*" strains

A total of three "*Ferrovum*" strains, two (JA12, PNG-185) isolated by Tischler *et al.* (2013) and strain Z-31 isolated by Mrs. Sarah Vogel (unpubl. results), were selected for genome analysis since they represent different phylogenetic clusters or sub-clusters within "*Ferrovum*" (Fig. 10). The genome of strain JA12 represents the most extensively analysed of the three "*Ferrovum*" genomes for two simple reasons: it was the first to be determined and it is almost completely closed (Table 2).

The in depth analysis of the genome of strain JA12 involved, in essence, its taxonomic and phylogenetic assignment, the reconstruction of its metabolic potential and the comparison of its genome with that of the proposed type strain '*Candidatus* Ferrovum myxofaciens' P3G (Johnson *et al.* 2014). The findings from this study and the subsequent comparison of the total of four available genome sequences of "*Ferrovum*" strains is described in great detail in Ullrich *et al.* (2016a) and Ullrich *et al.* (2016b), respectively. So, only a few representative conclusions are briefly mentioned here.

Phylogenetic analysis (Fig. 10) of the 16S rRNA gene sequences of "*Ferrovum*" strains together with those from other iron oxidising and non-iron oxidising members of the phylum *Proteobacteria* indicates that strain JA12 and PNG-185 represent cluster 3 of the "*Ferrovum*" clades suggest by Tischler *et al.* (2013) while strain Z-31 is very similar to the type strain P3G (i.e. cluster 1).

Fig. 10: Phylogenetic relationship of iron oxidising and non-iron oxidising members of the phylum *Proteobacteria* based on 16S rRNA marker (next page). Sequences of the 16S rRNA gene fragments were imported into the ARB software program and aligned to other proteobacterial 16S rRNA gene sequences using the automated alignment tool within ARB (Ludwig *et al.* 2004). Calculation of phylogenetic trees based on these sequence alignments (*E. coli* positions 887-1365) was conducted within MEGA6 (Tamura *et al.* 2013) using the neighbor-joining method with Jukes-Cantor corrections (Jukes and Cantor 1969). Numbers next to branches indicate the percentage of replicates (out of 1,000 bootstrap trees) in which the associated taxa clustered together (Felsenstein 1985). *Bacillus subtilis* was used as outgroup. Taken from Ullrich *et al.* (2016a).

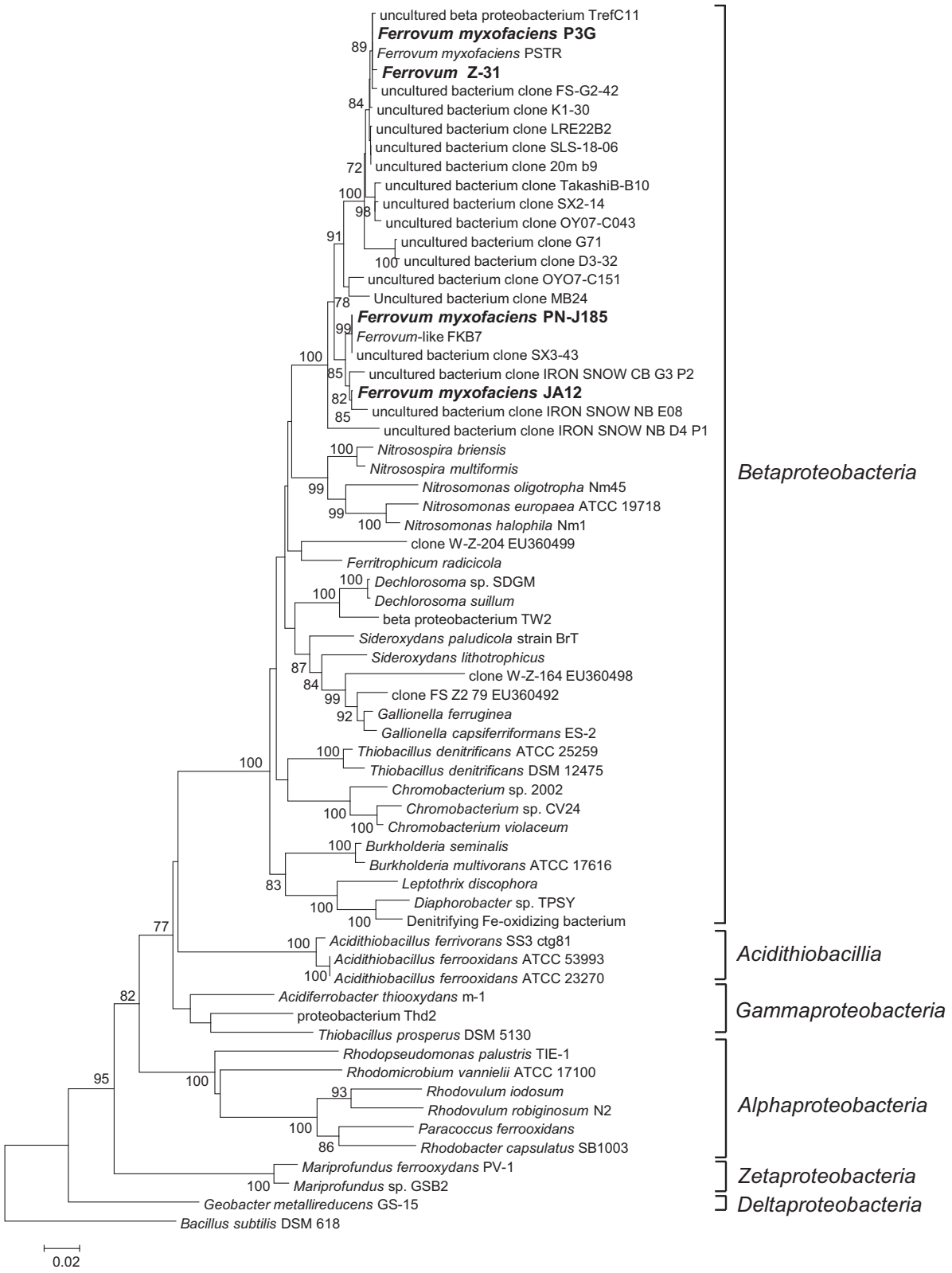

uncultured beta proteobacterium TrefC11
Ferrovum myxofaciens P3G
Ferrovum myxofaciens PSTR
Ferrovum Z-31
uncultured bacterium clone FS-G2-42
uncultured bacterium clone K1-30
uncultured bacterium clone LRE22B2
uncultured bacterium clone SLS-18-06
uncultured bacterium clone 20m b9
uncultured bacterium clone TakashiB-B10
uncultured bacterium clone SX2-14
uncultured bacterium clone OY07-C043
uncultured bacterium clone G71
uncultured bacterium clone D3-32
uncultured bacterium clone OYO7-C151
Uncultured bacterium clone MB24
Ferrovum myxofaciens PN-J185
Ferrovum-like FKB7
uncultured bacterium clone SX3-43
uncultured bacterium clone IRON SNOW CB G3 P2
Ferrovum myxofaciens JA12
uncultured bacterium clone IRON SNOW NB E08
uncultured bacterium clone IRON SNOW NB D4 P1
Nitrosospira briensis
Nitrosospira multiformis
Nitrosomonas oligotropha Nm45
Nitrosomonas europaea ATCC 19718
Nitrosomonas halophila Nm1
clone W-Z-204 EU360499
Ferritrophicum radicicola
Dechlorosoma sp. SDGM
Dechlorosoma suillum
beta proteobacterium TW2
Sideroxydans paludicola strain BrT
Sideroxydans lithotrophicus
clone W-Z-164 EU360498
clone FS Z2 79 EU360492
Gallionella ferruginea
Gallionella capsiferriformans ES-2
Thiobacillus denitrificans ATCC 25259
Thiobacillus denitrificans DSM 12475
Chromobacterium sp. 2002
Chromobacterium sp. CV24
Chromobacterium violaceum
Burkholderia seminalis
Burkholderia multivorans ATCC 17616
Leptothrix discophora
Diaphorobacter sp. TPSY
Denitrifying Fe-oxidizing bacterium
Acidithiobacillus ferrivorans SS3 ctg81
Acidithiobacillus ferrooxidans ATCC 53993
Acidithiobacillus ferrooxidans ATCC 23270
Acidiferrobacter thiooxydans m-1
proteobacterium Thd2
Thiobacillus prosperus DSM 5130
Rhodopseudomonas palustris TIE-1
Rhodomicrobium vannielii ATCC 17100
Rhodovulum iodosum
Rhodovulum robiginosum N2
Paracoccus ferrooxidans
Rhodobacter capsulatus SB1003
Mariprofundus ferrooxydans PV-1
Mariprofundus sp. GSB2
Geobacter metallireducens GS-15
Bacillus subtilis DSM 618
0.02
Betaproteobacteria
Acidithiobacillia
Gammaproteobacteria
Alphaproteobacteria
Zetaproteobacteria
Deltaproteobacteria

Table 2: Comparison of general genome features of four "*Ferrovum*" strains. (Taken from Ullrich *et al.* 2016b.)

	„*F. myxofaciens*“	**„*Ferrovum*“ strain from pilot plant**		
	P3G [1]	**Z-31**	**JA12**	**PN-J185**
Accession number	JPOQ01000000	LRRD00000000	LJWX00000000	LQZA00000000
Genome size (bp)	2,702,191	2,473,519	1,995,737	1,889,241
Number of contigs	647	212	3	6
G+C content (%)	54.3	54.3	44.5	39.9
Number of coding sequences	2,859	2,500	1,970	1,889
Number of protein-coding sequences	2,785	2,456	1,882	1,844
Number of RNA genes	53	44	43	45
Number of pseudogenes	20	?	45	?

[1] Moya-Beltrán *et al.* (2014)

Species nomination demands DNA-DNA hybridisation as proof (Wayne *et al.* 1987) which can not be conducted in the case of "*Ferrovum*" due to the lack of pure cultures. However, using the distance-to-distance calculator (Meier-Kolthoff *et al.* 2013) to simulate *in silico* DNA-DNA hybridisation (Wayne *et al.* 1987) the hypothesis was tested as to whether strains PN-J185 and JA12 denote novel species. The results obtained from this analysis and from a second mainstream genomic comparison, the average nucleotide identity (ANI: Konstantinidis and Tiedje 2005), suggest that strains JA12 and PNJ-185 represent (at a minimum) a novel species within the genus "*Ferrovum*" (Table 3). This notion is further supported by several genomic features that are common to JA12 and PNJ-185, while differing to those of the type strain: for example, the lower G+C content and the streamlined genome (see Table 2). As for the latter, it must, however, be considered that the genome sequence of the set of strains that represent these two species is of very different quality. That is, the precise genome size will change with closure of gaps between contigs. Nevertheless, it is noteworthy that the genomes of these strains are rather small when compared to other acidophilic iron oxidising bacteria (e.g. the 3.6 Mbp of *A. ferrooxidans*). Small genomes are certainly advantageous since phosphate is likely to be a limiting nutrient in environments rich in dissolved iron. Reducing the energetic cost of cellular housekeeping and to save nutrients by genome reduction in the hope that the required activities are provided by other members of the community – the Black Queen hypothesis: Morris *et al.* (2012) – appears, however, to be a dangerous strategy. Although the Black Queen hypothesis was developed based on the discovery that the planktonic cyanobacterium *Prochlorococcus* requires other bacteria to break down toxic hydrogen peroxide, Morris *et al.* (2012) also stress its density-dependent character. Unfortunately, information on cell numbers within the pilot plant Tzschelln have not been published, though spot checks indicate that it is rather low (10^{-4} – 10^{-5} cells/mL: S. Hedrich, personal communication). Nevertheless, the density-dependent character may also be regarded as support for the notion that both group 1 and group 2 strains (see below) have a biofilm associated life style for which cooperation among microorganisms within biofilms has been experimentally demonstrated (e.g. Drescher *et al.* 2014; Pande *et al.* 2016). Biofilm growth has been deduced from the observed production of copious amounts of extracellular polymeric substances (EPS) by the type strain P3G ('*Candidatus* Ferrovum myxofaciens': myxofaciens = "slime maker"; Johnson *et al.* 2014), but not by group 2 strains JA-12 and PN-J185.

Table 3: Genome-based phylogenetic indicators of the four "*Ferrovum*" strains. Calculation of ANIb (blastn-based average nucleotide identity) and tetra (tetranucleotide composition regression) values using JSpecies (Richter and Rosselló-Móra 2009).

indicator	strain	P3G	Z-31	PN-J185	JA12
ANIb (%) *	**Z-31**	99.31	100.00	65.40	65.75
	PN-J185	65.84	65.40	100.00	73.72
	JA12	66.40	65.75	73.72	100.00
Tetra *	**Z-31**	0.998	1.000	0.550	0.602
	PN-J185	0.555	0.550	1.000	0.951
	JA12	0.611	0.602	0.951	1.000

* Values below the thresholds of ≤ 95 % (ANI) and ≤ 0.99 (tetra) indicate that strains belong to different species (Richter and Rosselló-Móra 2009).

Comparing the reconstructed metabolism of the four available "*Ferrovum*" strains also supplies evidence for the existence of multiple species within "*Ferrovum*". In the absence of species nominations these are referred to as groups: group 1 for type strain P3G and Z-31 and group 2 for strains JA-12 and PN-J185 (Ullrich *et al.* 2016b). For example, group 1 strains (i.e. '*Candidatus* Ferrovum myxofaciens') encode genes that confer cell motility and chemotaxis and the ability to fix molecular nitrogen *via* a nitrogenase, while neither of these traits is encoded in the genomes of group 2 strains (Ullrich *et al.* 2016b; Fig. 11). In contrast to this, group 2 strains are able to use a wider range of nitrogen sources including nitrate and urea which can not be utilised by group 1 strains (Ullrich *et al.* 2016b). Apart from extending the availability of nitrogen the urease encoding gene cluster may have also been acquired by strain JA12 as a strategy in pH homeostasis. This hypothesis (Ullrich *et al.* 2016a) has been built on the findings obtained for the pathogen *Helicobacter pylori* which uses urea degradation via urease activity to survive the acidic conditions of mammalian stomach fluids (Eaton *et al.* 1991; Sachs *et al.* 2005). Experimental evidence for the same role in microorganisms (in this case *Thiomonas* sp. strain CB2) living within AMD has recently been provided (Farasin *et al.* 2015). Additionally to its role in pH homeostasis, this recent study further demonstrated that urea degradation also promotes the precipitation of metals and metalloids toxic to the cell (Farasin *et al.* 2015).

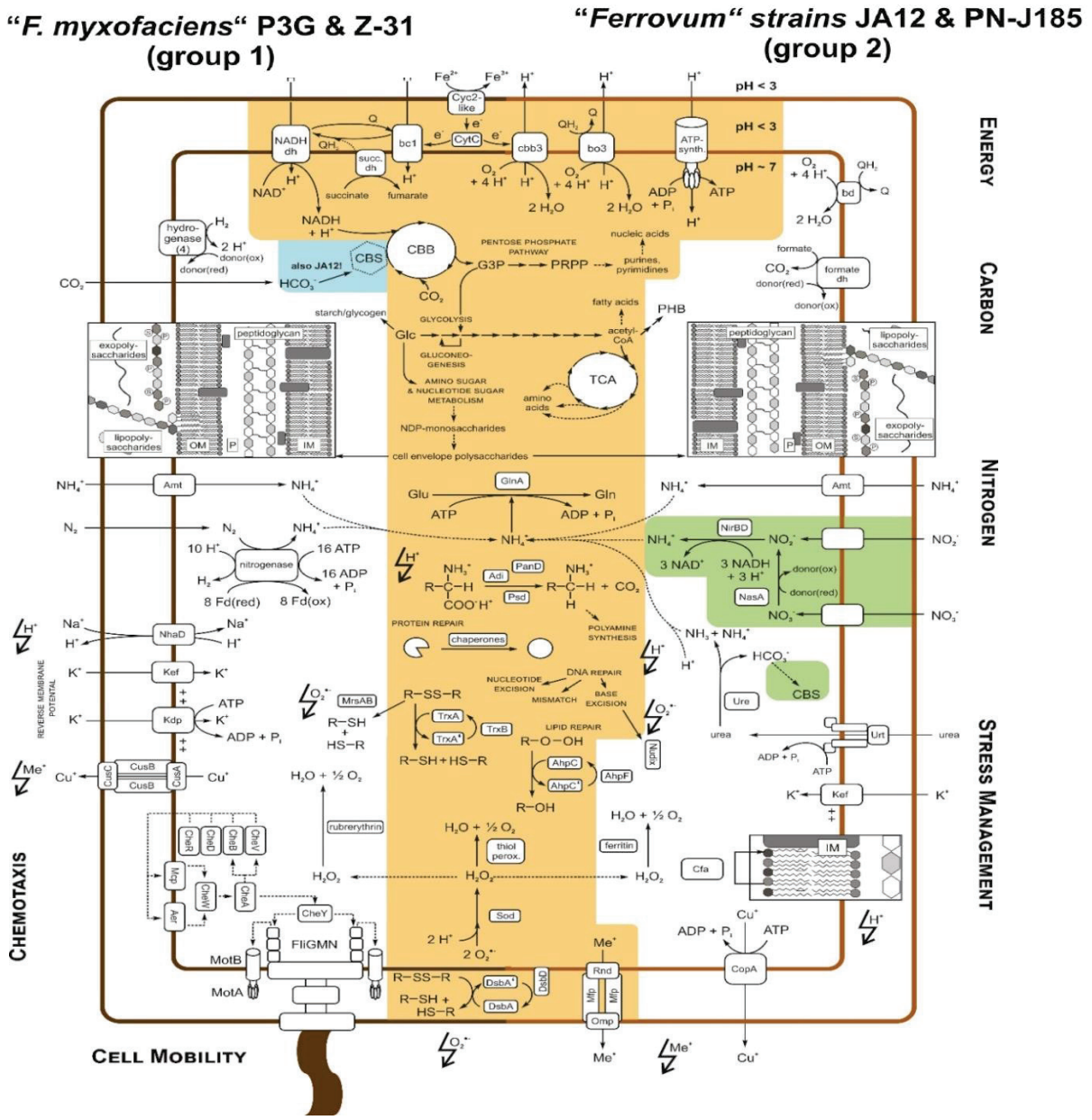


Fig. 11: Comparison of predicted metabolic potentials of group 1 and group 2 "*Ferrovum*" strains. The comparison is focused on energy metabolism, carbon and nitrogen metabolism, strategies of stress management with regard to the acidic pH (lightning with proton), oxidative stress (lightning with superoxide radical), high metal loads (lightning with metal ion) typical for their natural environment and chemotaxis and motility. Orange areas show metabolic traits predicted to be common in all four strains. Metabolic traits only predicted in group 2 strain JA12 are shown in green while traits that were only predicted in the group 1 strains and group 2 strain JA12 are colored in blue. Taken from Ullrich *et al.* (2016b) which is referred to for a detailed description and for a list of genes involved in the individual activities.

Genome analysis of two iron oxidising actinobacteria

The genomes of two further acidophilic iron oxidisers were sequenced in addition to the chemolithoautotrophic "*Ferrvoum*" strains. These strains, which belong to the phylum *Actinobacteria*, were isolated and provided by Prof. Johnson (Bangor University, UK) as pure (clonal and axenic) cultures. Both strains are either type strains or have been proposed as such for their respective genus: *Ferrimicrobium acidiphilum* $T23^T$ (Johnson *et al.* 2009) and *Acidithrix ferrooxidans* Py-F3 (Jones and Johnson 2015), respectively. Genome sequencing was performed using the Illumina platform at G2L (Goettingen University) resulting in draft genomes of 3.08 Mb and 4.02 Mb for *Acidithrix ferrooxidans* and *F. acidiphilum*, respectively. Several metabolic features predicted from the annotated genome sequences are summarised in two genome announcements (Eisen *et al.* 2015 a,b). To provide evidence for the value of this sequence information the following example shall briefly be outlined. Based on their characterisation both strains were judged to be obligate heterotrophes (Johnson *et al.* 2009; Jones and Johnson 2015) and the genome sequence of *F. acidiphilum* supports this finding. In contrast to this, the genome of *Acidithris ferrooxidans* Py-F3 encodes the two subunits of a type I RubisCO and several enzymes required for carbon fixation *via* the Calvin-Benson-Bassham cycle, indicating that strain Py-F3 might have the potential to grow autotrophically with CO_2 as the carbon source. This, however, stands in contrast to the fact that, using an established PCR-based approach to detect RubisCo encoding genes (*cbbl*, *cbbm*), Jones and Johnson (2015) did not obtain any PCR amplicons, thus indicating inappropriate PCR conditions (including the primer sequences). The concomitantly observed lack of growth of *Acidithrix ferrooxidans* Py-F3 in the absence of yeast extract may be due to a potential growth factor requirement which, if met, might facilitate autotrophic growth (Jones and Johnson 2015). Whether *Acidithrix ferrooxidans* strain Py-F3 is indeed capable of fixing CO_2 has, however, so far not been experimentally clarified.

Comparative analysis of the genome of acidophilic sulfate reducing bacteria

A second environmetal problem resulting from pyrite oxidation is, apart from the liberation of iron (and other metals), the oxidation of sulfide leading to sulfate concentrations in AMD waters (e.g. in the range of 1,200 mg/L in AMD from the Lausitz mining district). These sulfate levels are several folds above legal limits (250 mg/L) set by national and EU-wide

laws. Chemical precipitation of sulfate as gypsum ($CaSO_4$) through the addition of calcium carbonate appears to be the most reliable approach among the physical and chemical methods to reduce the sulfate load in AMD waters (Klein *et al.* 2014). However, while this approach results in the neutralisation of the acidity and, hence, precipitation of dissolved ferrous iron as ferric iron hydroxide, it does not lead to any sulfate removal from AMD waters in the lignite open-pit mine Nochten (Lusatia, Germany), let alone to meeting legal standards, due to the high solubility of gypsum (ca. 1,200 – 1,500 mg/L depending on pH and ionic composition). The most widely applied approach to resolving this problem utilises a so-called water management; that is, AMD water purified from iron *via* calcium carbonate addition is temporally stored within reservoirs from where it is discharged into various surface water ways so that the dilution compensates for the high sulfate concentrations. However, problems may result from reservoir or pit lake overflows or sulfate accumulation further downstream of water ways into which sulfate loaden water has been discharged.

A potential solution to reduce sulfate concentrations is provided by a process that utilises the dissimilatory sulfate reduction of microorganisms and, thus, results in chemical transformation. The removal of sulfate achieved by this biological and sustainable means depends on a variety of parameters that have been discussed and summarised in detail in a recent review co-written by the Author (Klein *et al.* 2014).

In this context, the Author has also been interested in studying acidophilic sulfate reducing prokaryotes (SRP) since these should enable the application of microbial dissimilatory sulfate reduction to the treatment of AMD (pH 3) that has passed the above mentioned pilot plant Tzschelln for the removal of much of the dissolved ferrous iron. Again, clonal and axenic cultures of three acidophilic sulfate reducing bacteria (SRB) were provided by Prof. Johnson (Bangor University): *Desulfosporosinus acididurans* strain M1, '*Candidatus* Desulfobacillus acidavidus' strain CL4 and *Peptococcaceae* bacterium strain CEB3. All these strains belong to the bacterial phylum *Firmicutes* and are mesophilic, anaerobic, motile and rod-shaped bacteria that form endospores. A particular aim of the genome analysis of the genome sequences of the three strains was to reveal possible metabolic features that enable these strains to live at low pH, thus supplementing the results from phenotypic analyses undertaken by Prof. Johnson. The results from the genome sequences of *Desulfosporosinus acididurans* strain M1, and *Peptococcaceae* bacterium strain CEB3 are summarised in two genome announcements (Petzsch *et al.* 2015a, b), while that of '*Candidatus* Desulfobacillus

acidavidus' strain CL4 is planned to be included in a publication by Prof. Johnson describing the proposed new species.

Although there are only two other genome sequences from acidophilic representatives (*Desulfosporosinus* sp. OT: Abicht *et al.* 2011; *Desulfosporosinus acidiphilus* DSM22704^T: Pester *et al.* 2012) among the approximately 80 partially or completely sequenced genomes of SRB that are freely available in online databases, the comparative analyses of the genomes of acidophilic SRB and those of neutrophilic and alkalophilic species may still contribute greatly to our understanding of the genetic basis for the physiological and biochemical adaptation to the acidic environment. This notion is substantiated by the whole genome comparison of the now five available genomes of acidophilic SRB with the genomes of five close neutrophilic relatives.

The results from this comparative analysis indicate that a set of 33 genes appear to be specific to the acidophilic strains (Petzsch *et al.* unpublished). The main fraction of these genes code for transporters (12 genes) with a possible role in ion homeostasis and the transport of heavy metals. Moreover, the comparison of the protein sequence of one of these transporters, a NhaC-like proton-sodium ion antiporter, to genomes of all sequenced SRB indicates a clear correlation to the acidophilic life style. Experiments have been undertaken to confirm these results. The approach taken involves both an overall analysis of the cellular response to acidic conditions using techniques in transcriptomics (RNA-seq) and a quantitative expression analysis targeted at the gene encoding the NhaC-like antiporter using quantitative reverse transcription PCR (RT-qPCR: Joffroy 2014). However, due to the difficult cultivation of acidophilic SRB the results proved so far not reproducible and require renewed attempts.

Interestingly, the genomes of all three acidophilic sulfate reducers also encode a cyanophycin synthetase and a cyanophycinase for the synthesis and degradation of cyanophycin, respectively. Both genes are also encoded in the genomes of neutrophilic representatives of sulfate reducing bacteria (e.g. *Desulfosporosinus youngiae*) where they may provide an interim storage polymer for assimilated nitrogen. In acidophilic sulfate reducers cyanophycin may, in addition to offering a mode for nitrogen storage, contribute to heavy metal resistance (Wood, 1983) and – a hypothesis further discussed in section 3.3 – provide a means to buffer high intracellular proton concentrations that result from life at low pH.

2.4.3 Genome analysis and characterisation of *Petrotoga* strains isolated from gas producing wells

[The results summarised in this chapter have not yet been written up for publication. The experimental work and the taxonomic classification of *Petrotoga* reported here has been carried out by MSc André Pollender as part of a postgraduate research project (Pollender 2014) and his Master thesis (Pollender 2015). The classical DNA-DNA hybridisation has originally been organised and managed by Dipl. Ang.-Nat. Claudia Gniese, but lateron by the Author. Genome and transcriptome sequencing together with the automated assembly and/or annotation of the sequence reads have been undertaken at the G2L (Goettingen University: Dr. A. Poehlein). The Author supervised A. Pollender and coordinated the overall project described in the following.]

Another example of the value of genome sequence analysis shall be provided in the context of research into the microbiology of the deep underground that has been undertaken at the Institute of Biological Sciences. This research had been embarked on with the aim to evaluate the potential impact of carbon capture and sequenstration (CCS) on the indigenous microbial life in subterrestrial natural gas reservoirs. It resulted – among other findings – in the isolation of two strains (isolate Sch_Z2_3, Sch_Z3_22) of the genus *Petrotoga* (Ehinger 2011). These are therefore the first strains of *Petrotoga* that were not isolated from subterrestrial petroleum reservoirs.

The genus *Petrotoga* is known to consist of anaerobic thermophilic microorganisms (Davey *et al.* 1993). Fermentation of various polysaccharides in the absence or presence of elemental sulfur as electron acceptor appears to be another common feature of the so far six described *Petrotoga* species, while only some of these species are able to respire thiosulfate (see Ehinger 2011 for an overview).

Phenotypic characterisation of the two isolates from natural gas reservoirs (isolate Sch_Z2_3, Sch_Z3_22) showed that growth of both is inhibited by high hydrogen concentrations (80% H_2, 20% CO_2) in the atmosphere (Ehinger 2011; Hofmann 2012). Addition of sulfur or thiosulfate was shown to releave this inhibition in strain Sch_Z2_3 (only sulfur in the case of strain Sch_Z3_22) leading to sulfide production and recovery of growth (Ehinger 2011; Hofmann 2012). A similar scenario had already previously been described for strains of *Petrotoga*, *Thermotoga*, *Thermoanaerobacter* and for the archaeon *Pyrococcus furiosus* and

is thought to be caused by the inhibition of the hydrogen evolving hydrogenase (Adams 1990; Fardeau *et al.* 1993; Fiala and Stetter 1986; Lien *et al.* 1998). However, the fact that addition of hydrogen to the gas phase of cultures in media supplemented with an appropriate electron acceptor did not result in better growth as compared to cultures in the same medium but growing under a nitrogen atmosphere, led to the notion that the reaction involving electron acceptor reduction represents (mainly) a detoxification mechanism rather than an energy-yielding process (Fiala and Stetter 1986; Malik *et al.* 1988; Adams 1990). In contrast to this, strains Sch_Z2_3 and Sch_Z3_22 proved to grow faster in heterotrophic media supplemented with an electron acceptor than in the same medium but without inorganic electron acceptors (Ehinger, 2011; Hofmann, 2012), thus indicating energy conservation from respiration. However, there is currently little known about the mechanisms involved in the alleviation of hydrogenase inhibition by sulfur and thiosulfate utilisation in any strain of the *Thermotogales*. Moreover, the mechanisms for sulfur and thiosulfate respiration and concurrent energy conservation is completely unresolved.

Aims

We therefore used genomic and transcriptomic approaches to reveal the genetic basis for sulfur and thiosulfate utilisation in strain Sch_Z2_3. The genome information was also planned to be scrutinised in order to clarify the taxonomic identity of the novel strains. Therefore, the genome of strains Sch_Z2_3 and Sch_Z3_22 together with those of the type strains of *P. miotherma* (Davey *et al.* 1993) and *P. mexicana* (Miranda-Tello *et al.* 2004) were also sequenced since these were recognised in previous studies to be the phylogenetically most closely related taxa (Erler 2009; Hofmann 2012).

Genome sequencing and assembly

Genome sequencing, assembly and automated annotation was again undertaken at the Göttingen Genomics Laboratory (G2L) and lead by Dr Anja Poehlein and Prof. Rolf Daniel. A complete genome sequence was obtained for strain Sch_Z2_3 employing a hybrid approach using the 454 GS-FLX system (Titanium GS70 chemistry; Roche Life Science) and the Genome Analyzer II (Illumina) with gap closure of the resulting contigs being achieved by Sanger sequencing of PCR amplified fragments. In contrast to this, only Illumina sequencing was applied in the other cases and resulted in draft genome sequences consisting of 86 to 280 contigs. Key genome properties are summarised in Table 4.

Table 4: Key genome features of the four *Petrotoga* strains sequenced in collaboration with the Göttingen Genomics Lab (G2L) and of *P. mobilis* (Lien *et al.* 1998) strain SJ95 which was sequenced in 2008 by the Joint Genome Institute (Department of Energy, USA). Data were retrieved from IMG/ER.

Eigenschaft	**Sch_Z2_3**		**Sch_Z3_22**		***P. miotherma***		***P. mexicana***		***P. mobilis SJ95***	
	value	**%-age**	**value**	**%-age**	**value**	**%-age**	**value**	**%-age**	**value**	**%-age**
Genome size (bp)	2,172,470	100%	2,611,526	100%	2,135,534	100%	2,142,506	100%	2,169,548	100%
Number of coding bases (bp)	1,982,554	91.3%	2,371,986	90.8%	1,928,422	90.3%	1,938,609	90.5%	1,969,308	90.8%
G+C content (bp)	738,973	34.0%	888,599	34.0%	727,408	34.1%	730,582	34.1%	740,319	34.1%
Number of contigs	1	100%	280	100%	152	100%	86	100%	1	100%
Total number of genes	2,063	100%	2,626	100%	2,075	100%	2,048	100%	2,012	100%
Protein coding genes	2,004	97.1%	2,539	96.7%	2,013	97.0%	1,982	96.8%	1,955	97.2%
RNA genes	59	2.9%	87	3.3%	62	3.0%	66	3.2%	57	2.8%
Protein coding genes…										
with function prediction	1,710	82.9%	2,085	79.4%	1,689	81.4%	1,703	83.2%	1,343	66.8%
with KEGG pathway	737	35.7%	835	31.8%	716	34.5%	736	35.9%	715	35.5%
with COG	1,705	82.7%	2,104	80.1%	1,682	81.1%	1,697	82.9%	1,591	79.1%
with KOG	754	36.6%	903	34.4%	751	36.2%	757	37.0%	741	36.8%
with Pfam	1,713	83.0%	2,086	79.4%	1,691	81.5%	1,703	83.2%	1,672	83.1%

Taxonomic characterisation

Phenotypic characterisation revealed evidence that strains Sch_Z2_3 and Sch_Z3_22 represent a novel putative species within the genus *Petrotoga* (Ehinger 2011). In particular, the strains differed, apart from the atypical environmental niche they were isolated from, also in their optimal growth temperature and sodium cloride concentration and in their inability to use xylan as organic substrate (Ehinger 2011). A two-tiered approach was employed to test the hypothesis that strains Sch_Z2_3 and Sch_Z3_22 indeed represent a novel putative species within the genus *Petrotoga* or, alternatively, to resolve their correct taxonomic assignment.

Firstly, the genomic information obtained from the genome sequencing of strains Sch_Z2_3 and Sch_Z3_22 and of *P. miotherma* and *P. mexicana* together with the genome data of *P. mobilis* available in the online database were interrogated using a MLSA approach similar to the approach discussed in section 2.2. Additionally, a recently developed bioinformatic approach that uses genome data to estimate the results from DNA-DNA hybridisation by calculating genome-to-genome distances (Meier-Kolthoff *et al.* 2013) was also employed to obtain supporting evidence for the assignment of strains Sch_Z2_3 and Sch_Z3_22 to particular species.

The results were further tested by DNA-DNA hybridisation (carried out at the DSMZ, Braunschweig).

16S rRNA based phylogenetic analysis of strains Sch_Z2_3 and Sch_Z3_22 and the six currently described *Petrotoga* species together with representatives from closely related genera of the *Thermotogales* indicates (Fig. 12) that strain Sch_Z3_22 and possibly also strain Sch_Z2_3 belongs to the species *P. mexicana*, a proposal that has already previously been put forward (Ehinger 2011). However, other 16S rRNA based phylogenetic trees indicate that strain Sch_Z2_3 may belong to *P. miotherma* (Hofmann 2012). The reason for this ambiguous result is most likely found in the high sequence similarity of the 16S rRNA genes among members of the genus *Petrotoga* (Table 5). So, it was not surprsing that a repeat of these analysis using each of the three algorithms (Neighbor-Joining, Maximum Likelihood, Maximum Parsimony) widely accepted by mainstream taxonomists for the calculation of phylogenetic relationship confirms again the close relation of strain Sch_Z3_22 to

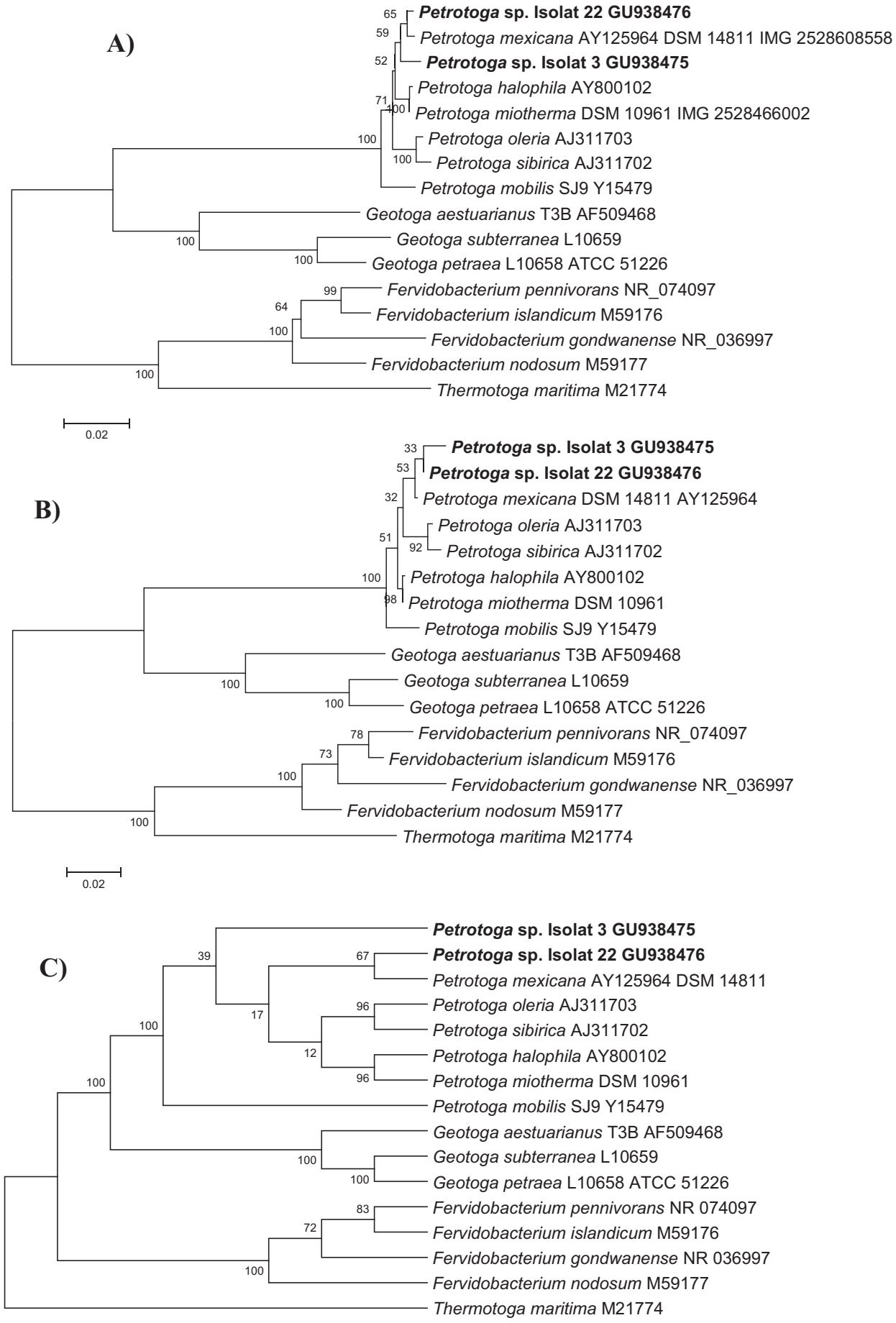
P. mexicana, while *Petrotoga* strain Sch_Z2_3 may represent either *P. mexicana* or belong to a novel putative species (Fig. 12).

Table 5: Results of the Blast searches with the 16S rRNA gene sequences of isolates Sch_Z2_3 and Sch_Z3_22 against the NCBI nr database. Shown are the closest hits to formally described species. Adapted from Hofmann (2012).

Isolate	best BLAST hits	Sequence similarity [1)]	
Sch_Z2_3	*Petrotoga miotherma*	1224/1236	(99,0%)
	Petrotoga halophila	1222/1236	(98,8%)
	Petrotoga mexicana	1218/1236	(98,5%)
Sch_Z3_22	*Petrotoga mexicana*	1220/1231	(99,1%)
	Petrotoga halophila	1216/1231	(98,8%)
	Petrotoga miotherma	1214/1231	(98,6%)

[1)] Sequence similarity is expressed as the number of identical positions / number of bases used for BLAST search and, in brackets, the percentage of identical positions.

Fig. 12: Phylogenetic trees based on 16S rRNA gene fragments (next page). Dendrograms were calculated using the Neighbor-Joining (A), the Maximum Likelihood (B) or the Maximum Parsimony (C) approach. The calculations were performed with MEGA 5 (Tamura *et al.* 2011) based on an alignment produced within ARB (Ludwig *et al.* 2004). Numbers on branches indicate results from bootstrap analyses (percentage of a total of 1000 trees). Isolates Sch_Z2_3 and Sch_Z3_22 are indicated in bold.


A)
Petrotoga sp. Isolat 22 GU938476
Petrotoga mexicana AY125964 DSM 14811 IMG 2528608558
Petrotoga sp. Isolat 3 GU938475
Petrotoga halophila AY800102
Petrotoga miotherma DSM 10961 IMG 2528466002
Petrotoga oleria AJ311703
Petrotoga sibirica AJ311702
Petrotoga mobilis SJ9 Y15479
Geotoga aestuarianus T3B AF509468
Geotoga subterranea L10659
Geotoga petraea L10658 ATCC 51226
Fervidobacterium pennivorans NR_074097
Fervidobacterium islandicum M59176
Fervidobacterium gondwanense NR_036997
Fervidobacterium nodosum M59177
Thermotoga maritima M21774
0.02
B)
Petrotoga sp. Isolat 3 GU938475
Petrotoga sp. Isolat 22 GU938476
Petrotoga mexicana DSM 14811 AY125964
Petrotoga oleria AJ311703
Petrotoga sibirica AJ311702
Petrotoga halophila AY800102
Petrotoga miotherma DSM 10961
Petrotoga mobilis SJ9 Y15479
Geotoga aestuarianus T3B AF509468
Geotoga subterranea L10659
Geotoga petraea L10658 ATCC 51226
Fervidobacterium pennivorans NR_074097
Fervidobacterium islandicum M59176
Fervidobacterium gondwanense NR_036997
Fervidobacterium nodosum M59177
Thermotoga maritima M21774
0.02
C)
Petrotoga sp. Isolat 3 GU938475
Petrotoga sp. Isolat 22 GU938476
Petrotoga mexicana AY125964 DSM 14811
Petrotoga oleria AJ311703
Petrotoga sibirica AJ311702
Petrotoga halophila AY800102
Petrotoga miotherma DSM 10961
Petrotoga mobilis SJ9 Y15479
Geotoga aestuarianus T3B AF509468
Geotoga subterranea L10659
Geotoga petraea L10658 ATCC 51226
Fervidobacterium pennivorans NR 074097
Fervidobacterium islandicum M59176
Fervidobacterium gondwanense NR 036997
Fervidobacterium nodosum M59177
Thermotoga maritima M21774

An MLSA approach was therefore developed in order to overcome the limiting resolution of the 16S rRNA marker. Using the online tool EDGAR (Blom *et al.* 2009) the core genome of the available five *Petrotoga* genomes and that of *Thermotoga maritima* (used as outgroup) was determined. Following an alignment of the 725 core genes the phylogenetic relationship was calculated using again each of the three above mentioned methods: Neighbor-Joining, Maximum Likelihood, Maximum Parsimony. The results indicate that strains Sch_Z2_3 and Sch_Z3_22 are more closely related to *P. miotherma* than to *P. mexicana* (Fig. 13). Trees based on the core genome (1513 genes) of only the five *Petrotoga* genomes resulted in identical branching, though with higher bootstrap values (data not shown).

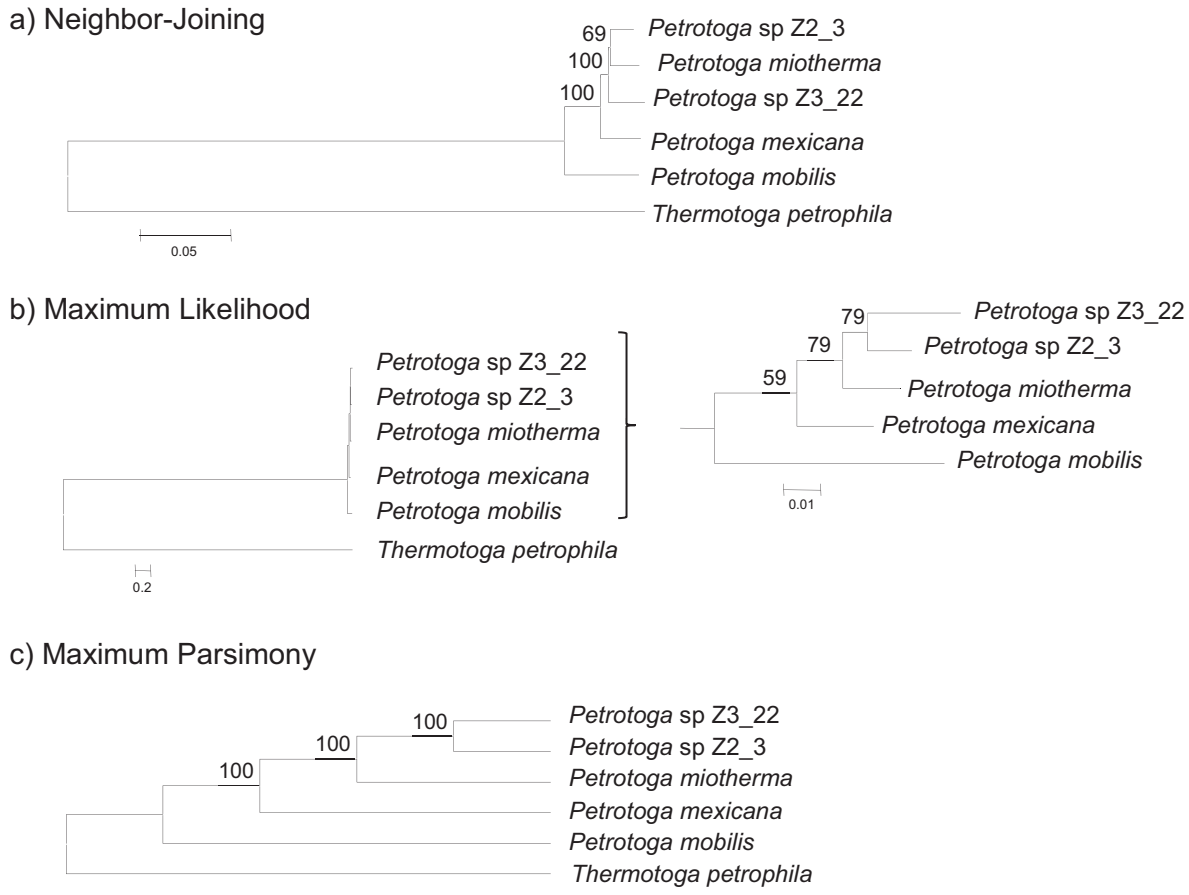


Fig. 13: Phylogenetic relationship of strains Sch_Z2_3 and Sch_Z3_22 within the genus *Petrotoga*. The calculations were based on an alignment of the core genome of 725 genes present in all of the strains. Bootstrap values indicate the percentage of identical branching (out of 1000 replicates). The neighbor-joining and the maximum parsimony dendrograms were calculated using MEGA 5 (Tamura *et al.* 2011) while the maximum likelihood dendrogram was computed with PHYML 3.0 (Guindon *et al.* 2009).

Furthermore, the hypothesis was tested that calculation of the phylogeny based on a small subset of the core genome results in a similar relationship to that obtained based on the core genome (Ivars- Martínez *et al.* 2008; Rokas *et al.* 2003: see section 2.2). The MLSA analysis was repeated with a subset consisting of the 16S rRNA gene and ten house-keeping genes (*recA*, *dnaK*, *glyA*, *pmg*, *metG*, *tufA*, *gyrB*, NAD synthetase, *polA*, *rpoB*), again using each of the three methods for estimating phylogenetic relationship. The dendrograms obtained from these analyses (Fig. 14) proved to be identical to those based on the core of 725 genes from the five available *Petrotoga* genomes and that of *T. maritima* (Fig. 13).

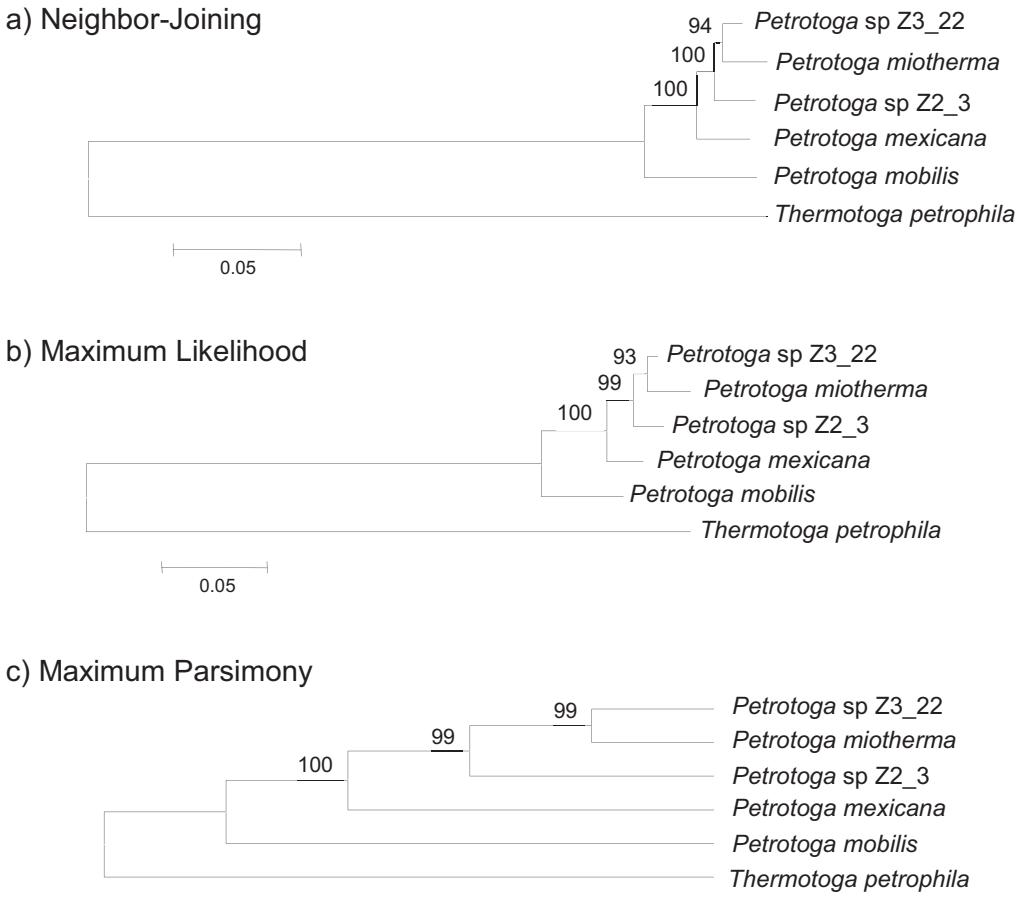


Fig. 14: Phylogenetic relationship of strains Sch_Z2_3 and Sch_Z3_22 within the genus *Petrotoga*. The calculations were based on an alignment of eleven marker genes present in all of the strains. Bootstrap values indicate the percentage of identical branching (out of 1000 replicates). The trees were calculated using MEGA 5 (Tamura *et al.* 2011).

We also employed the recently developed bioinformatic approach for the estimation of the results from DNA-DNA hybridisation based on genome data (Meier-Kolthoff *et al.* 2013). This method calculates genome to genome distances using various algorithms designed to cope with both complete and draft genome data (i.e. consisting of various numbers of contigs). The results from this *in silico* analysis confirmed the assignment of strains Sch_Z2_3 to *P. miotherma* (Table 6) which was also suggested by the results obtained by the MLSA approaches (Fig. 13, 14). In contrast to this, the taxonomic position of strain Sch_Z3_22 was less clear since the DNA-DNA hybridisation values were close to but below the threshold for taxonomic assignment to *P. miotherma* (Table 6).

Table 6: Results from the calculation of the DNA-DNA hybridisation values using the genome to genome distance calculator (GGDC 2.0) (Meier-Kolthoff *et al.* 2013). Shown are the results for various algorithms that were employed together with formula 2 which is recommended for draft genomes. The same results shown here were also obtained for the comparison of the closed genome sequences of strain Sch_Z2_3 and *P. mobilis* using formula 1 recommended for closed genome sequences (Meier-Kolthoff *et al.* 2013). Values >70 indicate that the strains belong to the same species (i.e. correspond to the 70% DNA-DNA hybridisation value).

			GBDP2_NCBI-BLAST	**GBDP2_ BLASTPLUS**	**GBDP2_ BLAT**	**GBDP2_ WU-BLAST**
	vs	Sch_Z3_22	73.5	100	77.9	100
Sch_Z2_3	*vs*	*P. miotherma*	72.0	72.6	72.2	17.0
	vs	*P. mexicana*	57.5	59.0	59.2	63.6
	vs	*P. mobilis*	43.6	45.9	46.1	54.3
	vs	*P. miotherma*	68.8	69.6	69	67.7
Sch_Z3_22	*vs*	*P. mexicana*	57.6	59.2	58.9	63.3
	vs	*P. mobilis*	45.8	48	47.8	55.5

Following successful cultivation to obtain biomass (3 g wet weight) sufficient for DNA-DNA hybridisation we went on to collect the experimental evidence that is required by the formal rules for the description of novel microbial species (Wayne *et al.* 1987). DNA-DNA hybridisation was carried out at the DSMZ (Braunschweig) and provided the final support for the assignment of strains Sch_Z2_3 and Sch_Z3_22 to *P. miotherma* (Table 7). Therefore, the results from the DNA-DNA hybridisation clarify the results obtained from the calculations based on the genome to genome distance calculator (GGDC 2.0). However, additional DNA-DNA hybridisation provides experimental evidence that *P. miotherma* and *P. halophila* (Miranda-Tello *et al.* 2007) are the same species, that is *P. halophila* is a synonym of *P. miotherma* which has been described first (Davey *et al.* 1993). Given the significance of the finding these tests are currently repeated (C. Spröer, DSMZ – Braunschweig).

Table 7: Results from the DNA-DNA hybridisation based on the recommendations of Wayne *et al.* (1987). DNA-DNA hybridisation was performed at the DSMZ, Braunschweig.

Strain/species	Sch_Z2_3	Sch_Z3_22	*P. miotherma*
Sch_Z2_3		84.7 (78.8)	
P. miotherma	78.8 (79.5)	86.9 (78.9)	
P. mexicana		41.8 (45.4)	44.5 (40.6)
P. halophila	88.1 (80.1)	78.9 (75.4)	88.7 (96.0)

Analysis of the mechanisms for sulfur and thiosulfate respiration

Reconstruction of the energy metabolism from the complete genome sequence (2.17 Mbp) indicated that cytosolic hydrogenases (HydABC) may be responsible for the regeneration of reducing equivalents in the absence of alternative electron acceptors (Fig. 15). This process results in the formation of H_2.

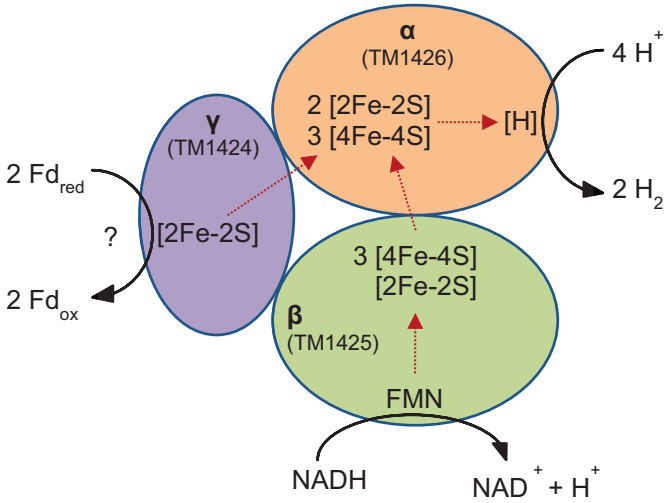


Fig. 15: Model of the hydrogenase HydABC in *Petrotoga* strain Sch_Z2_3 based on that suggested for *T. maritima* (Schut and Adams 2009). HydABC in *T. maritima* consists of 32 iron atoms in ten Fe-S clusters. Binding sites for FMN and NAD^+ are indicated, again as suggested by Schut and Adams (2009). Red arrows suggest the most likely path for electron transfer. Taken from Pollender 2015.

None of the five available *Petrotoga* genomes encodes cytochromes. Instead, the presence of an archaeal type gene cluster consisting of a membrane bound hydrogenase and an ion (e.g. proton or sodium ion) export function (Fig. 16) provides evidence for a situation similar to that known to be involved in sulfur respiration in the archaeon *Pyrococcus furiosus*, that is the coupling of proton reduction to proton translocation for energy conservation (Sapra *et al.* 2003; Schut *et al.* 2013). If correct, then this metabolic feature should result in increased growth rates in *Petrotoga* spp. in the presence of sulfur. That is, the increased proton export is subsequently used for additional ATP formation by ATP synthetases, of which *Petrotoga* sp. Sch_Z2_3 encodes two copies, one of the F0F1-type and an archaeal type ATP synthetase. Given the fact that lateral gene transfer between archaea and *T. maritima* has been observed (Nelson *et al.* 1999), this result only comes as little surprise.

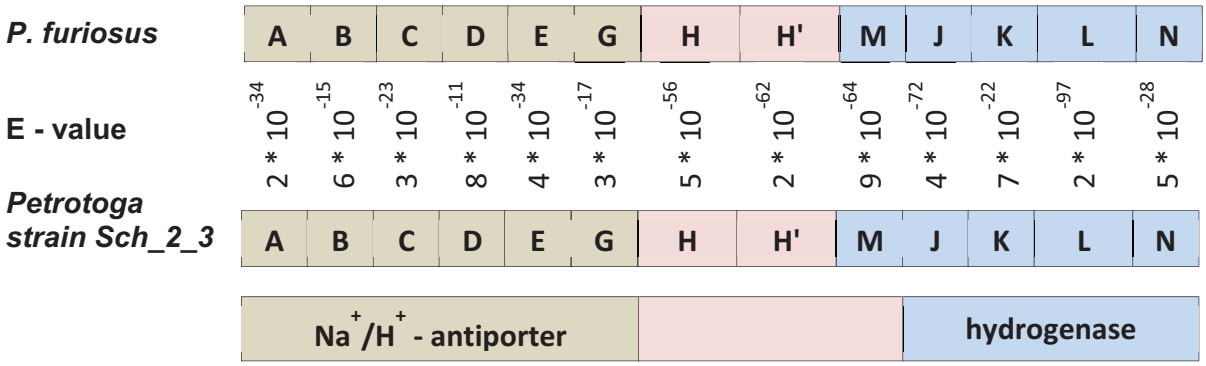


Fig. 16: Results from a BLAST based comparison of the individual genes of the Mbx gene cluster from *Pyrococcus furiosus* with those of the Mbx gene cluster in the genome of *Petrotoga* strain Sch_Z2_3. The pink coloured proteins have homology to both, antiporter and hydrogenase modules. Subunit L harbours the active site. (Taken from Pollender 2015.)

Apart from the sequence based similarity of the two gene clusters this hypothesis is further supported by the similarity of the predicted transmembrane helices within the individual proteins of the Mbx complex (Table 8).

Table 8: Transmembrane helices (TMH) in subunit proteins of the Mbx complexes in *Pyrococcus furiosus* and *Petrotoga* sp. strain Sch_Z2_3. TMH were determined using the TransMembrane Hidden Markov Model (TMHMM) prediction tool (Sonnhammer *et al.* 1998: http://www.cbs.dtu.dk/services/TMHMM/). Taken from Pollender (2015).

	P. furiosus			***Petrotoga* sp. strain Sch_Z2_3**	
Gene	**Locus tag**	**Locus tag**	**TMH**	**Locus tag**	**TMH**
mbxA	PF1453	PET_00985	3	PET_00985	3
mbxB	PF1452	PET_00986	3	PET_00986	3
mbxC	PF1451	PET_00987	3	PET_00987	3
mbxD	PF1450	PET_00988	3	PET_00988	3
mbxF	PF1449	PET_00989	6 (2+4)[1]	PET_00989	6
mbxG	PF1448	PET_00990	3	PET_00990	3
mbxH	PF1447	PET_00991	14	PET_00991	14
mbxH'	PF1446	PET_00992	14	PET_00992	16
mbxM	PF1445	PET_00993	8	PET_00993	7
mbxJ	PF1444	PET_00994	-	PET_00994	-
mbxK	PF1443	PET_00995	-	PET_00995	-
mbxL	PF1442	PET_00996	1	PET_00996	-
mbxN	PF1441	PET_00997	-	PET_00997	-

[1] Gene *mbxF* of *P. furiosus* encodes for two subunits: MbhE and MbhF. The sum of the TMH for the individual subunits equals that of PET_00989.

A transcriptome experiment was designed to test the involvement of this archaeal type mechanism in sulfur respiration and to also reveal the mechanism responsible for thiosulfate reduction. The approach taken was based on RNA-Seq analysis of batch cultures of *Petrotoga* strain Sch_Z2_3 incubated in maltose containing medium (DSMZ 718, with the following modifications: 0.5 g/L yeast, 0.5 g/L $NaHCO_3$) with or without sulfur or thiosulfate. Nine individual 50-mL cultures in serum bottles were grown for 48 hours at 60 °C in maltose-containing medium and under a CO_2/H_2 atmosphere (80%/20%) to reach exponential growth (Fig. 17A). The individual cultures were pooled and mixed. 40-mL of the pooled culture were transferred into each of nine serum bottles. These were supplemented with either 10 mL of fresh maltose containing medium (3 cultures), or maltose containing medium plus sulfur (final concentration: 1% (w/v); again three cultures), or maltose containing medium plus thiosulfate (final concentration: 20 mM; again three cultures). One culture of each of these three

experimental conditions was treated with phenolic stop solution (5 mL of a phenol-ethanol solution: 1:10) after 1 h and a second culture after 4 h (Fig. 17A). Additionally, the optical density (OD) was determined from an aliquot of these cultures to estimate their growth state. Cells were harvested by centrifugation and lysed by physical treatment (freeze-thawing: Pollender 2015). Total RNA extraction was performed using the RNeasy Kit (Qiagen). The third culture was used for a further measurements of cell growth (OD_{580}). This demonstrated that the test cultures were still in exponential growth following the 1-h and 4-h incubations (Fig. 17B, C).

Ribosomal RNA depletion and subsequent RNA-Seq analysis was performed by the G2L (Dr. A. Poehlein). Gene expression was quantified from RNA-Seq data by normalisation of the raw data as **N**ucleotide activity **P**er **K**ilobase of exon model per **M**illion mapped reads (NPKM; Wiegand *et al.* 2013). NPKM are functionally equivalent to the hitherto used RPKM (**R**eads **P**er **K**ilobase per **M**illion mapped reads; Mortazavi *et al.* 2008), though thought to be more accurate due to their single base resolution (Wiegand *et al.* 2013).

The transcriptome experiments were conducted without replication. [This is explained by the fact, that as part of the project CO2BioPerm, transcriptome experiments were to be conducted on cultures exposed to various pH levels exclusively adjusted by appropriate concentrations of CO_2. Since the provision of those culture by the project partner responsible for them were continuously delayed, it was decided to carry out a preliminary (hence no replicates) experiment to evaluate the involvement of the Mbx complex in sulfur and thiosulfate respiration.] Therefore, statistical evaluation as to the significance of differences in the expression levels of individual genes can not be made. Nevertheless, the data obtained from this experiment suffice as preliminary experimental evidence for the hypothesis that the Mbx complex is involved in sulfur respiration. That is, sulfur addition to the growth culture leads to the induction of the Mbx gene cluster which results, on average, in a 7.6-fold (1 h) and 22.6-fold (4h) increase of NPKM for the 13 genes (Table 9).

In contrast to this, no clear indication was obtained regarding the molecular answer of *Petrotoga* sp. Sch_Z2_3 to the availability of thiosulfate. That is, no obvious upregulation of genes possibly involved in thiosulfate respiration was observed. Interpreting modest changes in gene expression run the risk of being misleading due to the lack of replicates. The absence of a clear cellular response is surprising given that the cultures with added thiosulfate grew faster than those containing only maltose (Fig. 17C).

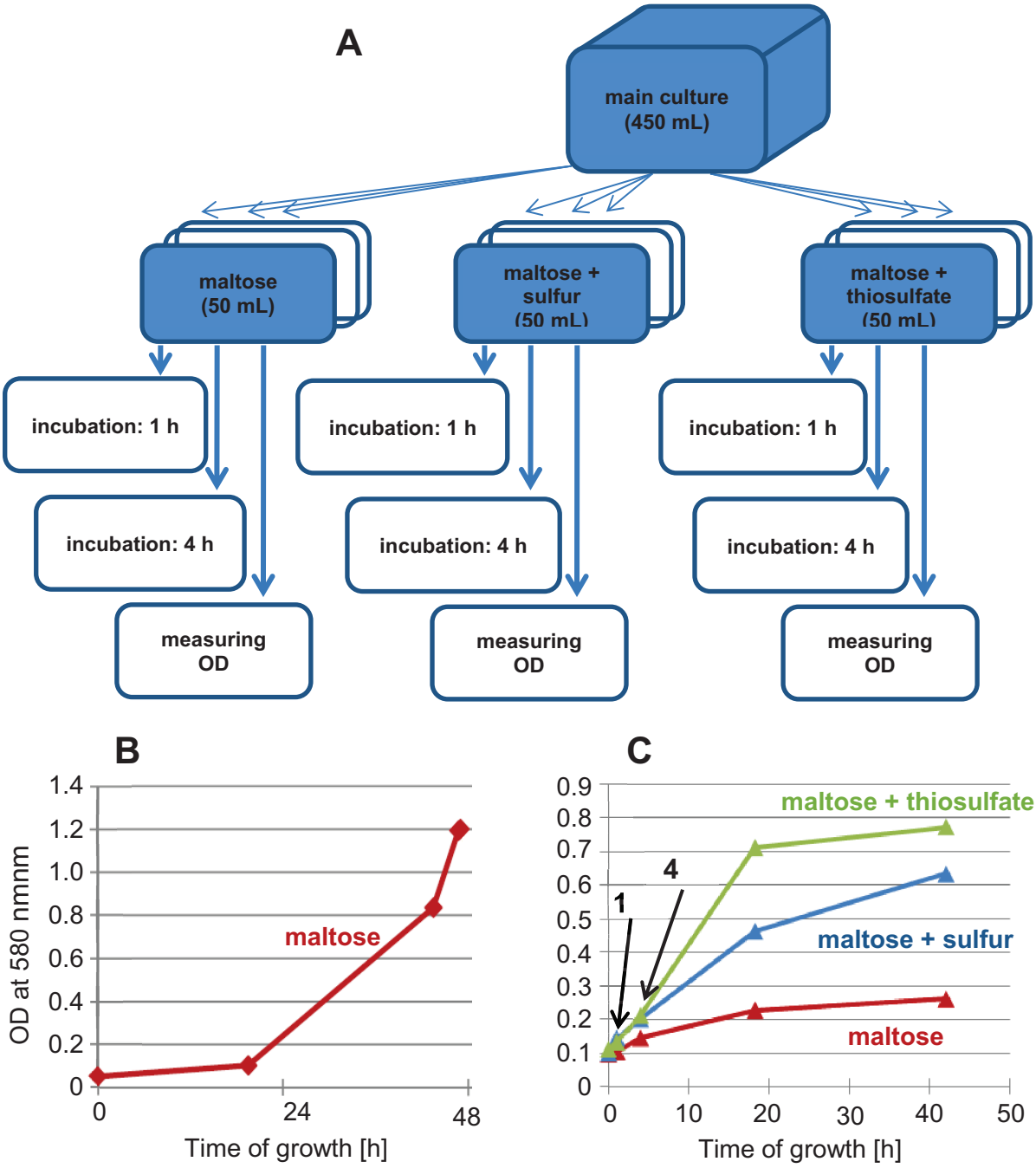


Fig. 17: Overview of the transcriptome experiment (A) and results from a growth curve analysis of the pre-culture (B) which was used after 48 hour growth as inoculum for the cultures analysed by RNA-Seq (C). Growth curve analysis of the those cultures confirmed that the cells were still in the exponential growth phase after 1h and 4h incubation in the presence or absence of an electron acceptor (S^o or thiosulfate). Setting up of the individual incubations for the experiment required the use of an anaerobic chamber. This experimental step therefore entailed approx. 45 min of exposure to room temperature prior to the start of the incubations at 60 °C for 1 h and 4 h. (Based on Pollender 2015.)

Table 9: Gene expression of the Mbx complex depending on the availability of electron acceptors. Gene expression is quantified as normalised sequence reads (NPKM). M - 10 mM maltose; T - 10 mM maltose plus 20 mM thiosulfate; S - 10 mM maltose plus 1 %(w/v) sulfur. Exposure times were 1h and 4h.

Locus-tag	**Gene**	**Gene expression (NRPKM)**						**Change in gene expression**	
		M 1h	M 4h	T 1h	T 4h	S 1h	S 4h	S *vs.* M 1h	S *vs.* M 4h
PET_00985	mbxA	1	3	0	4	67	145	67.0	48.3
PET_00986	mbxB	0	1	1	3	49	93	49.0	93.0
PET_00987	mbxC	2	5	0	5	60	165	30.0	33.0
PET_00988	mbxD	1	2	1	2	37	105	37.0	52.5
PET_00989	mbxF	2	1	2	2	44	110	22.0	110.0
PET_00990	mbxG	2	2	2	3	31	84	15.5	42.0
PET_00991	mbxH	4	5	7	7	27	88	6.8	17.6
PET_00992	mbxH’	7	4	12	10	36	88	5.1	22.0
PET_00993	mbxM	8	6	17	12	44	87	5.5	14.5
PET_00994	mbxJ	11	7	20	22	48	112	4.4	16.0
PET_00995	mbxK	12	10	25	23	58	119	4.8	11.9
PET_00996	mbxL	17	8	35	26	67	120	3.9	15.0
PET_00997	mbxN	14	8	24	25	45	93	3.2	11.6
	Average	**6.2**	**4.8**	**11.2**	**11.1**	**47.2**	**108.4**	**7.6**	**22.6**

The Rnf complex of Petrotoga *sp. Sch_Z2_3*

Genome analysis also uncovered an RnfABCDEG complex encoding gene cluster (see Pollender 2015) which, for instance in *Acetobacterium woodii* (Biegel *et al.* 2011), has been found to be involved in a membrane bound ion-motive electron transport chain as an alternative to cytochromes (Biegel *et al.* 2011). The remarkable feature of this complex is the hitherto unique link it provides between the cellular ferredoxin and the pyridine nucleotide

pool: this link extends the redox span of the electron transport chain to a more negative range than the -320 mV catered for by pyridine nucleotides (Kumagai *et al.* 1997; Biegel and Müller 2010; Biegel *et al.* 2011). In essence, electron transfer from reduced ferredoxin to oxidised NAD(P)H within the Rnf complex provides sufficient energy for the export of ions (protons or sodium ions) from the cytoplasm into the periplasm (Biegel *et al.* 2011), thus increasing the proton motive force and, as a consequence, energy production. The precise amount of ions exported for each ferredoxin that is oxidised depends on the prevailing membrane potential, but is cestimated to be in the range of one to two protons (Biegel *et al.* 2011). Furthermore, this electron transport is thought to also run reversibly, driven by ion translocation from the periplasm into the cytoplasm (Biegel *et al.* 2011). Therefore, it offers a mechanism for the recycling of reduced NADH that result from fermentation reactions. The number of protons that are required to be imported by the Rnf complex for the concomitant uphill electron transport from NADH to oxidised ferredoxin would again be dependent on the individual membrane potential, but should, based on the calculation for the ion export function (Biegel *et al.* 2011), not exceed one proton per uphill electron transfer in the reversed transport mode.

Apart from the high sequence similarity to the genes encoding the well characterised Rnf complex in *A. woodii* (Biegel *et al.* 2011), the Rnf complex in *Petrotoga* sp. Sch_Z2_3 also corresponds to that of *A. woodii* in terms of its transmembrane spanning structure (see Pollender 2015). Due to the lack of replicates in the transcriptomic analyses it remains however unsure whether the generally small increase in the expression of the Rnf complex encoding gene cluster (1.7 and 1.3 for thiosulfate and sulfur, respectively, after 4 h incubation) signifies a direct responds to the addition of the electron acceptors. Moreover, other genes likely to encode further transmembrane electron transport complexes or complexes for ion export through the cytoplasm also show low increase in their expression (Pollender 2015). However, the gene encoding ferredoxin was the second most expressed gene in any of the cultures. This is important since it underlines the relevance of ferredoxin as a “high energy” intermediate in electron transfer reactions (Biegel *et al.* 2011).

Based on the results from the genomic and transriptomic analyses a potential mechanism for the reduction of sulfur in *Petrotoga* sp. Sch_Z2_3 was developed (Fig. 18). Although the metabolic path of thiosulfate reduction could not convincingly be explained, genomic evidence is provided that energy conservation can be achieved similarly to the mechanism

described for sulfur dissimilation. Based on the proposed pathways for thiosulfate and sulfur respiration and under the assumption, that the import of one proton is required for the uphill transport of one electron (Biegel *et al.* 2011) sulfur and thiosulfate respiration would still lead to additional energy conservation in *Petrotoga* sp. Sch_Z2_3. Table 10 provides an overview of the overall energy gain from an exclusively fermentative life style and from one that is coupled to sulfur and thiosulfate respiration and compares this to that described for *Pyrococcus furiosus*.

Fig. 18: Energy conservation in *Petrotoga miotherma* strain Sch_2_3. Maltose is metabolised to acetate and CO_2 resulting in reduced NADH and ferredoxin (Fd) which are, in the absence of electron acceptors, recycled by a cytosolic hydrogenase (HydABC; the red arrows indicate the likely path of the electrons: based on Biegel *et al.* 2011). Upon addition of sulfur, the Mbx gene cluster is expressed. The Mbx complex catalyses the transfer of electrons from Fd to sulfur, thus leading to hydrogen sulfide formation. Energy conservation is achieved by concomitant proton export *via* the Mbx cluster and subsequent ATP formation through two types of ATP synthetases (proton import) whose encoding genes are also increasingly expressed upon sulfur and thiosulfate addition. NADH is re-oxidised in an uphill electron transfer to oxidised Fd. This redox reaction is catalysed by the reverse action of the Rnf cluster and driven by the proton gradient (proton import). It is assumed that the import of one proton suffices the energy demand for the uphill transport of one electron from reduced NADH to oxidised Fd (Biegel *et al.* 2011). The mechanism of thiosulfate reduction is less clear, but may involve an uncharacterised ABC transporter (PET_00811-PET_00813), a gene cluster predicted to encode an oxidoreductase (PET_00454-PET_00458), a predicted proton export complex (PET_00649-PET_00654). Additionally, the archaeal type ATP synthetase complex is shown (next to the proton exporter). However, no thiosulfite reductase has so far been detected. (Based on Pollender 2015.)

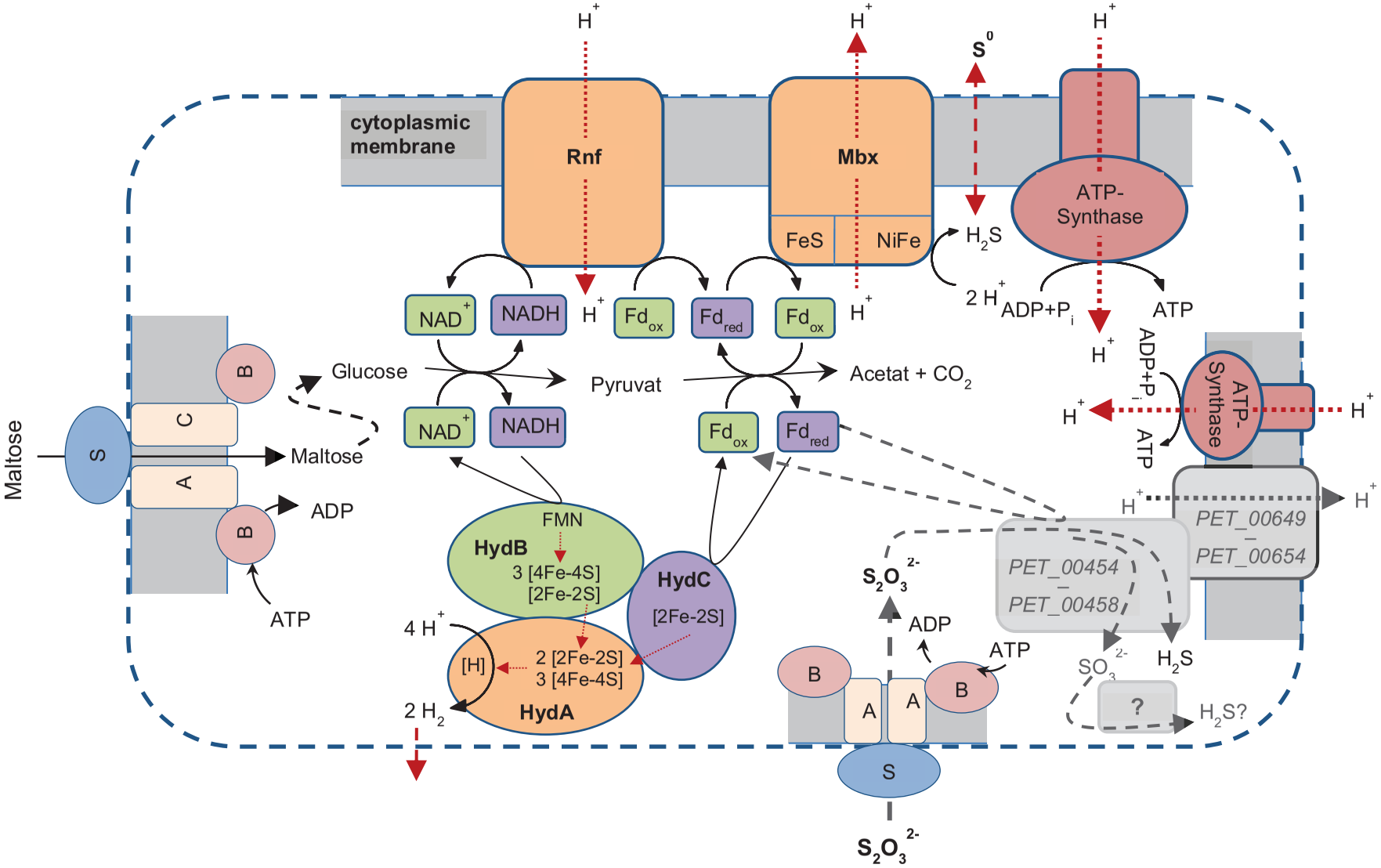

cytoplasmic membrane
Rnf
Mbx
FeS
NiFe
ATP-Synthase
H+
S0
H2S
2 H+
ADP+Pi
ATP
NAD+
NADH
Fdox
Fdred
Glucose
Pyruvat
Acetat + CO2
Maltose
S
C
A
B
ADP
ATP
FMN
HydB
3 [4Fe-4S]
[2Fe-2S]
HydC
[2Fe-2S]
4 H+
[H]
2 [2Fe-2S]
3 [4Fe-4S]
HydA
2 H2
PET_00649
PET_00654
PET_00454
PET_00458
S2O3 2-
SO3 2-
H2S
?
H2S?

Table 10: Comparison of the energy conservation in *P. miotherma* Sch_2_3 and *P. furiosus*. Details are provided for each metabolic step relevant for the energy budget of maltose fermentation which results in reduced NADH and ferredoxin. *P. furiosus* uses exclusively ferredoxin as reduction equivalent: glycerinaldehyd-3-phosphate oxido-reductase of *P. furiosus* catalyses the reactions performed in *P. miotherma* Sch_2_3 by two enzymes, glucose-6-phosphate isomerase and glyceraldehyde-3-phosphate dehydrogenase. $\Delta\mu_{H^+}$ = difference in electrochemical potential through proton translocation; Fd^- = reduced ferredoxin. (Based on Pollender 2015.)

Enzyme	***P. miotherma* Sch_2_3**	***P. furiosus***
Maltose import		
ABC transporter	– 1 ATP	– 1 ATP
Glycolysis		
Glucokinase	– 1 ATP	– 1 ATP
Glucose-6-phosphate isomerase	– 1 ATP	– 1 ATP
Glyceraldehyde-3-phosphate dehydrogenase	+ 2 NADH	
Phosphoglycerate kinase	+ 2 ATP	
Glyceraldehyde-3-phosphate ferredoxin oxidoreductase		+ 4 Fd^-
Pyruvate kinase	+ 2 ATP	+ 2 ATP
Pyruvate oxidation		
Pyruvate ferredoxin oxidoreductase	+ 4 Fd^-	+ 4 Fd^-
Acetate kinase	+ 2 ATP	+ 2 ATP
Regeneration of cofactors		
- without addition of electron acceptors		
HydABC	– 4 Fd^- – 2 NADH	
Mbh complex		– 8 Fd^- + 8 $\Delta\mu_{H^+}$
- with sulfur		
Rnf complex	– 2 NADH + 4 Fd^- – 4 $\Delta\mu_{H^+}$	
Mbx complex	– 8 Fd^- + 8 $\Delta\mu_{H^+}$	– 8 Fd^- + 8 $\Delta\mu_{H^+}$
- with thiosulfate		
uncharacterised membrane complex	min. – 8 Fd^- min. + 8 $\Delta\mu_{H^+}$	
ATP synthase	– 4 $\Delta\mu_{H^+}$ + 1 ATP	– 8 $\Delta\mu_{H^+}$ + 2 ATP
Sum: maltose only	**3 ATP**	**3 ATP**
Sum: maltose + sulfur	**4 ATP**	**3 ATP**
Sum: maltose + thiosulfate	**4 ATP (min.)**	

Outlook on future research

A repeat of this study, but under improved experimental conditions, is planned in the framework of a further project (depending on grant award). Improvements involve continuous culture of *Petrotoga* sp. Sch_Z2_3 in appropriate bioreactors and in replicates. Such an approach would, apart from statistical evaluation of the results, also permit to address the question regarding the time course of the gene expression of the Mbx complex encoding gene cluster and other genes involved in the regulation and function of sulfur respiration. These results could then be further quantified by RT-qPCR. Continuous culture would also permit to circumvent the need to use an anerobe chamber and, thus, avoid a temperature drop of the culture medium caused by the exposure to room temperature for approx. 45-min prior to the 1-h and 4-h exposure to electron acceptors (Fig. 17). It should be mentioned it this context that a cold shock protein encoding gene (PET_00011) showed the highest expression levels in any of six analysed cultures conditions. Although it remains unknown whether the exposure to room temperature prior to the final incubation period with or without electron acceptor addition (Fig. 17A) is responsible for this observation, continuous culture certainly allows for more control in culture conditions than does batch culture in small volumes.

Given the uncertainty regarding the mechanism of thiosulfate respiration the repeat experiment is also planned to be accompanied by physiological experiments (e.g. incubation with isotopically labelled thiosulfate) to support prediction of the fate of thiosulfate during thiosulfate reduction.

3. Culture independent approaches to access the genetic information of microbial communities

It is the Author's view that isolation of microorganisms persists as the most important method in Microbiology (see section 4). However, despite all systematic and creative attempts to improve the likelihood of success in the isolation of novel microorganisms, the rate of isolation and phenotypic characterisation will remain the rate limiting step in furthering our knowledge on the huge diversity of microorganisms in the natural environment. Identification of those species that have not yet been brought into (pure) culture can, however, be obtained by sequence analysis of individual clones from a library of cloned fragments of a particular marker gene, such as the 16S rRNA gene (Woese and Fox 1977), that has been amplified *via* the polymerase chain reaction (PCR; Saiki *et al.* 1985, 1986). This new culture independent approach has been applied for the study of the microbial diversity for much of the last 25 years (starting e.g. with Giovannoni *et al.* 1990; Ward *et al.* 1990; Amann *et al.* 1991; deLong 1992; Fuhrman *et al.* 1992). Although it may appear that it has sometimes been applied to satisfy researchers' particular interest (or paranoia: e.g. the analysis of the microbial community on religious icons that may be a result from the orthodox rite of "kissing of the icons": Lupan *et al.* 2014), it still remains evident that it has revolutionised the research of microbial ecologists and revealed that only the very tip of the iceberg of microbial diversity had been scratched for most environments that were analysed in this way. This approach, however, did not provide any data on possible phenotypic features of the uncultivated microbial fraction. Any hypothesis of their function was limited to existing knowledge of phylogenetically close relatives with representative isolates whose phenotype had been characterised and was, therefore, based on the assumption that phylogeny correlates to phenotype. Further culture independent methods were therefore developed in order to overcome this limitation too. An early approach involved the cloning of large fragments of genomic DNA from total environmental DNA preparations into vectors specifically designed for transformation large inserts and transformation into heterologous hosts. Those vectors that were most widely used are fosmids and cosmids (inserts up to ca. 35 kb in size) and bacterial artifical chromosomes (BACs: inserts up to ca. 150 kb in size). Apart from identifying novel genes that encode enzymes for unique phenotypic traits it was furthermore hoped that large fragments encode also a phylogenetic marker (ideally the 16S rRNA gene) in order to relate

diversity to function. Although the use of BACs led to extremely interesting findings, such as the presence of the proton pump proteorhodopsin in marine bacteria (Beja *et al.* 2000a), their handling proved to be difficult and insert sizes often remained within the range of those that can be cloned into fosmids and cosmids (an experience made by former colleagues of the Author).

The developemet of a new generation of sequencing techniques, in particular 454 pyrosequencing and *via* the Illumina approach, has provided the basis for a further revolution to our understanding of the extent of the microbial diversity. The drastic drop in sequencing costs and efforts while, at the same time, achieving unprecedented sequence output also meant that it was no longer prohibitively expensive to sequence total DNA isolated from an environmental sample. Moreover, this next generation sequencing approach also overcomes problems associated to the cloning of large genomic DNA fragments. These methods that not only analyse the genome of an individual isolate, but aim at the analysis of all genomes within a simple or complex microbial community, are summarised by the term metagenomics (Handelsman *et al.* 1998). The advancement in sequencing technologies also offered the possibility to determine all genes expressed by a single isolate or (a large fraction of those expressed) by whole microbial communities. In analogy to the term metagenomics this approach was named metatranscriptomics.

The Author has utilised both high capacity cloning vectors and next generation sequencing based approaches in his work and also contributed to the improvement of genetic fingerprint techniques.

3.1 Improvements to the culture independent methods for the study of complex microbial assemblages

The application of the culture independent techniques has, beyond doubt, revolutionised the work of microbial ecologists. However, there has been great dispute over the accuracy of the results achieved by genetic fingerprint techniques in terms of both the correct taxonomic assignment of microbial strains and the genetic resolution of the method (e.g. Bent *et al.* 2007).

3.1.1 Evaluating marker genes for their usefulness as phylogenetic marker

Although the occurrence of horizontal gene transfer (HGT) dates back more than 60 years (Tatum and Lederberg 1947; Zinder and Lederberg 1952), it has only more recently become clear that HGT is wide spread among prokaryotes (see Ochmann *et al.* 2000). In the era of microbial genome sequencing the extent to which microbial species are affected by HGT is now being delivered 'on a silver platter' and opens up a range of discussions on pertinent issues, including the question as to what constitutes a 'microbial species' (e.g. Konstantinidis *et al.* 2006).

From a more pragmatic viewpoint, it becomes furthermore evident that HGT also throws a monkey wrench into the works of single-marker based diversity studies. The 16S rRNA gene has so far not really been under suspicion to be horizontally transferred between species. However, apart from the application of marker genes to assess a specific functional diversity, there is also the relatively high sequence conservation and the occurrence of multiple ribosomal RNA operons with sequence variations between them that sometimes necessitates the use of alternative markers to evaluate microbial diversity or to better determine strain identity (see also section 2.2, Ivars-Martinez *et al.* 2008 and Mühling 2012).

Apart from testing whether the selected marker genes are indeed similar in terms of their phylogenetic information, such an assessment should also reveal the extent of genetic resolution that the various marker genes provide, e.g. in relation to the 16S rRNA gene. An understanding of the relative genetic resolution is important for both, the comparison between the results from studies that used different marker genes and for the statistical analysis of correlation with environmental and biotic parameters. For example, Martiny and colleagues (2009) found that the correlation between *Prochlorococcus* diversity and environmental variables and dispersal time was dependent on the scale of sequence identity that was used to group sequences into taxonomic units. Similarly, the Author found different values for the correlation between environmental parameters and *Prochlorococcus* diversity dependent on whether the detected *Prochlorococcus* (*rpoC*) sequences were analysed as individual RFLP-types or grouped into phylogenetic (sub)clades (Jameson *et al.* 2010).

Therefore, a serious assessment of marker genes with available PCR primers for amplification of gene fragments is a prerequisite for successful analyses of the overall microbial diversity or of that of particular phylogenetic groups.

A suggestion on how to perform such an assessment has been outlined by the Author (Mühling 2012) and was applied in a multi locus sequence analysis (MLSA) study of marine *Alteromonas macleodii* (Ivars-Martinez *et al.* 2008). In essence, it involves the statistical comparison of the differences in nucleotide sequences between every pair of strains and for each marker gene. Such an analysis can be performed by calculating distance matrices for each of the marker genes from the individual nucleotide sequence alignments. These individual matrices are then compared using, for instance, a second stage multidimensional scaling (MDS) analysis (Fig. 19); that is, by calculating the rank correlations between every pair of distance matrices (Ivars-Martinez *et al.* 2008; Mühling 2012). Moreover, this approach also permits the direct comparison of e.g. the phylogenetic clustering as obtained from molecular sequence data with the grouping of the same set of strains obtained from e.g. phenotypic analyses since rank correlations do not compare numerical values between distance matrices, but merely the ranks of samples within each of the matrices (Ivars-Martinez *et al.* 2008).

Apart from the phylogenetic message inherent to a particular marker gene its genetic resolution is also of importance for the selection procedure. The probably most simple procedure to assess the genetic resolution of a particular marker gene is performed by calculating the ratio of polymorphic to monomorphic sites, thus also normalising the data against sequence length. In this context and to underline the recurrent need for alternatives to the 16S rRNA gene as molecular marker for diversity studies it should be noted that the 16S rRNA gene proved to be the most conserved one out of ten marker genes in the case of *Prochlorococcus* (Mühling 2012), thus requiring more variable marker genes to differentiate, for instance, between geno- and ecotypes of a species.

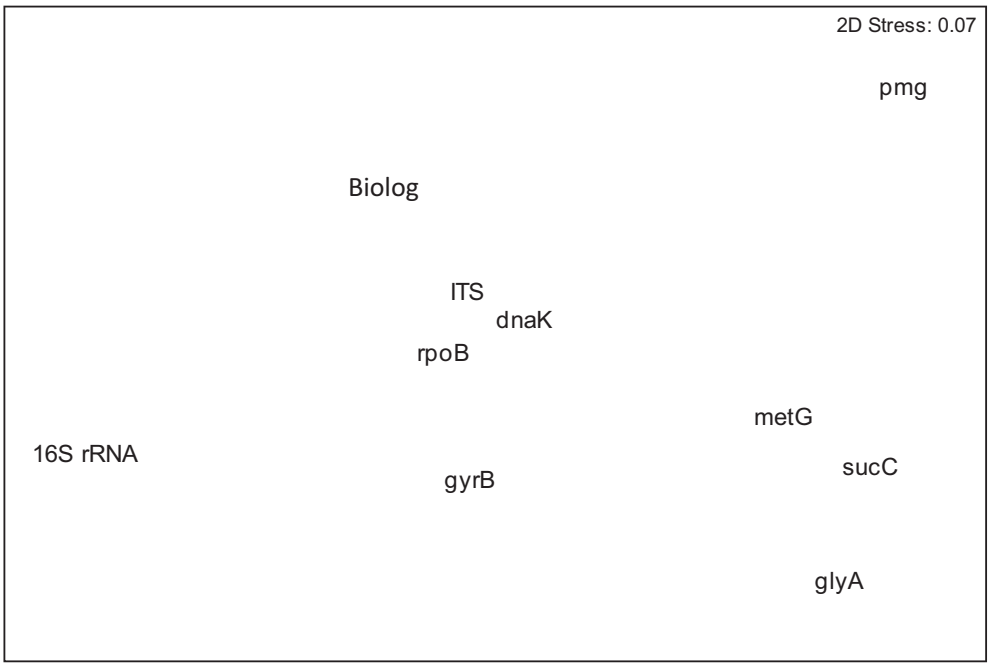


Fig. 19: Second-stage MDS analysis of the phylogenic relationships derived from nine marker genes of 23 *Alteromonas macleodii* isolates. The MDS plot is based on the rank correlations between every pair of distance matrices calculated from the alignments of the individual marker genes. Marker genes that group closely show more similar relationships (based on their individual distance matrices) than those that are more distant in the MDS plot. Additionally, the similarity between strains concerning their individual substrate utilisation assessed using Biolog plates was also included in the analysis. The stress value of 0.07 indicates that the distances in the multi-dimensional space are well represented in this two-dimensional plot. (Taken from Ivars-Martínez *et al.* 2008).

3.1.2 Development of PCR primers specific for phylogenetic groups

The 16S rRNA gene has been established long time ago as the 'gold standard' for taxonomic identification and for phylogenetic analyses of microorganisms (e.g. Amann *et al.* 1995). The approach taken in such analyses is (Sanger) sequence analysis of PCR amplified 16S rRNA gene fragments followed by bioinformatic comparison of the sequence with those of, among others, closely related taxa. Therefore, efforts have been undertaken to identify the less abundant components of microbial populations. Increasing the number of clones per clone

library has been successful in detecting novel bacterial clades (e.g. Chouari *et al.* 2005). However, despite the decreasing costs for nucleotide sequencing using the Sanger method, the success of this approach has long been limited because of the huge scale of bacterial diversity. (In contrast to this, the archaeal diversity is much lower and rarefaction analysis of sequence data obtained by next generation sequencing techniques indicate that sequence analysis of approx. 20,000 16S rRNA fragments is sufficicent to cover most of the archaeal diversity in a coastal marine sample (in this case collected in the English Channel about 15 km off Plymouth, UK). So, prior to the arrival of next generation sequencing techniques analyses of bacterial diversity required a solution that enaled the researcher to achieve an in-depth investigation of, at least, those taxa that were specifically under investigation. A simple solution to this problem is provided by the use of PCR primers that specifically anneal to the 16S rRNA genes of those particular phylogenetic groups. Such group specific PCR primers have long been in use (see Amann *et al.* 1995). As 16S rRNA gene sequence databases continue to grow, group-specific PCR primers must be continually re-evaluated for their specificity range of sequence matches (Baker *et al.* 2003). Such an attempt has been undertaken by the Author in a study (Mühling *et al.* 2008) that resulted in the design of group-specific primer pairs for the specific amplification of strains of three phyla (*Bacteroidetes*, *Planctomycetes*, *Firmicutes*), three classes (*Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*) and for the *Cyanobacteria* (including chloroplast 16S rRNA sequences). The evaluation of the primer pairs was conducted by sequence analysis of clone libraries produced from DNA samples of a mesotrophic (Norwegian coastal) and an oligotrophic (Northern Atlantic gyre) environment. The results of these tests indicate that five of the seven primer pairs specifically amplified target 16S rRNA gene sequences. Exceptions were the *Betaproteobacteria*- and *Firmicutes*-specific primers, which were relatively successful with coastal water mesocosm samples but less so with the Northern Atlantic Gyre sample. Phylogenetic analysis of sequences from the *Gammaproteobacteria* clone library revealed that the coastal sample yielded a number of clones that clustered within clades that belong to the oligotrophic marine *Gammaproteobacteria* (OMG) group, indicating that this group is not confined exclusively to the oligotrophic environment (see also below under setion 3.1.3). The group-specific primer pairs were also tested for their feasibility in genetic fingerprint application Denaturing Gradient Gel Electrophoresis (DGGE). Comparison of the bacterial diversity of the environmental DNA sample from the coastal and the open ocean using a two- or three-step nested PCR-DGGE process revealed significant differences in the bacterial communities. Overall, the results indicated that the application of the group-specific

primers provides a higher resolution genetic fingerprinting approach than existing DGGE primer sets.

In summary, prior to the broad availability of next generation sequencing techniques in microbial ecology studies the application of group-specific primers provided a simple and effective solution to the problem of in-depth analysis of bacterial diversity. Moreover, since Sanger sequence analysis of individual clones of a clone library prepared with group-specific libraries is still a sensible approach to achieve precise taxonomic assignment of 16S rRNA gen fragments. That is, using the primers suggested by Klindworth *et al.* (2013) Illumina sequence analyses on a MiSeq sequencer still only provide ca. 427 bp of the 16S rRNA gene (the V3-V4 region) *via* the paired end sequence analysis of PCR amplified 16S-tags (i.e. 600 bp maximum sequence length minus the primer sequences that sum up to 105 bp: Klindworth *et al.* 2013). In contrast to this limitation, Sanger sequencing using, equivalent to paired end Illumina reads, sequence analysis with both forward and reverse primers (e.g. those used to amplify the 16S rRNA gene fragment or primers binding to binding sites on the vector used for cloning) routinely results in full length sequences of the amplified fragment (up to the ca. 1450-bp fragment obtained with primer pairs 9bfm/1512uR: Mühling *et al.* 2008). So, while 16S-tag analysis provides a more comprehensive analysis of the range of genetic diversity within a sample, the traditional Sanger sequence analysis of individual clones from a clone library achieves much more precise taxonomic assignment. A combination of both approaches therefore may appear to be necessary in some instances.

Given the advances made in the development of next generation sequencing techniques a discussion on their potential to overcome the limitations mentioned above one should not shut oneself off progress, but have a sensible discussion on their potential for microbial ecology. As already mentioned, Illumina sequence analysis of PCR amplified 16S-tags, while being the simplest approach for the average microbial ecologist, does and will for (presumably) the medium-term future not provide sequence length that are required to obtain reliable taxonomic identification of microbial taxa within yet uncharacterised samples. (This stands in contrast to, for instance, samples from clinical studies composed of a restricted set of known microbial pathogens.) Pyrosequencing of yet uncharacterised samples was hoped to provide much better chances of a robust taxonomic assignment of sequences due to the longer sequence reads (e.g. 650 bp, though sequence reads of genomic DNA libraries have been shown to reach 1kb). However, pyrosequencing is likely to become less available over the

next years since Roche, who owns the patents on the technology, has announced to stop its support for this technology as of 2016. A real alternative, though to the best of the Author's knowledge not used to any extent so far, is provided by the application of the Single Molecule Real Time (SMRT) sequencing technology developed by Pacific Biosciences Inc. Using SMRT sequence analysis of circular consensus sequences (CCS) it is currently possible to analyse more than 10,000 16S rRNA gene sequences (ca. 1.5 kb fragments) in a single run with >99.5% predicted accuracy (Bowman *et al.* 2013). Multiplexing of samples may also be considered for SMRT sequencing since the price tag for this approach is still rather high (ca. 2,500-3,000 euros per sample including library preparation and sequence analysis by a commercial service provider: personal inquiries by the Author – December 2015). Although multiplexing results in a reduction of sequences obtained per sample (e.g. at 3 x multiplexing, each SMRT cell generates ca. 5000 1500-bp 16S rRNA gene sequences per sample with >99% predicted accuracy: ASM poster), this still represents a much cheaper option than Sanger sequence analysis of clones from a clone library. Application of the SMRT technology, possibly in combination with the use of PCR primers for the specific amplification of 16S rRNA gene fragments of the bacterial group under investigation may therefore offer a relevant and realistic alternative to 16S-tag sequencing in cases where identification below the genus level is required.

In the context of the design and evaluation of 16S rRNA gene targeting PCR primers it should be noted that some of the primers in use are, based on today's knowledge, indeed not appropriate for applications in microbial ecology studies. An example of this is provided by primer 1387r (reference) that has hitherto been used in the microbial ecology studies at the Author's current work place. Comparison of the results of an Illumina sequence analysis of 16S-tags amplified with the universal primer pair suggested by Klindworth *et al.* (2013) with those from the sequence analysis of a clone library prepared with primer pair 27F/1387r from the same DNA sample indicated that sequences derived from strains belonging to the phylum *Bacteroidetes* are present in the former, but absent in the latter dataset (Voitel 2015). The presence of *Bacteroidetes* strains in the enrichment culture is further corroborated by the presence of at least one contig (GALL contig002851; see Fig. 31A in section 3.3.2) within the metagenomic dataset. Detailed comparison (Table 11) of selected sequences detected by the 16S-tag approach demonstrated that the absence of those sequences within the clone library is most likely due to the mispriming of primer and target 16S rRNA gene binding site at two positions including the 3'-end position of primer 1387r (i.e. the 5'-site of the binding site). A

cloning bias against 16S rRNA gene fragments from taxa of the *Bacteroidetes* into standard cloning vectors seems, in comparison, to be rather unlikely.

Table 11: Binding site primer 1387r which targets the 16S rRNA gens of the whole range of bacterial diversity (i.e. a 'bacterial 16S primer'). Taken from Voitel (2015).

Taxonomische Gruppe	**Bindungssequenz**
Primerbindungssequenz	5'-GCCTTGTACACWCCGCCC-3'
Sideroxydans lithotrophicus	
***Sediminibacterium salmoneum*, *Flavobacteria bacterium* KF 030**	5'-ACCTTGCACACACCGCCC-3'

3.1.3 OEZY – a new software tool for the selection of the most suitable restriction endonucleases for TRFLP analyses

Prior to the availability of next generation sequencing techniques microbial ecologists have developed various crafty techniques to overcome the limitation associated to the sequence based approach, in particular the relatively high costs for Sanger sequencing of clones from clone libraries (e.g. of 16S rRNA gene fragments), and to faster obtain results from a large number of samples. These genetic fingerprint techniques employ, by and large, PCR amplification of a fragment of a marker gene followed by electrophoretic separation of those (or fragments thereof). The fingerprint techniques that have probably been most widely used by microbial ecologists are denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (TRFLP) analyses. The Author's contribution to the development of the DGGE approach has already been cited above (section 3.1.2). His contribution to the advance of the ***T***erminal ***R***estriction ***F***ragment ***L***ength ***P***olymorphism (TRFLP) has been achieved in the framework of a Bacherlor thesis (Müller 2011) and with the support of the informatician MSc Rico Beier (University of Applied Sciences, Mittweida, Saxony) a new approach for the ***O***ptimising ***E***n***ZY***me selection for best performing restriction fragment length polymorphism (TRFLP) analysis using ARB (OEZY). In addition to encompassing most of the advantages provided by the currently available programmes, OEZY

also overcomes their limitations (Mühling *et al.* 2016). This new and user friendly approach is based on ARB (Ludwig *et al.* 2004) and the previously developed TRF-CUT (Ricke *et al.* 2005) since the ribosomal RNA genes remain to be the most used markers for TRFLP analysis. The availability of extensive SILVA databases of aligned 16S and 23S ribosomal RNA gene sequences within ARB (Pruesse *et al.* 2007; Quast *et al.* 2013) therefore simplifies the selection of appropriate representative sequences and facilitates RE selection and TRF calculation in direct comparison to their phylogenetic resolution without the need of additional software (Ricke et a., 2005). Moreover, the fact that all ribosomal sequences are aligned within ARB additionally permits the prediction of lacking sequence data for individual sequences included in the set used for RE selection (Ricke *et al.* 2005). However, apart from the ribosomal RNA gene sequences, OEZY is also applicable for the analysis of user added sequences that are intended to be used as marker genes instead of the ribosomal RNA genes.

Another strength of the new software lies in the evaluation and scoring of the identified sets of REs. For instance, OEZY predicts the phylogenetic resolution in form of a dendrogram resulting from the TRFLP analysis of the selected sequences with a specific set of REs. This dendrogram is calculated from a distance matrix that is based on the TRF patterns from each individual input sequence after in silico digest with each RE of a specific RE set. The results are scored in form of '0' and '1' depending on whether the RE results or does not result in different TRF between two sequences, respectively. The sum of these individual scores based on each of the various REs of a specific set of REs forms the distance matrix. The distance matrix therefore contains the distance based on the differences in the resulting TRFs between each pair of two sequences. Clustering of the sequences (i.e. the TRF pattern for each of the sequences) in form of a dendrogram is then carried out using the unweighted pair group method with arithmetic mean (UPGMA: Sokal and Michener 1958).

Dendrograms can be directly compared to the phylogenetic information of the same set of sequences (i.e. the 'sequence based dendrogram', that is the dendrogram based on rRNA gene sequence data within ARB which should also be calculated using also the UPGMA approach). The conclusions from this visual comparison are further substantiated by the correlation between the distance matrix calculated with the sequence data within ARB and that based on the TRF pattern (Spearman rank coefficient) and by further scores (see below) that provide quantitative measures for the resolution achieved by TRFLP analysis with the selected set of REs. In summary, the new program not only provides a user friendly tool for the selection of

optimal sets of REs of ranked target sequences but also analyses the results in terms of the achieved resolution and assesses it by directly comparing it to that based on the complete sequence information. The knowledge of the correlation between the TRFs and the phylogeny based on the marker genes thus has the potential to identify novel TRFs to the level of genetic resolution achieved by the TRFLP approach.

3.2 Applications of genetic fingerprint techniques

'Classical' genetic fingerprint techniques and sequence analysis of clone libraries

Apart from contributions to the further development of genetic fingerprint techniques the Author has also applied those techniques for the analysis of microbial communities. Given today's possibilities provided by the highthroughput sequencing techniques for the analyses of PCR amplified 16S-tags, the more 'classical' fingerprint approaches may almost appear to be 'trivial' in terms of their genetic resolution and size of datasets produced. However, they still resulted in scientific findings that are, in some cases, of significance even today. Examples for such studies involve the analysis of the –

i) microdiversity of abundant groups of marine cyanobacteria and the potential impact of viral control on those (Mühling *et al.* 2005, 2006; Jameson *et al.* 2008, 2010, 2011);
ii) potential danger of environmental pollution by anthropogenically produced nanoparticles (Bradford *et al.* 2009);
iii) impact of increasing atmospheric pCO_2 on the microbial diversity in the marine environment (Arnosti *et al.* 2011);
iv) microbial population dynamics in sulfidogenic bioreactors used for bioremediation of AMD waters (Klein *et al.* 2013).

Particular emphasis is placed on one of these studies (Mühling *et al.* 2005) since this demonstrated for the first time viral control of their cyanobacterial host (*Synechococcus*) in a natural oligotrophic marine environment (that is, analysing samples from the marine environment rather than from a laboratory or a mesocosm experiment: e.g. Wilson *et al.*

1996). Moreover, the fact that the genetic diversity of host and virus were found to be mutually interactive factors also indicates that cyanophage-resistant mutants do not dominate the host population in the oligotrophic regions of the oceans as has been suggested (Waterbury and Valois, 1993), but is in keeping with the study of Suttle and Chan (1994) who demonstrated lack of resistance in offshore environments. In contrast to this, the physiological burden from sustaining resistance against co-occurring viruses may be a worthwhile investment for *Synechococcus* living in coastal environments where higher nutrient concentrations result in higher host cell densities and, hence, a higher encounter rate of host and virus.

The 'next generation fingerprint techniques': use of high-throughput sequencing methods

In order to improve the genetic resolution with which microbial communities were analysed the Author also utilised next generation DNA sequencing techniques that, during the last15 years, have been developed to a level that made them applicable to also answer typical questions arising in microbial ecology. Sequence analysis of hundreds or thousands of PCR amplified 16S rRNA fragments, so-called 16S-tags, per sample using high-throughput sequencing techniques not only provides higher genetic resolution of a microbial assemblage than achieved by 'classical' fingerprint approaches, but also offers partial taxonomic identification. That is, by comparison with ribosomal sequence databases (e.g. RDP II, SILVA) the 16S-tag sequence reads can be assigned to microbial taxa, though the relatively short sequence does, in general, still not permit species identification. (Some exceptions exist, such as the analysis of the microbial community of samples that are typically characterised by a low and/or known genetic diversity.)

Examples of the efforts undertaken by the Author include the analysis of the –

i) seasonal and diel temporal variation of the microbial diversity in a temperate coastal marine environment (Gilbert *et al.* 2010a);
ii) impact of ocean acidification on the bacterial community structure in high-latitude marine coastal mesocosms (Roy *et al.* 2013);
iii) the bacterial diversity in groundwater used for the abstraction of drinking water.

Key findings from the latter of these studies shall be outlined in the following in some detail since these are, in contrast to the former, not yet published.

Assessment of the microbial diversity in groundwater used for drinking water abstraction by 16S-tag pyrosequencing and physiological analysis of enrichment cultures

Of the Earth's total water ressources freshwater represents 3% with approximately two thirds of it being confined in ice caps and glaciers. Almost 99% of the remaining freshwater is formed by groundwater (Younger 2007) which makes groundwater an invaluable source for the abstraction of drinking water. This importance for humankind is demonstrated, for instance, by the fact that on average 60% of the drinking water used in Europe derives from groundwater sources (Steube *et al.* 2009). Therefore, groundwater produced for drinking water abstraction undergoes constant monitoring for a variety of geochemical parameters. This assessment, which is regulated by national and international (e.g. EU) legislation and which follows standardised procedures, includes in particular the quantitative determination of the inorganic (cations, anions) and organic (xenobiotics) fraction and derived variables (e.g. calcite dissolution capacity). In contrast to the wide range of chemical and physical measurements (e.g. turbidity) the quantitative determination of microbiological variables is limited to *Escherichia coli*, coliformes and enterococci (as proxy for feacal contamination).

Groundwater systems around the Earth have been shown to harbour complex microbial assemblages which, in turn, are shaped by the chemical and physical nature of the groundwater, thus underlining the need to be considered as ecosystem. This has also been recognised by several national regulatory agencies. As a consequence, appropriate measures have been proposed in national monitoring programmes (Steube *et al.* 2009; Griebler *et al.* 2010; Stein *et al.* 2010; Korbel and Hose 2011). Nevertheless, the detailed analyses of the indigenous microbial fraction in groundwater have so far only been performed at national scale in New Zealand (Sirisena *et al.* 2013). Reasons for this are likely to be connected to the technically demanding analyses and associated costs, both preventing routine monitoring beyond research programmes and outside research institutions. In cases where contamination of groundwater aquifers or depleting production levels for drinking water abstraction are observed detailed analyses are, however, required to quantitatively determine hydrological, physical, chemical and microbial variables. Comparing the various datasets obtained from such a monitoring program may permit deduction of potential causes for the observed changes in water quality and productivity. One of such instances is groundwater with high concentrations of iron and sulfate which is either caused by the geological formation of the groundwater aquifer or by, for instance, anthropogenic mining activities (e.g. Eberle and

Razem 1985). However, the potential effects on or contributions to the problem by the indigenous microbial community have not yet been demonstrated.

The typical approach to access the groundwater is *via* groundwater wells which consist of a well of various depth ranging, for example, from 15 to 150 m or more. At the lower end the well wall is perforated by slits that permit the water from the aquifer to enter (Fig. 20). From there it is usually abstracted *via* a centrically positioned pipe and transferred for purification and subsequent distribution to households, workplaces and for recreational purposes. In the presence of high concentrations of ferrous iron in the groundwater this approach is, however, often impacted by the formation of ferric iron precipitates. This process, also referred to as iron ochre formation, occurs upon contact of oxygen to the (almost) anoxic groundwater with circumneutral pH since this results in abiotic oxidation of ferrous iron and, as a consequence, in declining abstraction of groundwater. Iron ochre formation is thought to be accelerated by microaerophilic iron oxidising bacteria. Strains belonging to the genus *Gallionella* are generally regarded as key players in this process. Mechanical cleaning is then required to liberate the well perforations from iron ochre deposits. Such remediation actions are costly and also interrupt continuous abstraction and delivery of groundwater.

This problem has been recognised by the German Federal Ministry for Education and Research (BMBF) which funded a consortium project (*"Entwicklung und Erprobung einer Technologie für die automatisierte Erkennung von Ablagerungsprozessen in Trinkwasser-brunnen und eine nachhaltige Brunnenbewirtschaftung" – BRUNNTEC*) that aimed at a better understanding of the microbial parameter associated with iron ochre formation. Consortium partners involved in the microbiological investigations of this project were the water company ZBL Bornaer Land (access to water wells and sampling) and the geo-engineering firm G.E.O.S. Ingenieurgesellschaft mbH, Halsbrücke (coordination, sampling and geochemical analyses). [Experiments leading to the results presented in the following have been undertaken by Dr. Christine Steinbrenner and M.Sc. Marlen Liebig under the supervision of the Author who also planned and designed the project. 16S-tag pyrosequencing has been carried out at the G2L, Goettingen University.]

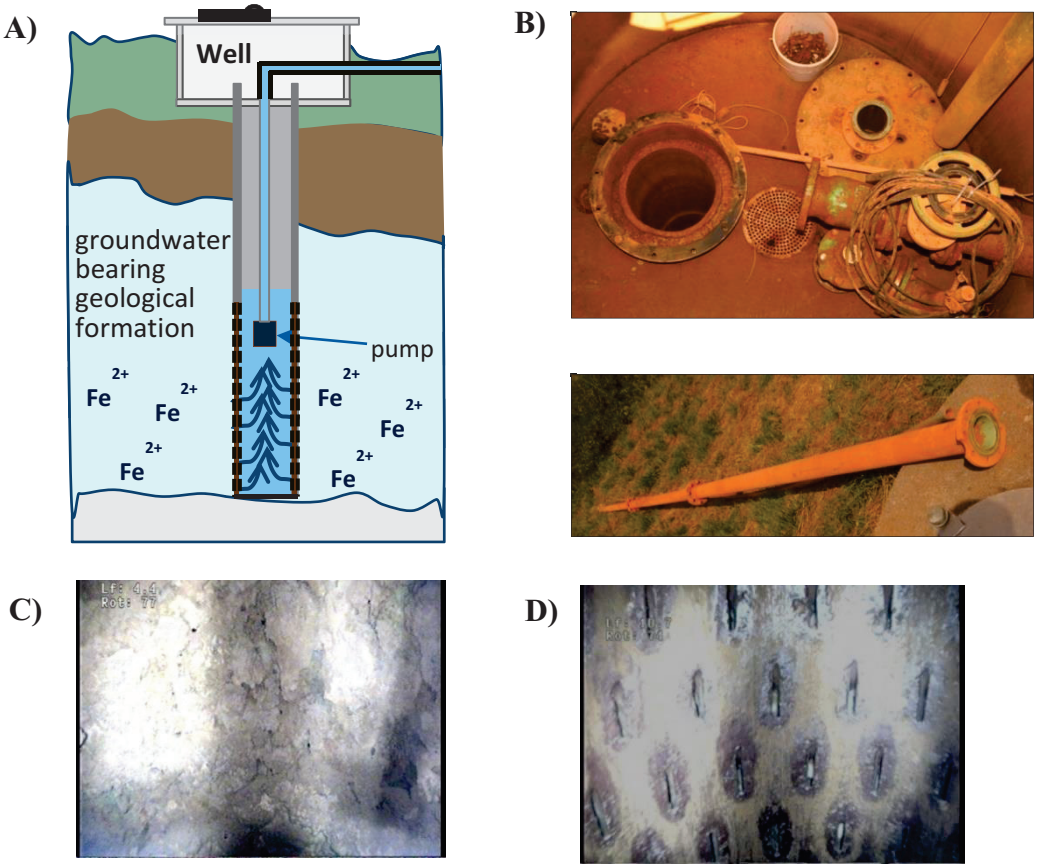


Fig. 20: Iron ochre formation in groundwater well Kitzscher 2 of the catchment area Kitzscher (see Fig. 21) which is managed by the water company ZBL Bornaer Land. A) Illustration of a typical groundwater well. Based on www.walterfiedler.de. B) Rust-like deposits within the well head and at the outer side of the central pipe used for water abstraction. C) and D) Photographs taken during *in situ* well inspection with TV-camera to control the well status prior (C) and following (D) a mechanical cleaning. Courtesy of ZBL Bornaer Land.

Results

Study area and sampling

The four groundwater wells investigated in this study are managed by the drinking water provider 'Zweckverband Wasser/Abwasser Bornaer Land' and are located approximately 25 km south-east of the city of Leipzig (Fig. 21). Groundwater produced by these wells derives from two groundwater catchments, Kitzscher (Ktz) consisting of a total of 14 water producing wells, and Kesselshain (Kh) with nine wells.

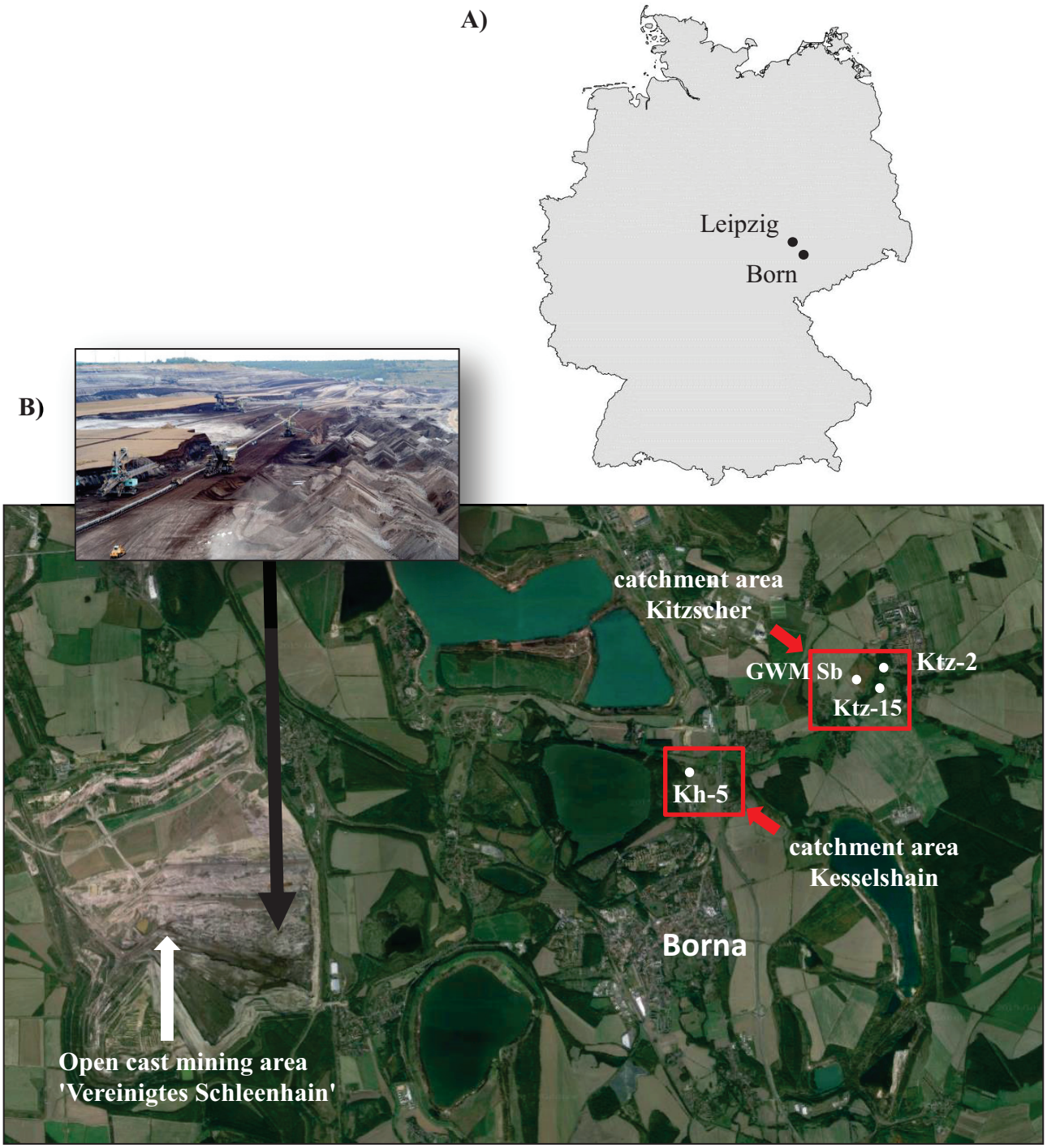


Fig. 21: Map indicating the study area (A) and aerial view showing the catchment areas Kitzscher and Kesselshain (red squares) that are managed by the water company ZBL Bornaer Land (B). The town of Borna is located approximately 25 km south-east of the city of Leipzig (A). The blue-green areas in (B) represent pit lakes that result from the lignite mining (inlet) within the mid-German lignite mining district. Samples were collected from groundwater catchment areas Kitzscher (wells 2 and 15), from the groundwater monitoring site Steinbach (GWM Sb) and from Kesselshain (well 5). Retrieved from Google Maps.

Selection of water wells for this study was based on representative differences of their water quality and productivity (Fig. 21). Three of the four wells are located in catchment area Kitzscher: two water producing wells (Ktz-2, Ktz-15) and the groundwater monitoring well Steinbach (GWM Sb). Catchment area Kesselshain was represented by well 5 (Kh-5). Table 12 summarises well and groundwater characteristics and highlights main differences.

Importantly, catchment Kitzscher is of tertiary origin (depth ca. 20m) and lies within an area that is impacted by land rehabilitation of parts of the mid-German open cast lignite mining district. In particular, the observed rise of the groundwater level following closure of parts of the open cast mining area and the flooding of pit lakes near the catchment area likely result in chemical changes to the groundwater bearing aquifers (Rösner 2011a). In contrast to catchment Kitzscher, tritium measurements indicate that the groundwater aquifer of catchment Kesselshain, which lies within the bedrock (depth ca. 130 m), is approx. 40 to 100 years old and practically unaffected by anthropogenic influences due to the overlaying massive stratum comprising conglomeratic sandstone followed by claystone and siltstone with unconsolidated sandstone (Rösner 2011b; Fig. 22). This conclusion is further corroborated by the lack of detectable nitrite (Rösner 2011b).

Sampling and chemical analysis

Ten litres of groundwater were filtered through two Sterivex cartridges with 0.2 µm pore size filter membrane (Millipore). The cartridges were stored at -80 °C until further analysis. Additionally, samples of iron ochre, either precipitated at the inner wall of the water producing well or sedimented at the bottom of the well, were collected at various time points.

Chemical and microbiological (*E. coli*, total coliforms, enterococci) analyses of water samples collected at the groundwater producing wells were performed by certified laboratories (Eurofins Ost GmbH or Kommunale Wasserwerke Leipzig GmbH) according to DIN (German Institute of Standards) guidelines.

Inorganic anions in samples collected from enrichment cultures and other laboratory experimental cultures (see below) were determined by ion chromatography using an ICS-5000 (4 mm system, Thermo Scientific) equipped with suppressed conductivity detection.

Table 12: Well characteristics of groundwater producing (Ktz-2, Ktz-15, Kh-5) and monitoring (GWM SB) wells within groundwater catchment areas Kitzscher and Kesselshain investigated in this study. Shown is the range of only the most relevant variables over the time period of the study. (nd = not determined)

	Kitzscher			**Kesselshain**
Well characteristic ↓	**Kitzscher 2**	**Kitzscher 15**	**GWM Steinbach**	**Kh-5**
Depth	21 m	20.5 m	10 m	139 m
Geological formation	tertiary	tertiary	tertiary	Lower buntsandstein
Groundwater retrieved from	main aquifer	top and footwall of main aquifer	main aquifer	main aquifer
Water quality (overall)	good, constant	medium, variable	bad	very good, constant
Water production	high	Low	none	high, but reducing
Temperature [°C]	10.7 – 12.0	10.5 – 11.4	10.7–11.0	14.0 – 15.1
pH	6.2 – 6.4	6.3 – 6.5	6.45	7.5
Fe^{2+} [mg/L]	4.0 – 5.3	3.9 – 4.0	5.9 – 6.8	0.27 – 0.35
Fe^{2+}/Fe^{3+} [mg/L]	4.5 – 11.3	4.1 – 4.6	6.7 – 8.0	0.31 – 0.37
$Manganese_{(dissolv.)}$ [mg/L]	0.09	0.3 – 0.3	0.1 – 0.2	0.05
Manganese [mg/L]	0.09	0.2 – 0.3	0.1	0.05
Sulfate [mg/L]	298 – 372	312 – 346	393 – 402	64 – 71
Schwefel [mg/L]	120 – 350	110 – 340	140 – 370	21 – 61
Oxygen [O_2/mg]	0.6 – 8.9	1 – 14	nd	0.7 – 1.1
Calcite dissolution capacity	67.6 – 92	83.5 – 123	60.1 – 83.8	-7.4 – 13.3

Anions were separated using an IonPac® AS11-HC column (Thermo Scientific) with 30 mM sodium hydroxide at a flow rate of 1.5 mL/min at 30 °C. Identification and quantification of anions was achieved by comparison to external standards. Soluble and precipitated sulfides were quantified according to the method of Cord-Ruwisch (1985). Ferrous iron and total iron was photometrically determined using the ferrozin based approach of Lovely and Phillips (1987).

PCR amplification of bacterial 16S-tags

A PCR approach was employed to amplify the V3 – V5 region of the 16S rRNA gene spanning approximately 570 bp. PCRs were carried out in 50-µl volumes and contained 200 µmol L^{-1} dNTPs, 1.0 U of Phusion DNA polymerase (in HF Buffer, ThermoScientific), 1.5 mmol L^{-1} $MgCl_2$, 500 nmol L^{-1} of each of the primers and 12.5 µg bovine serum albumin (BSA). All PCRs used the same cycle protocol: following an initial denaturation step of 3 min at 98ºC, 35 PCR cycles were performed (98ºC for 10 sec, 52ºC for 30 sec, 72ºC for 90 sec) followed by a final extension step at 72ºC for 5 min. Primers used in PCRs contained 10-bp sample specific indeces (10-bp MID), a key (underlined) and the Roche 454 pyrosequencing adaptors at their 5'ends. To maintain similar annealing temperatures the reverse primer also contained 10-bp MID sequences, though this was the same for each of the samples undergoing 16S-tag PCR amplification: V3for-MID: 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG(10-bp MID)TACGGRAGGCAGCAG-3'; V5rev-MID: 5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG(10-bp MID)GTGCTCCCCCGCCAATTCCT-3'.

Ten 50-µL PCRs were carried out in parallel in order to obtain sufficient product for subsequent library preparation. PCR products were concentrated *via* a purification step using the SureClean PCR purification kit (Bioline) prior to electrophoretic separation on a 1.5% (w/v) agarose gel (25V for 16h). The approximately 700-bp band (including PCR primers) was excised from the gel and purified using the QIAEX II gel extraction kit (Qiagen). The quality and quantity of the extracted DNA was determined using the NanoDrop ND-1000 spectrophotometer (NanoDrop) and the Bioanalyzer (Agilent) with the DNA 1000 kit (Agilent).

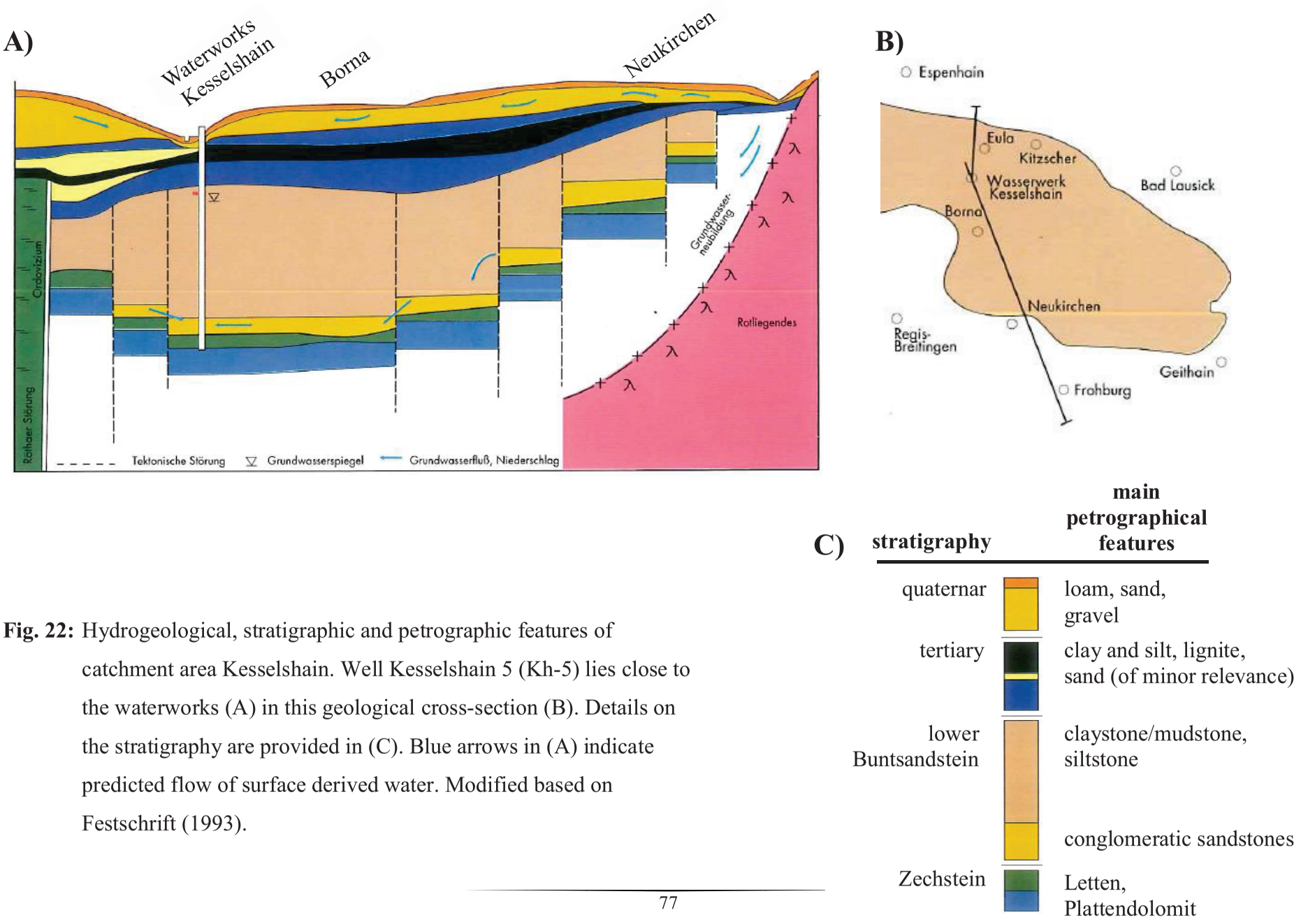


Fig. 22: Hydrogeological, stratigraphic and petrographic features of catchment area Kesselshain. Well Kesselshain 5 (Kh-5) lies close to the waterworks (A) in this geological cross-section (B). Details on the stratigraphy are provided in (C). Blue arrows in (A) indicate predicted flow of surface derived water. Modified based on Festschrift (1993).

Pyrosequencing of 16S-tags

Pyrosequence analysis of the 16S-tag amplicons was performed at the Göttingen Genomics Laboratory (G2L; Goettingen University) using a Roche GS-FLX 454™ pyrosequencer and the titanium chemistry (Roche). Sequence reads were analysed by Dr Bernd Wemheuer (G2L) using Qiime (Caporaso *et al.* 2010) that is part of an in-house software pipeline.

Culturing

The gradient culture technique of Kucera and Wolfe (1957) was employed to enrich for microaerophilic iron oxidisers using a modified version of Wolfe's mineral medium (800 mL: 0.8g NH_4Cl, 0.272g $MgSO_4$-7H2O, 0.08g CaCl-2H_2O, 0.032g KH_2PO_4; pH6.5; Wolfe 1958). Following sterilisation by autoclaving the medium was gassed with CO_2 for 20 min. Iron sulfide or iron carbonate obtained *via* the method of Emerson and Floyd (2005) or Hallbeck *et al.* (1993), respectively, and concealed within 0.5% (w/v) agarose was used as iron source and placed at the bottom of a 10-mL screw tab test tube overlayed by either 4 or 6 mL of liquid or semiliquid (0.15% (w/v) agarose) Wolfe's medium. Media in the test tubes (screw caps only loosely attached) were equilibrated to microaerobic conditions for 24h within a 2.5-L AnaeroJar incubator (ThermoScientific) containing a CampyGen sachet (ThermoScientific). The medium was subsequently inoculated with groundwater samples collected at various sampling occasions and incubated in the AnaeroJar incubator with a fresh CampyGen sachet.

TRFLP analysis of enrichment cultures

TRFLP analysis was undertaken to assess the microbial diversity within enrichment cultures or iron ochre samples from the groundwater wells. In principle, 16S rRNA gene fragments were amplified using primer pair 27f/1387r (5'-AGTGTTTGATCCTGGGTCAG-3' / 5'-GGGCGG(A/T)GTGTACAAGGG-3') with 27f as Cy5-labelled primer. PCR amplification was performed using the same protocol as above except for a higher annealing temperature (55 ºC) and fewer rounds of amplification (30 cycles). Amplicon purification and quantification were achieved as described above.

Restriction digestion of 20 ng of purified PCR product was performed in 10-µL reaction volumes with three restriction endonucleases (AluI, HaeIII, HhaI: Fermentas; 2U each, 2.5 h at 37 ºC). Terminal restriction fragments (TRF) of 60-600 bp were detected and quantified using a Beckman Coulter CEQ8000 genetic analyser. Peaks below the defined threshold of

10% were regarded as background noise and excluded from further analysis. TRFs were judged to be indicative of separate strains if they differed by a minimum of two nucleotides.

Statistical analysis

The data on the bacterial diversity within the 43 samples were evaluated for skewness using Statgraphics Centurion XVI.II (StatPoint Technologies, Inc.). Since this test revealed non-uniform distribution the data were interrogated using non-parametric multivariate analyses which were carried out using the Primer version 6.0 software package (Primer-E Ltd.: Clarke and Warwick 2001). A Bray-Curtis distance matrix was calculated from log transformed data on the occurrence of the 785 taxonomic groups in the 43 samples (11 samples for each of the wells except Ktz-15 for which data of only 10 samples were available). Distances between samples were visualised using multidimensional scaling (MDS) analysis. The significance of differences between *a priori* defined groups was tested by one-way analysis of similarities (ANOSIM), a test of the hypothesis that rank similarities within groups are greater than rank similarities between groups.

Principal component analysis (PCA) was used to visualise differences between the wells based on the 45 geochemical parameters determined in the framework of the monitoring programme.

Results and discussion

Well and catchment specific groundwater characteristics based on geochemical parameters

PCA of 45 geochemical parameters analysed during the 21-months monitoring programme underlines the expected difference between the quality of the groundwater from the two catchment areas (Fig. 23). The groundwater abstracted from Kesselshain well 5 has lower concentrations of sulfur compounds and iron and its pH is continuously higher than that of the groundwater within catchment area Kitzscher (Table 12, Fig. 23). Overall, Kesselshain 5 represents a deep bore well at which groundwater of good quality has been continuously abstracted for the last 20 years (Rösner 2011b). The chemical quality of the groundwater from the three Kitzscher wells investigated in this project also varied with oxygen und turbidity as the main variants. While it is easy to conclude that the water chemistry of GWM Steinbach is different to that of the two groundwater producing wells, the reason for the observed differences between the groundwater from Kitzscher wells 2 and 15 is less obvious.

Individual groundwater abstraction schemes employed (i.e.higher and constant groundwater abstraction at Kitzscher well 2 versus well 5) and access to different layers of the main aquifer (i.e. main aquifer versus top and footwall of main aquifer in Kitzscher wells 2 and 15, respectively: see Table 12) may contribute to this difference.

PC	Eigenvalues	%Variation	Cum.%Variation
1	9.66	82.1	82.1
2	0.92	7.8	89.9

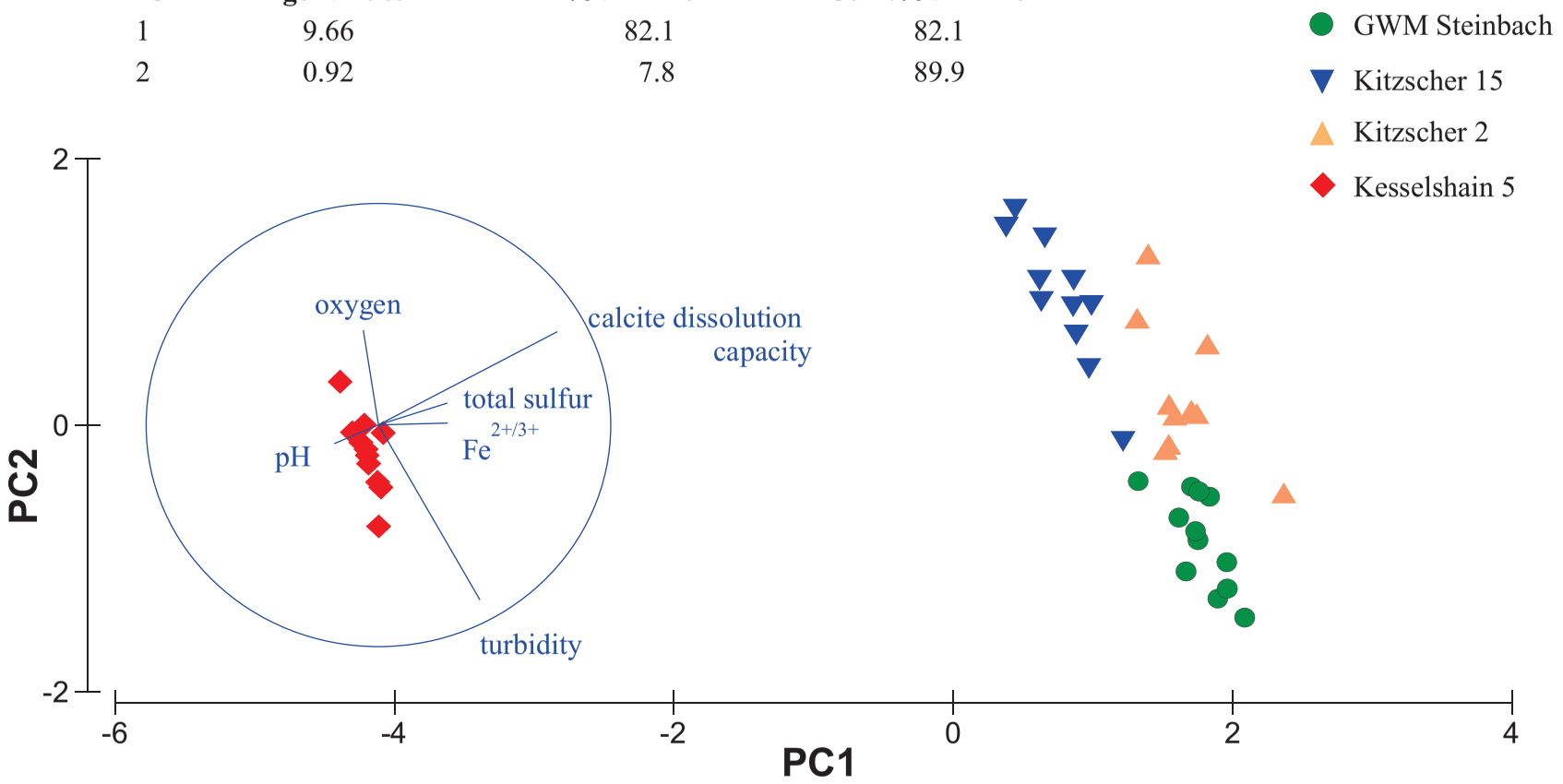


Fig. 23: Principal compnent analysis (PCA) of the 45 geochemical parameters analysed in the study. The first two axes explain approximately 90% of the variation observed between the wells. The five parameters that contribute most to the variability are indicated within the plot.

Bacterial diversity within groundwater from two catchment areas

Data on the bacterial diversity were obtained *via* 16S-tag pyrosequence analysis for 43 of the 44 samples collected in the study. Due to lack of successful DNA isolation no sequence data are available for Ktz-15 from 7. July 2013. An average of 2313 sequence reads (≥350 nt) per sample were obtained after quality trimming, totalling 99472 reads. The sequence reads were clustered into OTU at 97% sequence similarity resulting in 3994 OTU (data not provided) which, based on comparison to the SILVA database (Pruesse *et al.* 2007; Quast *et al.* 2013), were clustered into 785 taxonomic groups (data not provided).

Diversity measures generally represent two components – richness and evenness. Richness reflects the number of entities present, whereas evenness reflects the distribution of individuals among entities. The bacterial population detected by 16S-tag pyrosequencing varied between temporal samples and wells, yet still indicates that richness in samples from Kesselshain 5 and GWM Steinbach was generally higher than in samples from Ktz-15 and Ktz-2 (Fig. 24).

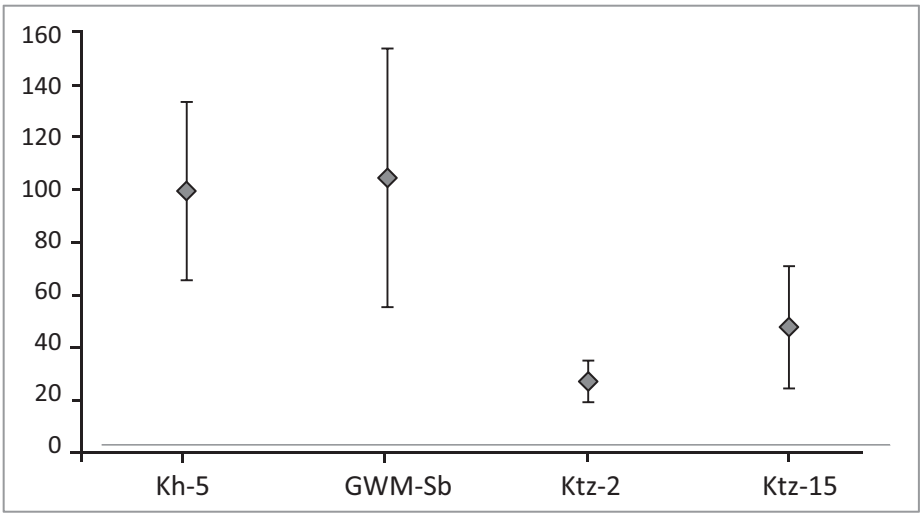


Fig. 24: Genetic diversity of bacterial populations in groundwater samples from the four wells. The genetic diversity is expressed for each of the four wells as the average of the number of OTU detected in the samples collected at different time points (i.e. richness) with the standard deviation indicating the variation among samples.

Bacterial taxa were also more evenly distributed in samples from wells Kesselshain 5 and GWM Steinbach (Fig. 24), though members of the *Nitrospiraceae* and *Lachnospiraceae* (Kesselshain 5) or strains belonging to the genus *Gallionella* (GWM Steinbach) dominated at various time points (Fig. 25). It should be noted in this context that the analysis of the bacterial diversity did not include data for those samples that were collected on 10. June 2013 from Kitzscher 15 and GWM Steinbach. The dominance of members of the genus *Flavobacterium* (Kitzscher 15) or of the *Escherichia coli – Shigella* group (GWM Steinbach) indicates that these samples were affected by contamination from surface soil caused by flooding of the well area after heavy rainfalls prior to the time of sampling (details not provided). The unique bacterial composition of these two samples is further corroborated by the results of the multi-dimensional scaling (MDS) analysis of the data obtained by the 16S-tag pyrosequencing approach (Fig. 26).

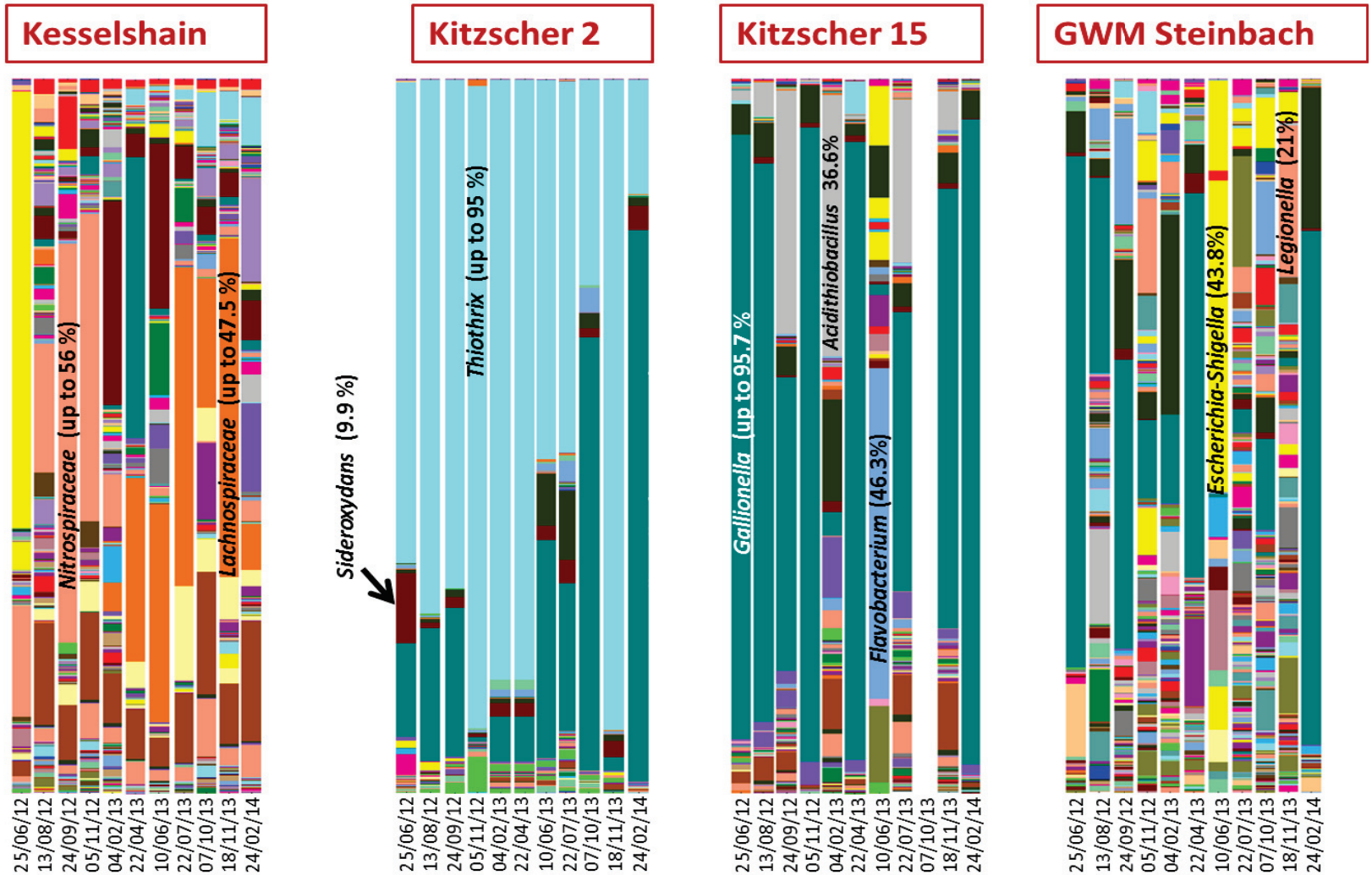


Fig. 25: Taxonomic identification of bacteria in groundwater samples from the four wells based on 16S-tag sequence reads. The abundance of the individual taxonomic groups is provided as percentage of total 16S-tag reads per sample. For the purpose of clarity only the names of the most abundant or relevant taxonomic groups (out of a total of 785) are shown. No data are available for the sample Kitzscher-15 from 7. October 2013 due to unsuccessful DNA extraction.

Apart from this, the MDS analysis also demonstrates that the rest of the samples roughly group according to their origin; that is, to the well the groundwater was collected from (Fig. 26). The generally higher similarity of the bacterial composition among samples from the same well as compared to samples from another well was further confirmed by an ANOSIM test. The samples (Kitzscher 15 and GWM Steinbach from 10. June 2013) that presumably suffered from contamination due to flooding of the surface soil after heavy rainfalls were again omitted from this analysis. The results from the ANOSIM test shows that the difference between these *a priori* defined groups (i.e. samples grouped according to their origin, that is the individual wells) was highly significant based on their bacterial diversity ($R = 0.81$, $P = 0.001$). In contrast, a lower correlation ($R = 0.543$, $P = 0.001$) was found if catchment area was chosen as *a priori* defined group. Although the 16S-tag analysis revealed temporal changes of the microbial composition the dynamics of these remain unpredictable.

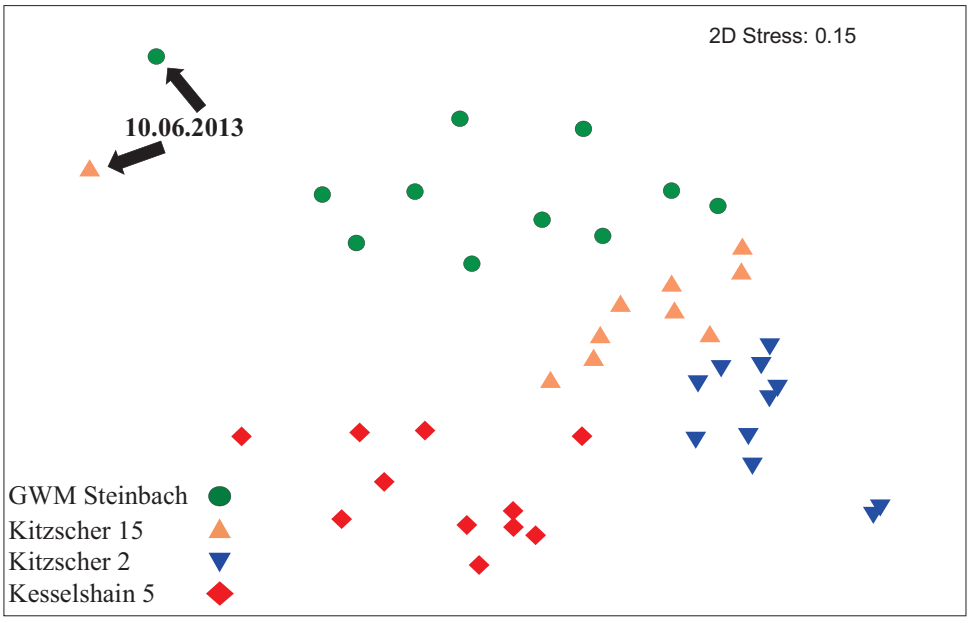


Fig. 26: Nonmetric multidimensional scaling (MDS) ordination plots of Bray-Curtis similarities calculated from log transformed data on the occurrence of the 785 taxonomic groups detected by 16S-tag pyrosequence analysis. The MDS plot illustrates the relative variation in the bacterial composition within the groundwater samples from the four wells throughout the sampling period. The time points of sampling are not indicated within the plot in order to maintain clarity. Relative distances between samples reflect inter-sample similarities, with samples close together being similar in terms of their bacterial composition. The stress value (Kruskal's stress formula 1) < 0.15 indicates acceptable representations of multivariate relationships in two dimensions.

The dataset of geochemical parameters collected for the same groundwater samples analysed by 16S-tag pyrosequencing was further interrogated to assess the potential impact of the geochemical quality of the groundwater on the bacterial composition. These analyses, excluding again the samples collected from Kitzscher 15 and the GWM Steinbach on 10 June 2013, revealed that bacterial diversity also correlates with the geochemical composition (log-transformed data) of the corresponding samples (Spearman rank correlation: $\rho = 0.591$, $P = 0.001$). Thus, it appears that well specific geochemical parameters shape the bacterial community. However, additional statistical tests (e.g. the BioEnv approach: Clarke and Warwick 2001) did not identify particular geochemical parameters that correlate with and, thus, are likely to be the drivers for the observed bacterial composition.

Presence of members of candidate divisions and of microbial indicators for feacal contamination

In total 22 of the currently accepted 30 bacterial phyla were detected among the 43 samples, though some of those are – as mentioned above – likely to be the result of contamination from sampling after heavy rain caused flooding of the well area (i.e. samples Kitzscher 15 and GWM Steinbach collected on 10.06.2013). Additionally, 17 proposed phylogenetic groups at phylum level that are comprised of yet uncultured and uncharacterised bacteria (i.e. candidate phyla) were also detected.

Based on the percentage of 16S-tag reads the groundwater collected from Kesselshain well 5 harboured a higher fraction of cells belonging to these candidate divisions or phyla (on average 18.1%) than the groundwater collected from the wells of catchment area Kitzscher (well 15: 5.9%, well 2: 2.3%, GWM Steinbach: 4.1%). Candidate divisions OP11 (Kesselshain 5: 10.2%, Kitzscher 15: 3.5%, Kitzscher 2: 0.4%, GWM Steinbach: 0.2%) and BD1 (Kesselshain 5: 3.1%, Kitzscher 15: 1.3%, Kitzscher 2: 0.2%, GWM Steinbach: 0.2%) were the two most abundant uncultured groups within the groundwater samples. Apart from these high level phylogenetic divisions there were also further candidate groups at lower taxonomic level in many of the 43 samples, though these made up only a small fraction in any of those cases.

Although candidate divisions such as the OP11 have been detected in a variety of environments (e.g. Hugenholtz *et al.* 1998; Harris *et al.* 2004), their observed abundance within groundwater from Kesselshain 5 is possibly also supported by the depth of the aquifer (approx. 130 m) which may have prevented isolations endeavours in the past. Additionally, the depth also means that changes in environmental conditions and regular input of soil microbes is prevented and, if it

occurs, the growth of indigenous microbes is likely to be favoured over that of invaders which tend to be adapted to different environmental conditions (e.g. higher nutrient loads).

Nevertheless, the bacterial diversity in samples from Kesselshain 5 collected from April to November 2013 was characterised by the presence of a large fraction of members of the *Lachnospiraceae* (Fig. 25: up to 47%). This raises questions as to the quality of the groundwater from Kesselshain 5 since *Lachnospiraceae* are so far only known to be associated with the digestive tracts of mammals, including cattle, which led to them being proposed as microbial indicators for feacal contamination (Meehan and Beiko 2014). A possible explanation for feacal contamination of the groundwater at this depth may possibly be provided by the prevailing hydrogeological features resulting from tectonic fractures and displacements which may allow surface water to enter geological formations below the conglomeratic sandstone formation and, hence, the aquifer that feeds Kesselshain well 5 (Fig. 22). However, since this hypothesis is in contrast to the above mentioned results from previous tritium measurements and the lack of detectable nitrite (see above) an alternative yet unknown life style of the *Lachnospiraceae* strains detected in this study may be the more likely explanation.

Enrichment culture confirm the viability and activity of iron oxidising bacteria

Since members of the genus *Gallionella* seemed to dominate the bacterial community in the groundwater samples collected from Kitzscher well 15 and also represent an abundant group in samples from wells Kitzscher 2 and from the GWM Steinbach, attempts were undertaken to further corroborate their identity and physiological activity. This is important because these three wells suffer from iron ochre formation which *Gallionella* is likely to contribute to. However, quantifying the microbial fraction of iron ochre formation at circumneutral pH represents a very delicate task due to the concurrent abiotic (i.e. purely chemical) oxidation of ferrous iron to ferric iron. Therefore, the approach chosen here was based on enrichment cultures using the gradient tube technique developed by Kucera and Wolfe (1957) and successfully applied by Engel and Hanert (1967) for the isolation of microaerophilic *Gallionella ferruginea* (Fig. 27). Microscopic analysis of these enrichment cultures revealed the presence of *Gallionella* typically recognised by the presence of long stalks (Fig. 27). This observation was further substantiated by the results from analyses using TRFLP which also demonstrated the presence of *Gallionella* within the enrichment cultures and within iron ochre precipitates collected from the inner wall of well Kitzscher 15 (data not shown). In summary, these findings corroborate the involvement of

Gallionella strains in iron ochre formation since ferrous iron oxidation is the only hitherto known energy providing process of these chemolithotrophic bacteria.

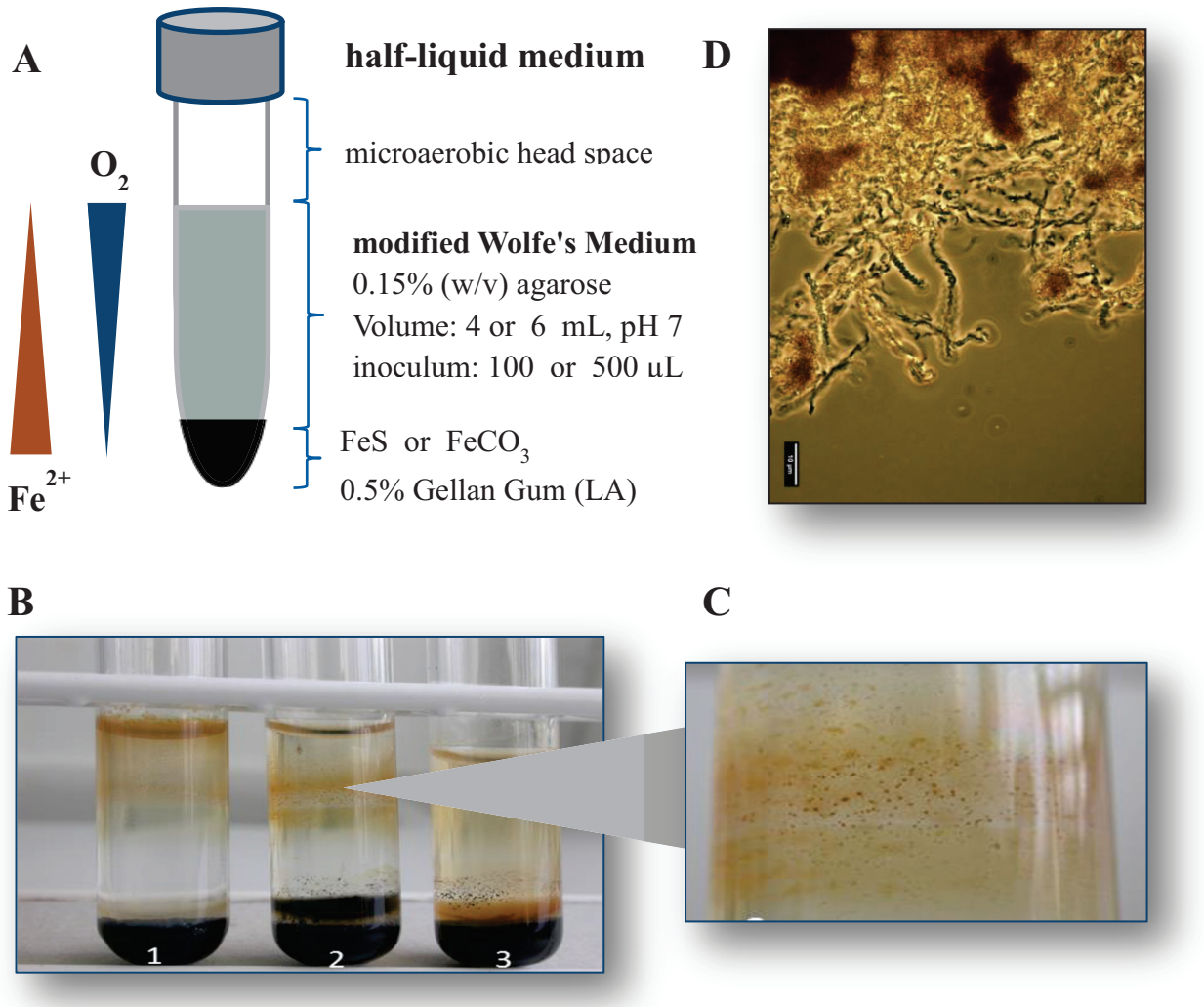


Fig. 27: Illustration of the gradient tube approach (A), the *Gallionella* enrichment cultures (B) with colonies growing at the surface of the inner test tube wall (C) and micrograph showing stalks typical for *Gallionella* (D).

The 16S-tag pyrosequence analysis of the bacterial community also exposed the presence of members of the genera *Acidithiobacillus* and "*Ferrovum*", both comprising acidophilic iron oxidising bacteria. Using TRFLP analysis these bacteria were also detected within iron ochre precipitates from the inner wall of well Kitzscher 15 (data not provided).

To confirm this observation attempts were again undertaken to obtain (enrichment) cultures of *Acidithiobacillus* and of "*Ferrovum*" using media known to selectively enrich for these acidophiles (iFe medium, pH 2.5 and APPW, pH 3.0 for *Acidithiobacillus* and "*Ferrovum*", respectively: Tischler *et al.* 2013). One of those enrichment cultures which, using TRFLP analysis, was found to be dominated by *Acidithiobacillus* (Fig. 28), was monitored for ferrous iron oxidation in iFe medium at pH 2.5 (optimum for *Acidithiobacillus* and to acidic for neutrophilic iron oxidisers such as *Gallionella* species). This test revealed that ferrous iron within the iFe medium (5 mM) was completely removed during the first 16 days of culture (data not shown). Since under acidic conditions (here pH 2.5) ferrous iron is stable also in the presence of oxygen its observed loss during this experiment (Fig. 28) must be the result of microbial activity. This conclusion was further corroborated by the results from the monitoring of an appropriate control experiment (iFe medium without inoculum) for which no decrease in ferrous iron concentration was detected over the same 16-day time period. Unfortunately, none of the attempts to enrich for "*Ferrovum*" strains proved successful in these experiments.

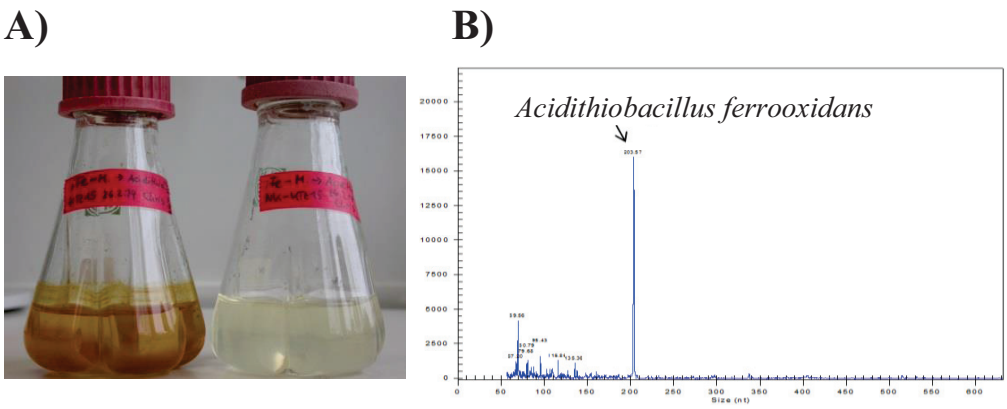


Fig. 28: Example of acidophilic enrichment cultures from groundwater samples of Kitzscher well 15. The culture (A: left flask) was obtained in iFe medium (pH 2.5) which favours the growth of iron oxidising *Acidithiobacillus* strains. No growth was observed in the uninoculated control (A: right flask). TRFLP analysis revealed that the iron oxidising culture was dominated by one or more strains of *Acidithiobacillus ferrooxidans* (B: TRFLP analysis was conducted with the restriction endonuclease HhaI).

The presence of acidophilic bacteria and the activity of at least *Acidithiobacillus ferrooxidans* like strains in enrichment cultures raises a number of questions regarding their origin. Clearly, incidences of locally confined formation of acidic habitats in which acidophiles like *Acidithiobacillus ferrooxidans* then thrive cannot be ruled out, though are difficult to explain with their discontinuous detection. An alternative and provoking hypothesis might be that the origin of those acidophilic strains is likely to be found in the mid German lignite mining district. Their detection, in turn, might therefore open an avenue for biomonitoring the spread of acid mine waters in aquifers. Further experimental evidence may be attainable from a detailed comparison of *Acidithiobacillus ferrooxidans* genotypes present in acidic mine waters of the mid German lignite mining district and the surrounding pit lakes with those retrieved from groundwater sampled within the catchment area Kitzscher. Parallel hydrogeological analyses including development of transport models and, possibly, tracer experiments might also be needed to supplement such an investigation.

3.3 Metagenomic approaches

Despite these interesting findings it must, however, be stressed that the results from any study that utilises the PCR method (preparation of 16S clone libraries, applications involving genetic fingerprinting techniques or 16S-tag sequencing) are potentially biased towards particular groups of microorganisms while excluding others (see section 3.1.2 and Table 11) and affected by artefacts connected to the PCR (e.g. von Wintzingerode *et al.* 1997).

A molecular approach that permits access to the genetic pool of microorganisms but circumvents PCR amplification is provided by improvements to the cloning of large DNA fragments and the advances made to DNA sequencing techniques. In microbial ecology these approaches are referred to as metagenomics and metatranscriptomics and are a synonym for the analysis of the total fraction of DNA and RNA, respectively, in a sample. However, the Author likes to note in this context that the term “meta” derives from the Greek prefix and preposition *μετά* meaning “after” (also: “with” or “among”), thus rather indicating that metagenomics and metatranscriptomics have evolved beyond analysing single genes and gene transcripts and their apparent diversity in a given environment to the large scale analysis of significant fractions of the total gene or transcript pool in an ecosystem.

3.3.1 Use of high capacity cloning vectors

Cloning of metagenomic DNA fragments was one of the first approaches used to access the genetic information of microorganisms that are recalcitrant to isolation. This approach is particularly attractive in cases in which the cloned fragments are predominantly large in size. While 'standard' laboratory cloning vectors are, on average, limited to the cloning of DNA fragments < 10 kb, so-called high capacity cloning vectors aid the transfer of DNA fragments in a size range of 35-45 kb (fosmids, cosmids) or even >100kb (e.g. *B*acterial *A*rtificial *C*hromosomes – BACs) into a heterologous host for replication. The particular charm of large fragments recovered by cloning is that their sequence analysis has the potential of providing insights into complete metabolic pathways and, in cases where they also harbour ribosomal RNA operons, offers a link of those metabolic functions to phylogenetic identification. One of the most prominent examples of the application of high capacity cloning vectors was the discovery of photosynthesis in bacteria *via* the light-driven proton pump proteorhodopsin (Béjà *et al.* 2000a). However, while this and other studies produced large insert BAC clones with an average insert size of e.g. 80 kb (e.g. Béjà *et al.* 2000b) such an outcome proves to be rather rare (pers. communication of former colleagues of the Author who unsuccessfully attempted to clone large DNA fragments). Given this caveat cloning smaller environmental DNA fragments into high capacity vectors, in particular into cosmids and fosmids, may become equally attractive (particularly since the use of high copy number fosmid vectors may provide the additional benefit of functional screening approaches *via* biochemical assays).

Proteorhodopsin-supported primary production was also detected in the genome of '*Candidatus* Pelagibacter ubique' HTCC1062 (Giovannoni *et al.* 2005a), a member of the SAR11 clade which consists of small heterotrophic marine *Alphaproteobacteria* and accounts for approximately 25% of total marine bacterioplankton (Morris *et al.* 2002; Mary *et al.* 2006). This together with its ability to proteorhodopsin-supported primary production positioned SAR11 as a major contributor to ocean energy budgets (Giovannoni *et al.* 2005a).

However, despite its abundance, clones containing 16S rRNA sequences of SAR11 bacteria have very rarely been detected in libraries of high capacity cloning vectors with total environmental DNA from SAR11 dominated sites (Suzuki *et al.* 2004; DeLong *et al.* 2006). This puzzling observation was also made in a study involving the Author. Fosmids were used as the cloning vector with the aim to obtain genomic information on the ubiquitous group of '*Candidatus* Pelagibacter ubique' (Gilbert *et al.* 2008). That is, PCR screening of a fosmid library of 10,000

clones constructed with surface water from the English Channel revealed only one clone (clone 01-003783) that contained a 16S rRNA gene with high sequence similarity to that of '*Candidatus* Pelagibacter ubique' (Gilbert *et al.* 2008). The low likelihood in cloning large AT-rich DNA was confirmed in a subsequent study, again involving the Author (Temperton *et al.* 2009). This repeated observation led the authors to suggest that reduced stability of AT-rich DNA, possibly due to lower hydrogen bonds weakening DNA against non-perpendicular shear forces, might provide an explanation for this (Temperton *et al.* 2009). More frequently occurring strand breakage of AT-rich DNA reduces the number of approx. 40-kb fragments required for fosmid vector insertion and, thus, lowering the representation of AT-rich DNA in fosmid libraries (Temperton *et al.* 2009). Although a number of other reasons have also been suggested to explain the observed bias (e.g. toxicity from expressed genes: Sorek *et al.* 2007; Feingersch and Béjà 2009), more recent research indicates that distinct mechanisms may add to the observed cause of the cloning bias in *E. coli* with transcription of cloned DNA possibly playing a particular role (Lam and Charles 2015).

Nevertheless, sequence analysis of the 32-kb fosmid clone 01-003783 that contained a 16S rRNA gene of '*Candidatus* Pelagibacter ubique' (Gilbert *et al.* 2008) confirmed the genetic plasticity in the genome of this ubiquitous bacterium (Rusch *et al.* 2007; Wilhelm *et al.* 2007). Apart from the ribosomal genes only four functional genes were among the total of 28 genes encoded by clone 01-003783 that showed high sequence similarity to SAR11 homologues. Half of the 22 other genes showed little sequence similarity to, while the other half did not even have representative homologues in the genome of '*Candidatus* Pelagibacter ubique' (Gilbert *et al.* 2008). It was therefore hypothesised (Gilbert *et al.* 2008) that the extreme abundance of '*Candidatus* Pelagibacter ubique' is intrinsically linked to hypervariable regions within the otherwise highly conserved and streamlined SAR11 genomes (Giovannoni *et al.* 2005b).

3.3.2 Use of high-throughput sequencing techniques without the need of cloning

The development of next generation sequencing techniques in combination with advanced computing tools for the analysis of sequence reads (e.g. assembly, annotation) not only provides, as mentioned above, a means to avoid biases caused by the application of the PCR, but also those caused by cloning of genomic DNA. As a result large fractions of the complete genomic information present in a sample can now be assessed. The Author has been involved in two studies that utilised this line of experimental approach (see also Gilbert *et al.* 2011).

An example from the marine environment

The aim of the first study (Gilbert *et al.* 2010a,b) was designed by the Author to test the hypothesis that bacteria exist in definable communities and that these communities are important for biogeochemical function. This was a follow-up of the NERC (UK) funded consortium project 'Aquatic Microbial Metagenomics' in which the construction of fosmid libraries, 16S rRNA sequencing, microarrays and stable isotope probing were employed to address the same hypothesis. However, since next generation sequencing techniques, at that time 454 pyrosequencing, had just become available an additional strategy was chosen to address the question as to which bacterial genes are present in the environment (metagenome analysis) and which of these genes are actually expressed in natural populations of marine bacteria (metatranscriptome analysis).

In order to test the above general hypothesis two specific objectives were designed. Firstly, previous analyses of long-term data collected at the sampling station L4 (50.2518° N, 4.2089° W; i.e. approx. 9 miles off the coast of Plymouth in the English Channel), part of the Western Channel Observatory (WCO, http://www.westernchannelobservatory.org.uk) in the English Channel had demonstrated the regular occurrence of blooms of various phytoplankton species throughout the year. The first objective therefore was to test the hypothesis that bacterial communities differ when the dominant phytoplankton species changes; that is, the occurrence of the different groups of the larger phytoplankton will influence the composition of the co-occurring bacterioplankton which is likely to utilise phytoplankton-derived organic matter. Additionally, the analysis of expressed genes should reveal why some species dominate with specific phytoplankton taxa. Secondly, the hypothesis should be tested as to whether day light impacts on the composition of marine bacterioplankton. Therefore, samples should also be taken from day and night time.

In order to approach these objectives pre-filtered (<1.6 μm) pelagic samples were collected in 2008 from station L4 under contrasting environmental conditions, in particular from different seasons and at different day time: during the winter season (at 2 pm and 10 pm on January 28th: low abundance of phytoplankton), in spring (at 2 pm and 10 pm on April 22nd : diatom bloom) and on two days in summer (at 4 pm and 10 pm on August 26th and at 4 pm and 10 pm of August 27th: bloom of coccolithophores, in particular of *Emiliania huxleyi*). (Due to contractual reasons it was not possible to sample the dinoflagellate bloom that has been found to regularly occur in September.) In addition to metagenomic and metranscriptomic sequence analyses bacterial 16S-

tag pyrosequencing (Sogin *et al.* 2006) was also performed in order to compare the taxonomic profile of the microbial populations.

The results from this microbial 'multi-omic' study combining 16S rRNA amplicon sequencing with metagenomic and metatranscriptomic profiling can be summarised in five main conclusions (Gilbert *et al.* 2010a,b): ***i)*** archaea follow the same seasonal patterns as bacteria, but show lower relative diversity; ***ii)*** higher 16S rRNA sequence diversity also reflects a higher diversity of gene transcripts; ***iii)*** diversity is highest in winter and at night, a result lateron confirmed by a six-year time series of 16S-tag data from the WCO (Gilbert *et al.* 2012); ***iv)*** community-level changes in 16S-based diversity and metagenomic profiles are best explained by seasonal patterns (with samples closest in time being most similar), while metatranscriptomic profiles are best explained by diel patterns and shifts in particular categories (i.e., functional groups) of genes; ***v)*** changes in key genes are seasonal and diel in nature (i.e., photosynthesis); but these samples contain large numbers of orphan genes without known homologues and it is these unknown gene sets that appear to contribute most towards defining the differences observed between times.

An example from a mining related acidic environment

The second study was designed by the Author to use metagenomics in order to circumvent the problems associated with the isolation of acidophilic *Gallionella*-like strains from the pilot plant Tzschelln; this plant uses microbial iron oxidation for the bioremediation of AMD (see section 2.4.2). Based on their low 16S rRNA sequence similarity (≤94%) to other *Betaproteobacteria*, these strains most likely represent a novel genus, most closely related to *Gallionella ferruginea* (Heinzel *et al.* 2009a,b). Additionally, these strains were found to form the second most abundant group within the microbial community of the pilot plant (Heinzel *et al.* 2009b). In this context, it should however be noted that the information which supports the presumed abundance of *Gallionella*-like acidophiles in the pilot plant Tzschelln is based almost exclusively on TRFLP analyses of samples throughout an annual cycle (Heinzel *et al.* 2009b). The three samples that were analysed by real-time qPCR did, however, not confirm the results from the TRFLP analyses in all cases (Heinzel *et al.* 2009b). Additionally, the identification of this abundant group of strains as belonging to a *Gallionella*-like taxon is solely based on the recovery and phylogenetic analysis of only two clones (W-Z-164: EU360498; clone FS_Z2_79: EU360492) from a 16S rRNA gene fragment library (Heinzel *et al.* 2009a).

The availability of genomic information of these acidophilic *Gallionella*-like strains may allow conclusions on their physiological and biochemical requirements for growth and activity. This,

in turn, may reveal parameters useful as a means to monitor and improve the performance of the pilot plant.

The results from this study will be presented in more detail in the following since these have not yet been published. Additional details can be retrieved from Voitel (2015).

[The microaerophilic culture ADE-12-1 was established by M.Sc. Anna Drechsel. The analysis of the bacterial diversity within enrichment culture ADE-12-1 and the subsequent assignment of metagenomic contigs was undertaken by B.Sc. Matthias Voitel in the framework of his Bachelor thesis. Both were supervised by the Author who also planned and designed the project.]

The experimental approach taken for this investigation was similar to that employed for the analysis of "*Ferrovum*" strains (section 2.4.2), though the enrichment of the target microorganisms resulted in a far more complex assemblage.

Enrichment of microaerophilic and acidophilic iron oxidising microorganisms was achieved using gradient tubes of semi-solid APPW (plus additional phosphate; see section 2.4.2) and incubation in microaerobic chambers (2.5 L Anaerojar with Campygen pads, OXOID; see section 3.2). This approach was based on the assumption that the acidophilic *Gallionella*-like strains are physiologically similar to neutrophilic *Gallionella* which have long been known to only occur under ferrous iron rich and oxygen limiting conditions (Engel and Hanert 1967). Gradient tubes originally developed by Kucera and Wolfe (1957) were produced by encapsulating iron sulfide (prepared according to Emerson and Floyd 2005; see section 3.2) within agarose or gellan gum (0.5% w/v) at the bottom of a glass tube with a semisolid (0.15% (w/v) agarose) layer of APPW medium (pH 3.5) atop (see Fig. 29). Tests showed that this setup led to better results than those using, for instance, iron carbonate as source of ferrous iron (unpubl. results). Additionally, a semisolid layer proved also to be superior for the isolation of microaerophilic enrichment cultures in comparison to a liquid layer of APPW medium, as has been suggested by Hallbeck *et al.* (1993).

Extraction of total environmental nucleic acids from the microaerophilic enrichment culture (Fig. 29B) provided 1.2 µg of DNA which were submitted to sequence analysis using the Illumina approach (2 x 112 bp paired end run; performed at the G2L). Assembly and annotation of the individual sequence reads resulted in a total of 36 Mbp of unique sequence on a total of 7031 contigs. Only 17 of these contigs contained a 16S rRNA gene or gene fragment, with only one instance in which the 16S rRNA sequence was part of a large contig (contig_002851 assigned to *Sediminibacterium salmoneum* based on the 16S rRNA gene sequence). In most cases the 16S rRNA sequences were individual sequence reads or within contigs smaller than

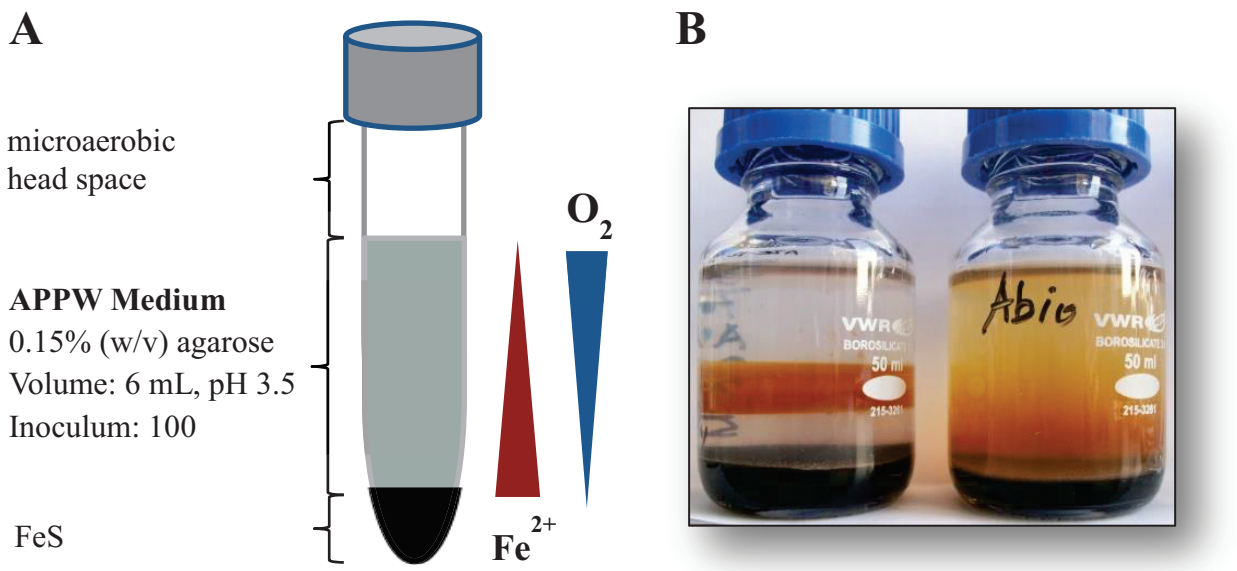


Fig. 29: Basics of gradient tubes for the culture of microaerophilic acidophilic microorganisms (A; similar to that used for the enrichment of neutrophilic *Gallionella* except for medium composition and pH: see Fig. 27A) and enrichment culture ADE-12-1 of microaerophilic iron oxidising microorganisms in 50-mL scale (B). A sample of the inflow of AMD water into the pilot plant Tzschelln (see section 2.4.2) was used as inoculum. The orange-brown ring in the left bottle in (B) is formed at the presumed optimal oxygen (and ferrous iron) condition for the microaerophilic iron oxidisers enriched in this experiment. The right bottle shows the abiotic control that was not inoculated. Highest ferric iron precipitation occurs there near the bottom of the bottle (i.e. the source of ferrous iron). The photograph was taken after 35 days of incubation at room temperature (courtesy of Anna Drechsel.)

ca. 30 kb. This is not surprising since ribosomal gene clusters often occur in multiple copies within genomes which, in turn, affects the assembly of the raw data. Other ribosomal rRNA genes which are part of larger contigs were too short to permit assignment to specific taxa within the *Betaproteobacteria* (e.g. 5S rRNA gene). This necessitated employment of alternative marker genes to support taxonomic assignment of contigs. The absence of extensive sequence databases of such alternative markers, however, meant that the assignment would be limited to taxa whose genome sequences are publicly available. Therefore, a three-tier 16S rRNA gene fragment based approach was chosen to obtain a robust identification of the bacterial taxa

present in the enrichment culture ADE-12-1 for subsequent selection of reference genomes. This approach comprised, apart from the analysis of the 17 sequences of 16S rRNA gene fragments within the metagenomic dataset (first tier) also a 16S-tag analysis using Illumina sequencing (second tier) and the preparation and (Sanger) sequence analysis of a library of PCR amplified 16S rRNA gene fragments (third tier). The latter approach was necessary since the 16S-tag analysis rarely results in taxonomic assignment below genus level, though it permits the detection of even those taxa that occur at low abundance due to the high sequencing effort typically inherent to the methodology. Both PCR based approaches (second and third tier) used as DNA template the same PCR amplicons obtained with primers 27f (5'-AGAGTTTGATCCTGGCTCAG) and 1387r (5'-GGGCGG(AT)GTGTACAAGGC). That is, the results of the approaches should be comparable, though subsequent experimental steps may have led to some degree of bias (i.e. cloning and index PCR in the case of the preparation of the clone library and the Nextera library for Illumina sequence analysis, respectively).

The results (Table 13) from the sequence comparison of the 17 16S rRNA genes or gene fragments from the metagenomic dataset revealed the presence of eight genera (*Sideroxydans, Telmatospirillum, Cellulomonas, Sulfuritalea, Sediminibacterium, Thiomonas, Methylotenera, Opitutus*) from four phyla (*Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia*). However, the 16S rRNA gene fragments of neither neutrophilic *Gallionella* nor acidophilic *Gallionella*-like strains (Heinzel *et al.* 2009a) were closest (BLAST) hits to any of the 17 16S rRNA sequences. However; three of the 16S rRNA gene fragments showed highest sequence similarity to *Sideroxydans lithotrophicus* strain ES-2 (Table 13). Moreover, the fact that these 16S rRNA sequences differed substantially (93%, 96%, 98%) indicates that the enrichment culture contains three different acidophilic (or at least acidotolerant) strains (or species) of *Sideroxydans* or *Sideroxydans*-like genera.

This notion was further supported by the results from the analysis of both the 16S rRNA clone library and the 16S-tag Illumina reads (Fig. 30). That is, none of the sequences showed highest similarity to the 16S rRNA sequences of either *Gallionella* or acidophilic *Gallionella*-like strains (Heinzel *et al.* 2009a). Two or three groups of *Sideroxydans* or *Sideroxydans*-like strains formed in its entirety either the most abundant (in the case of the 16S rRNA gene fragment library: representing 69% of the clones screened clones) or one of the most abundant taxa (in the case of the 16S-tag analysis: with 29% as abundant as *Telmatospirillum* and only slightly more abundant than *Opitutus*: ca. 21%) within the enrichment culture. (The reader is referred to Voitel (2015) for further details of the results and methodology used for the preparation and screening of the 16S rRNA clone library and the analysis of the sequence data.)

Table 13: Closest blastn hits of the 17 sequences of 16S rRNA gene fragments from the metagenome dataset to isolated strains and strains with sequenced genomes. The E-value for the hits was zero in all cases. Accession numbers of the hits can be retrieved from Voitel (2015).

Contig no.	Closest match to isolated representative	% similarity	Closest match to isolated representative isolate with genome sequence available	% similarity
contig000399	*Thiomonas intermedia* strain K12	96%	*Thiomonas intermedia* strain K12	96%
contig000631	*Sideroxydans lithotrophicus* strain ES-1	98%	*Sideroxydans lithotrophicus* strain ES-1	98%
contig000768	*Telmatospirillum siberiense* strain 26-4b1	96%	*Magnetospirillum gryphiswaldense* strain MSR-1 v2	91%
contig001087	*Cellulomonas cellasea* strain DSM	98%	*Cellulomonas fimi* strain ATCC 484	97%
contig001607	*Sulfuritalea hydrogenivorans* sk43H	96%	*Sulfuritalea hydrogenivorans* sk43H	96%
contig001717	*Thiomonas cuprina* strain Hoe5	97%	*Thiomonas intermedia* strain K12	94%
contig001988	*Telmatospirillum siberiense* strain 26-4b1	96%	*Magnetospirillum gryphiswaldense* strain MSR-1 v2	91%
contig002544	*Sulfuritalea hydrogenivorans* sk43H	96%	*Sulfuritalea hydrogenivorans* sk43H	96%
contig002851	*Sediminibacterium salmoneum* NBRC 103935	95%	*Rubrivivax gelatinosus* strain IL144	95%
contig003822	*Sideroxydans lithotrophicus* strain ES-1	93%	*Sideroxydans lithotrophicus* strain ES-1	93%
contig004326	*Thiomonas cuprina* strain Hoe5	99%	*Thiomonas intermedia* strain K12	95%
contig004349	*Telmatospirillum siberiense* strain 26-4b1	96%	*Magnetospirillum gryphiswaldense* strain MSR-1 v2	91%
contig004460	*Sideroxydans lithotrophicus* strain ES-1	96%	*Sideroxydans lithotrophicus* strain ES-1	96%
contig004966	*Methylotenera mobilis* strain JLW8	96%	*Methylotenera mobilis* strain JLW8	96%
contig005444	*Methylotenera mobilis* strain JLW8	96%	*Methylotenera mobilis* strain JLW8	96%
contig006346	*Opitutus terrae* strain PB90-1	92%	*Opitutus terrae* strain PB90-1	92%
contig006588	*Telmatospirillum siberiense* strain 26-4b1	95%	*Magnetospirillum gryphiswaldense* strain MSR-1 v2	91%

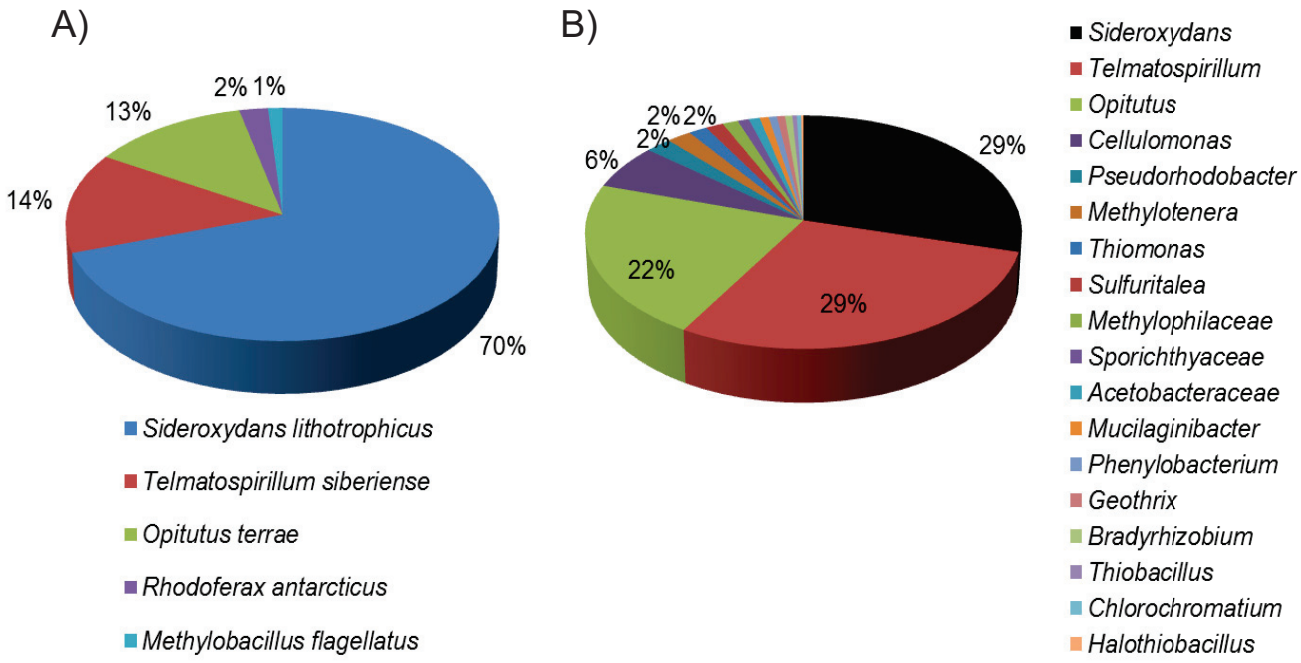


Fig. 30: Summary of the bacterial diversity detected by sequence analysis of PCR amplified 16S rRNA gene fragments. Sanger sequence analysis was conducted on cloned 16S gene fragments (A), while Illumina sequencing was applied to the analysis of the 16S-tags (B).

Further evidence for the taxonomic assignment of the 16S rRNA sequences from the metagenomic dataset and the clone library is provided in form of the phylogenetic analysis (Fig. 31) which also supports the concept of the presence of two to three different *Sideroxydans* and *Sideroxydans*-like groups of strains.

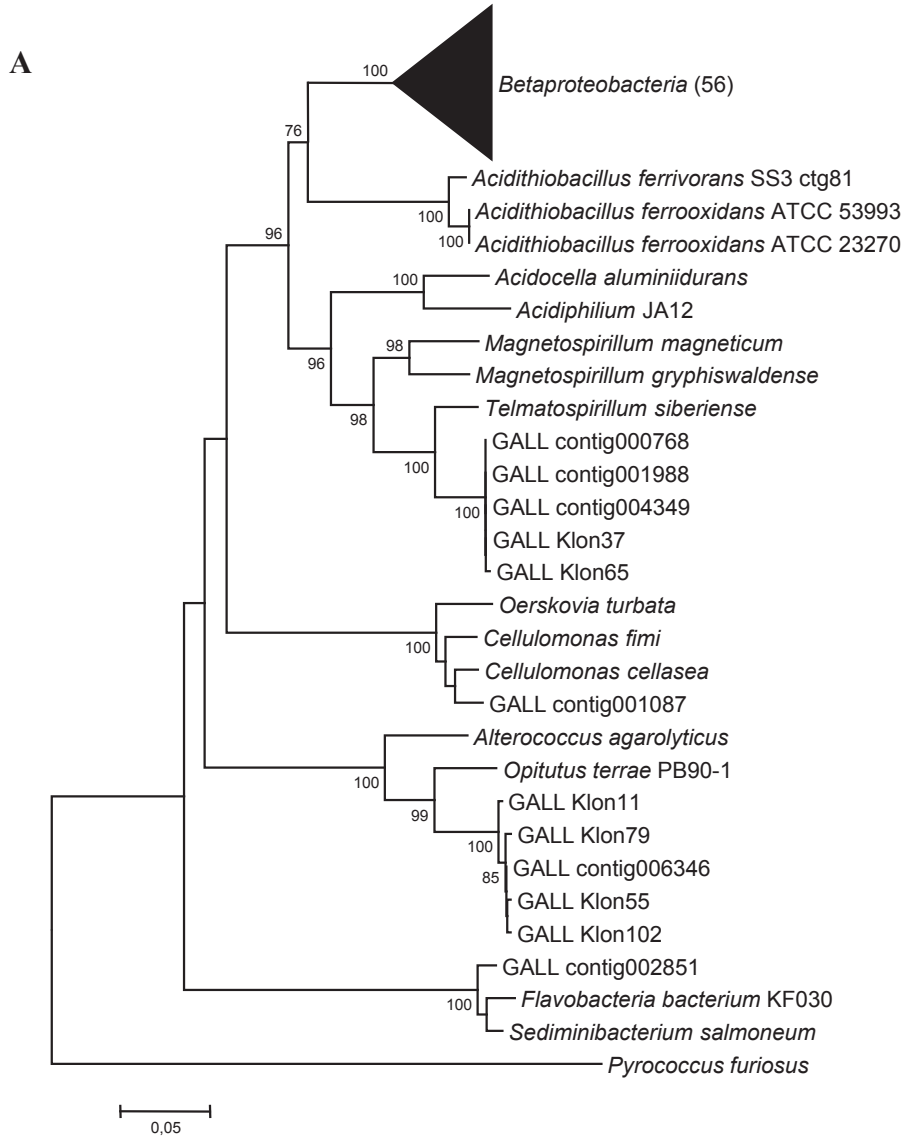
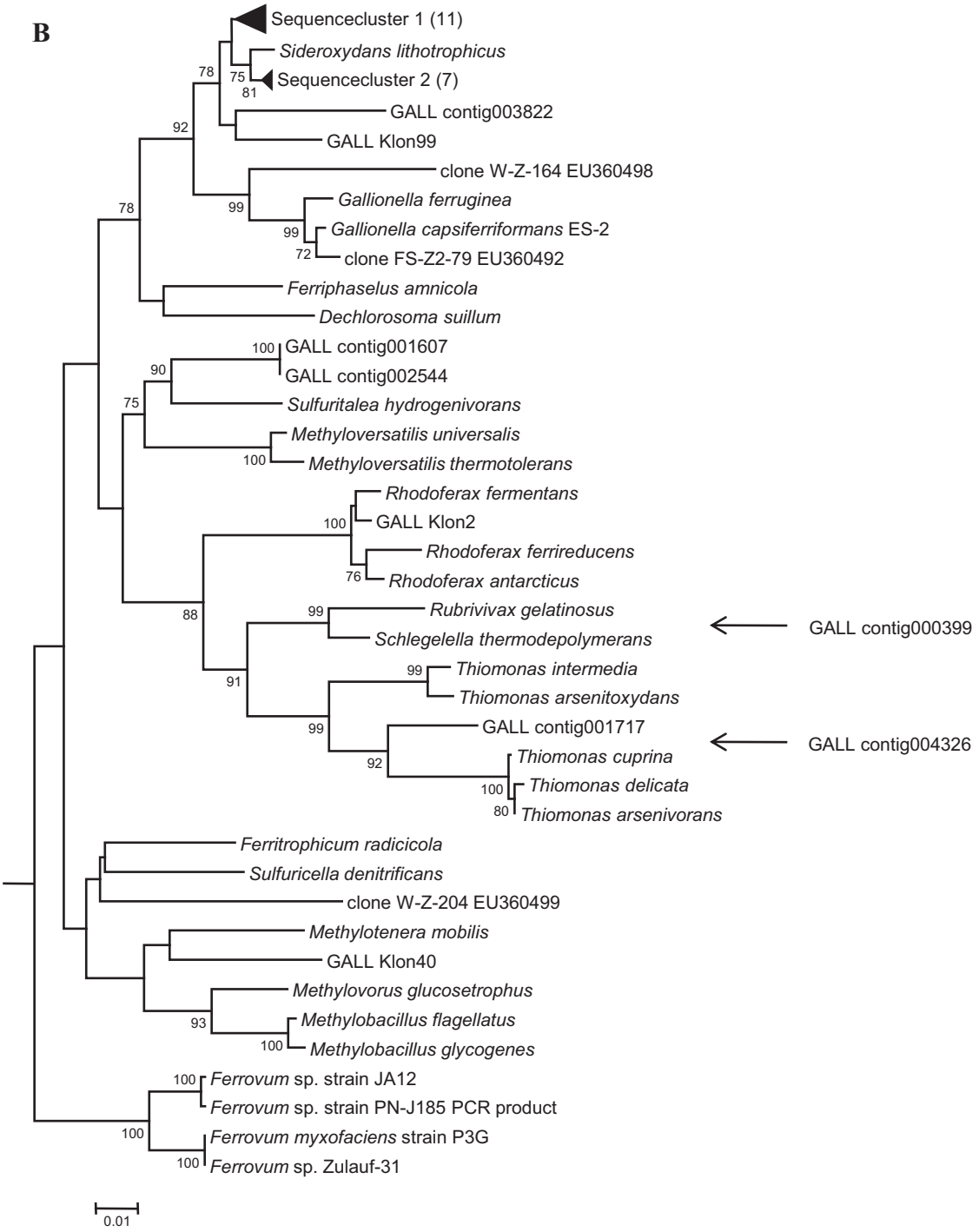
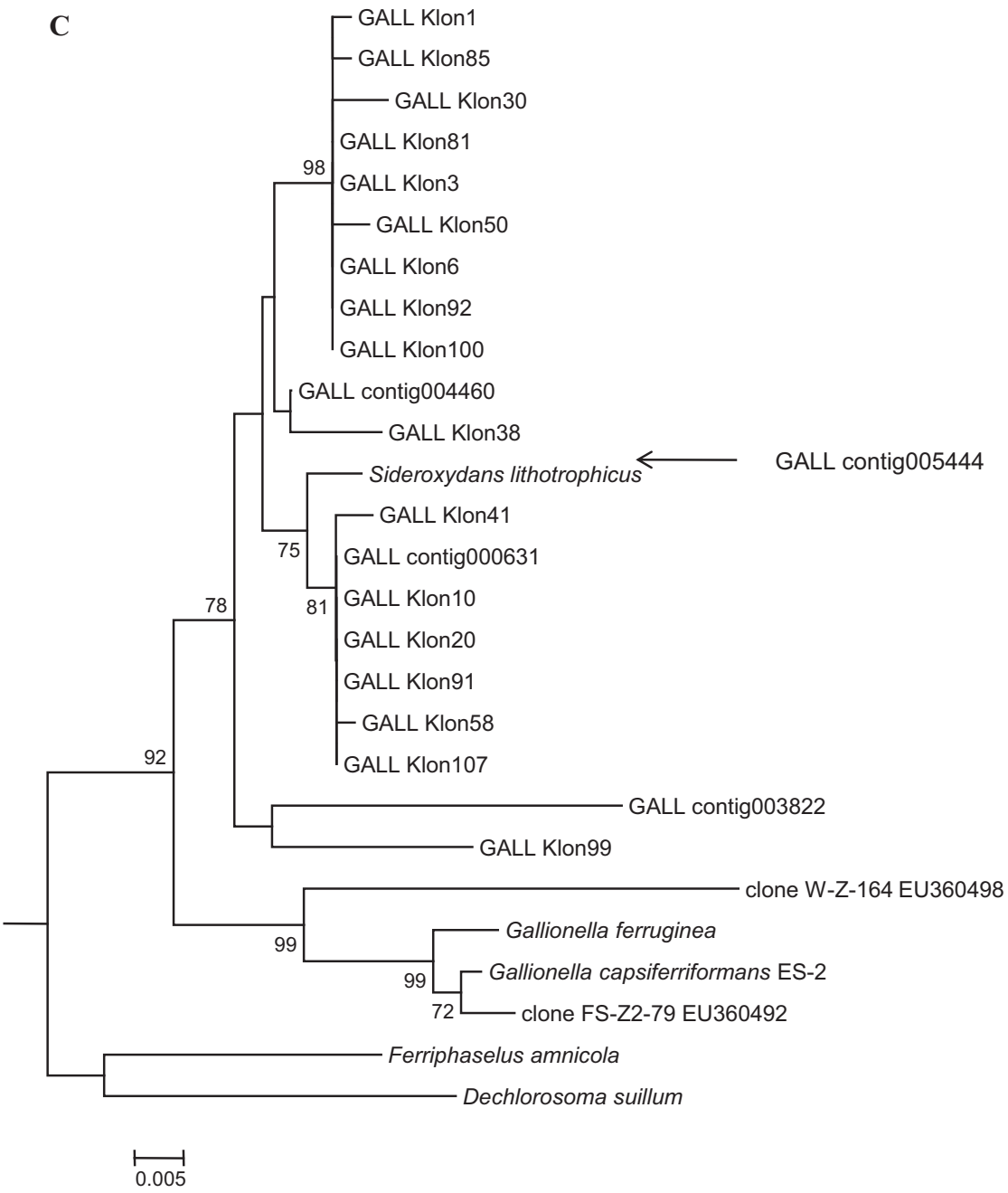


Fig. 31: Phylogenetic analysis of The evolutionary history was inferred using the Neighbor-Joining method based on the Jukes-Cantor model (Jukes and Cantor 1969) and an alignment of a 84 sequences of 16S rRNA gene fragments produced within ARB (Ludwig *et al.* 2004). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. There were a total of 612 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.* 2013). The 56 *Betaproteobacteria* (A) are dissected in (B) and (C). The latter provides a detailed view on the relationship of the *Gallionellaceae*. The position of three of the 17 16S rRNA gene fragments from the metagenomic dataset (see B and C) was determined separately due to the short overlapping region (220 – 446 bp) with the other sequences. This was achieved by adding them to an identical neighbor-joining tree using the parsimony option within ARB.


B
Sequencecluster 1 (11)
Sideroxydans lithotrophicus
Sequencecluster 2 (7)
GALL contig003822
GALL Klon99
clone W-Z-164 EU360498
Gallionella ferruginea
Gallionella capsiferriformans ES-2
clone FS-Z2-79 EU360492
Ferriphaselus amnicola
Dechlorosoma suillum
GALL contig001607
GALL contig002544
Sulfuritalea hydrogenivorans
Methyloversatilis universalis
Methyloversatilis thermotolerans
Rhodoferax fermentans
GALL Klon2
Rhodoferax ferrireducens
Rhodoferax antarcticus
Rubrivivax gelatinosus
Schlegelella thermodepolymerans
GALL contig000399
Thiomonas intermedia
Thiomonas arsenitoxydans
GALL contig001717
GALL contig004326
Thiomonas cuprina
Thiomonas delicata
Thiomonas arsenivorans
Ferritrophicum radicicola
Sulfuricella denitrificans
clone W-Z-204 EU360499
Methylotenera mobilis
GALL Klon40
Methylovorus glucosetrophus
Methylobacillus flagellatus
Methylobacillus glycogenes
Ferrovum sp. strain JA12
Ferrovum sp. strain PN-J185 PCR product
Ferrovum myxofaciens strain P3G
Ferrovum sp. Zulauf-31
0.01


C
GALL Klon1
GALL Klon85
GALL Klon30
GALL Klon81
98
GALL Klon3
GALL Klon50
GALL Klon6
GALL Klon92
GALL Klon100
GALL contig004460
GALL Klon38
GALL contig005444
Sideroxydans lithotrophicus
GALL Klon41
75
GALL contig000631
78
81
GALL Klon10
GALL Klon20
GALL Klon91
GALL Klon58
92
GALL Klon107
GALL contig003822
GALL Klon99
clone W-Z-164 EU360498
Gallionella ferruginea
99
Gallionella capsiferriformans ES-2
99
72
clone FS-Z2-79 EU360492
Ferriphaselus amnicola
Dechlorosoma suillum
0.005

Based on the knowledge of the bacterial diversity within the enrichment culture a set of 17 genomes was selected as the closest representative taxa with publicly available genome sequences. A list of single copy genes either encoding ribosomal proteins or house keeping genes, including some of those suggested by Raes *et al.* (2007), were used as alternative markers to determine the taxonomic identity of the strains from which the individual contigs derived. The approach taken for this was based on the calculation of the phylogenetic relationship based on amino acid alignments of contig-derived marker sequences and those from each of the 17 reference genomes. Details on the resulting dendrograms can be found by Voitel (2015). The relevant conclusions from these analyses are summarised in Table 14.

Table 14: Summary of the bacterial taxa from which the individual contigs were derived based on sequence comparison of marker genes encoded on these contigs with homologues from the 17 bacterial reference genomes. Marker genes used were those described by Raes *et al.* (2007) plus further single copy genes encoding cellular housekeeping functions (see Voitel 2015 for details). Sequence comparison was achieved by calculating dendrograms for each marker gene using the neighbor-joining method (see Voitel 2015 for details). This analysis has only been conducted for contigs larger than 100 kb and seven smaller contigs since these were found to encode functions of particular interest (e.g. urease, H^+/Na^+-antiporter etc.).

Contig	contig size [kb]	marker genes tested	taxonomic assignment
contig000124	118	1	*Sediminibacterium*
contig000135	100	2	*Opitutus terrae*
contig000173	107	1	*Sideroxydans lithotrophicus*
contig000587	115	1	*Opitutus terrae*
contig000597	109	1	*Opitutus terrae*
contig000625	173	2	*Sideroxydans lithotrophicus*
contig000757	274	3	*Sediminibacterium*
contig000977	166	1	*Sideroxydans lithotrophicus*
contig001031	106	2	*Opitutus terrae*
contig001288	149	1	*Methylotenera mobilis*
contig001341	108	1	*Methylotenera mobilis*
contig001403	270	1	*Magnetospirillum gryphiswaldense*
contig001550	84	1	*Sediminibacterium*
contig001784	123	1	*Methylotenera mobilis*
contig001871	298	1	*Sideroxydans lithotrophicus*

Contig	contig size [kb]	marker genes tested	taxonomic assignment
contig001878	108	1	*Opitutus terrae*
contig002429	136	1	*Opitutus terrae*
contig002578	207	2	*Sediminibacterium*
contig002635	113	1	*Methylotenera mobilis*
contig002650	195	1	*Sideroxydans lithotrophicus*
contig002668	105	2	*Methylotenera mobilis*
contig002696	152	1	*Sideroxydans lithotrophicus*
contig002717	103	3	*Opitutus terrae*
contig002737	149	2	*Sideroxydans lithotrophicus*
contig002774	125	1	*Sediminibacterium*
contig002785	186	5	3x *Methylotenera mobilis,* 1x *Sideroxydans lithotrophicus, 1x Rhodoferax ferrireducens*
contig002815	150	1	*Methylotenera mobilis*
contig002817	312	3	*Opitutus terrae*
contig002843	212	2	*Opitutus terrae*
contig002851	156	1	*Sediminibacterium*
contig002876	12	2	*Sideroxydans lithotrophicus*
contig002985	42	2	*Sideroxydans lithotrophicus*
contig003056	55	2	*Methylotenera mobilis*
contig003168	192	2	*Opitutus terrae*
contig003202	171	2	*Sideroxydans lithotrophicus*
contig003459	103	1	*Sideroxydans lithotrophicus*
contig003549	130	1	*Sediminibacterium*
contig003665	146	2	*Opitutus terrae*
contig003750	138	1	*Sediminibacterium*
contig003766	223	2	*Methylotenera mobilis*
contig003788	48	3	*Sideroxydans lithotrophicus*
contig003979	112	1	*Sideroxydans lithotrophicus*
contig004062	156	1	*Sediminibacterium*
contig004270	202	2	*Sediminibacterium*
contig004599	430	1	*Methylotenera mobilis*
contig005689	146	1	*Sediminibacterium*
contig005805	128	1	*Sideroxydans lithotrophicus*
contig005997	164	1	*Sediminibacterium*
contig006137	246	2	*Opitutus terrae*
contig006157	139	1	*Sideroxydans lithotrophicus*
contig006248	52	1	*Sideroxydans lithotrophicus*
contig006268	138	1	*Opitutus terrae*

Contig	contig size [kb]	marker genes tested	taxonomic assignment
contig006346	159	1	*Opitutus terrae*
contig006503	136	1	*Opitutus terrae*
contig006601	129	1	*Sideroxydans lithotrophicus*
contig006742	177	2	*Sideroxydans lithotrophicus*
contig006788	105	1	*Sideroxydans lithotrophicus*
contig006902	7	1	*Magnetospirillum gryphiswaldense*
contig006920	123	1	*Sediminibacterium*

Most of the 59 contigs were assigned to *S. lithotrophicus* (19) followed by *Opitutus terrae* (15), *Sediminibacterium* (13) and *Methylotenera mobilis* (10). Only two contigs were found to derive from *Magnetospirillum gryphiswaldense*, which was the closest relative to *Telmatospirillum siberiense* with a publicly available genome sequence (Table 13). This result stands in contrast to the fact that four of the 17 16S rRNA genes or gene fragments detected within the metagenomic sequence information seems to derive from this species. Moreover, sequence analysis of PCR amplified 16S rRNA gene fragments also indicates that 14% (clone library) or 29% (16S-tag Illumina sequencing) of the amplicons derive from *Telmatospirillum siberiense* (Fig. 30). Similarily, none of the contigs larger than 100 kb were found to have derived from members of the genus *Thiomonas* which three of the 17 16S rRNA genes were assigned to (Table 14) and which was also detected by 16S-tag Illumina sequencing (Fig. 30). Bias introduced by PCR amplification of 16S rRNA gene fragments seems likely to cause the observed differences between the metagenomic data and those from the clone library and 16S-tag sequencing. However, the reasons for the detected differences between the number of 16S rRNA gene fragments in the assembled metagenomic dataset and the distribution of the 59 contigs screened in this study is less obvious, though is likely to be related to the approach employed for the assembly of the sequence data.

Overall, the cumulative size of the contigs that were assigned to the specific taxa amounts to between 6% and 80% of the size of the five reference genomes (Table 15). This calculation does, however, not take into account the fact that the contigs might need to be assigned to more than one strain or species of the reference taxon (as is likely the case with *S. lithotrophicus*).

Table 15: Summary of the genome sizes and accession numbers of the five reference taxa which metagenomic contigs were assigned to and the total size of the various contigs. Contig002785 was assigned to *Methylotenera mobilis* (see Table 14).

	sum of individual contigs assigned to specific taxa	**genome size**	**accession number**
Sideroxydans lithotrophicus	2.46 Mb	3.00 Mb	CP001965.1
Sediminibacterium	2.02 Mb	3.25 Mb	NZ_AXZP00000000.1
Opitutus terrae	2.32 Mb	5.96 Mb	CP001032.1
Methylotenera mobilis	1.46 Mb	2.55 Mb	CP001672.1
Magnetospirillum gryphiswaldense	0.28 Mb	4.37 Mb	HG794546.1

Comparison with previous studies

The comparison of the findings from the presented investigation with those from previous studies on samples from the same environmental habitat raises the question as to the reasons for the observed discrepancy; that is, the abundance of *Sideroxydans* or *Sideroxydans*-like strains versus that of *Gallionella*-like strains reported previously (Heinzel *et al.* 2009a, b). Clearly, a definite answer can not be provided since the original samples are no longer available for renewed analysis using techniques that are now available and less affected by experimental bias (see below). However, a number of possible reasons shall be provided to support the argument that *Sideroxydans* is likely to not only be abundant in the microaerophilic enrichment culture ADE-12-1, but also within the AMD water sampled at the pilot plant Tzschelln. For a start, the samples tested in previous studies (Heinzel *et al.* 2009a,b) and that used for the enrichment in the presented study (Voitel 2015) were sampled at different times (2007 versus 2014, respectively). So, the bacterial composition within the AMD water may, therefore, differ as observed. There are, however, also issues related to the methodology used for the identification and quantification of the abundant taxa (ARDRA analysis of clones from a library of 16S rRNA gene fragments: Heinzel *et al.* 2009a; TRFLP and real-time quantitative PCR: Heinzel *et al.* 2009b). Detailed analysis of those methods demonstrated that neither the ARDRA analysis used for the reduction of redundancy within the clone library nor the TRFLP approach proved to be adequate to distinguish between sequences of 16S rRNA gene fragments of *Gallionella* and

Gallionella-like strains and those from *Sideroxydans* or *Sideroxydans*-like strains (for details see Voitel 2015). In contrast to this, the real-time qPCR approach withstands this scruntiny, in particular due to a mismatch between the 3'-end base of reverse primer 384r (Heinzel *et al.* 2009b) and its corresponding binding site within *Sideroxydans* and *Sideroxydans*-like strains (Voitel 2015). The discrepancy between the results obtained by the real-time qPCR (specific for *Gallionella* and *Gallionella*-like strains) and TRFLP (both genera are indistinguishable) analyses may, therefore, explain the possible detection of *Sideroxydans* or *Sideroxydans*-like strains in the samples from the pilot plant by one (TRFLP)b but not the other (real-time qPCR) methodological approach (see Heinzel *et al.* 2009b; Voitel 2015).

pH homeostasis

Apart from the recognition that *Sideroxydans* may be another abundant species within the pilot plant Tzschelln the metagenomic study, though based on a relatively low sequencing effort (ca. 36 Mb), also provides first insights into the metabolic potential of acidophilic *Sideroxydans* strains. Although analyses in this direction are currently still underway, first results of the comparison of the metabolic potential encoded by the genome of the neutrophilic *S. lithotrophicus* to that encoded by the metagenomic contigs assigned to *Sideroxydans* or *Sideroxydans*-like strains indicate three acidophile specific mechanisms to maintain intracellular pH homeostasis. These include, firstly, a K^+-transporting ATPase of the Kdp-type to support the production of a reversed (inside positive) membrane potential (Baker-Austin and Dopson 2007). A second example includes a urease encoding gene cluster within a metagenomic contig assigned to *Sideroxydans* (GALL_contig006248). Similar to the mechanism discussed above for "*Ferrovum*" strains belonging to group 2 (section 2.4.2), urease may also play a role in the ability of acidophilic *Sideroxydans* or *Sideroxydans*-like strains to live at acidic pH.

The third phenotypic feature possibly involved in the adaptation to acidic environmental conditions is based on the presence of cyanophycin synthetase (CphA1, CphA2) encoding genes within a contig assigned to an acidophilic *Sideroxydans* or *Sideroxydans*-like strain. Therefore, it is likely that this strain also produces cyanophycin, a polymer consisting of the two amino acids aspartic acid and arginine (Simon and Weathers 1976). Although long thought to be specific for *Cyanobacteria* targeted analysis of genome sequences revealed their presence in a wide range of eubacteria outside the *Cyanobacteria* (Krehenbrink *et al.* 2002), including *Betaproteobacteria* such as *G. capsiferiformans* strain ES-2. There it may serve as nitrogen storage that is produced during, for instance, phosphor limitation. The presence of two alleles (*cphA1*, *cphA2*) of the cyanophycin synthetase encoding genes

may furnish *Sideroxydans*-like strains with functional redundancy (e.g. the ability to respond to various environmental triggers). However, the presence of this physiological trait in an acidophilic (GALL_contig_000625), but not a neutrophilic representative (*S. lithotrophicus* strain ES-1) of the same taxonomic group may indicate acquisition of the trait and its subsequent maintenance to fulfil (also) a role in pH homeostasis: decarboxylation of aspartic acid and arginine moietes of cyanophycin to provide an effective means of buffering intracellular proton levels. Decarboxylation of amino acids has already been reported as a strategy in pH homeostasis (Castanie-Cornet *et al.* 1999; Baker-Austin and Dopson 2007). However, the presence of large amounts of cyanophycin would provide the cell with a particularly extensive buffer capacity against high proton concentrations. In this context, it is interesting to speculate on whether cyanophycin occurs in inclusions or as soluble polymer within the cell (Frommeyer and Steinbüchel 2013) since this is likely to affect its availability (i.e. particularly high in the case of the soluble form) as instant buffer against acidity. Additionally, cyanophycin has been reported to contribute to heavy metal resistance (Wood, 1983).

Utilisation of cyanophycin requires cyanophycinase (CphB) activity which is neither encoded on the same contig (GALL_contig000625) that harbours the cyanophycin synthetase nor on any other contig assigned to *Sideroxydans* (Table 14). However, the fact that genes coding for this function have also not been detected within genome sequences available from any of the other *Betaproteobacteria* indicates the presence of an alternative path within this phylogenetic group (Krehenbrink *et al.* 2002). (The best hits from a blastp search of a cyanophycinase from *Nostoc punctiforme* against all betaproteobacterial genome sequences within the NCBI database on 7 January 2016 did again not reveal any significant hits: e-values $\geq e^{-10}$ over less than the full length of the query sequence.)

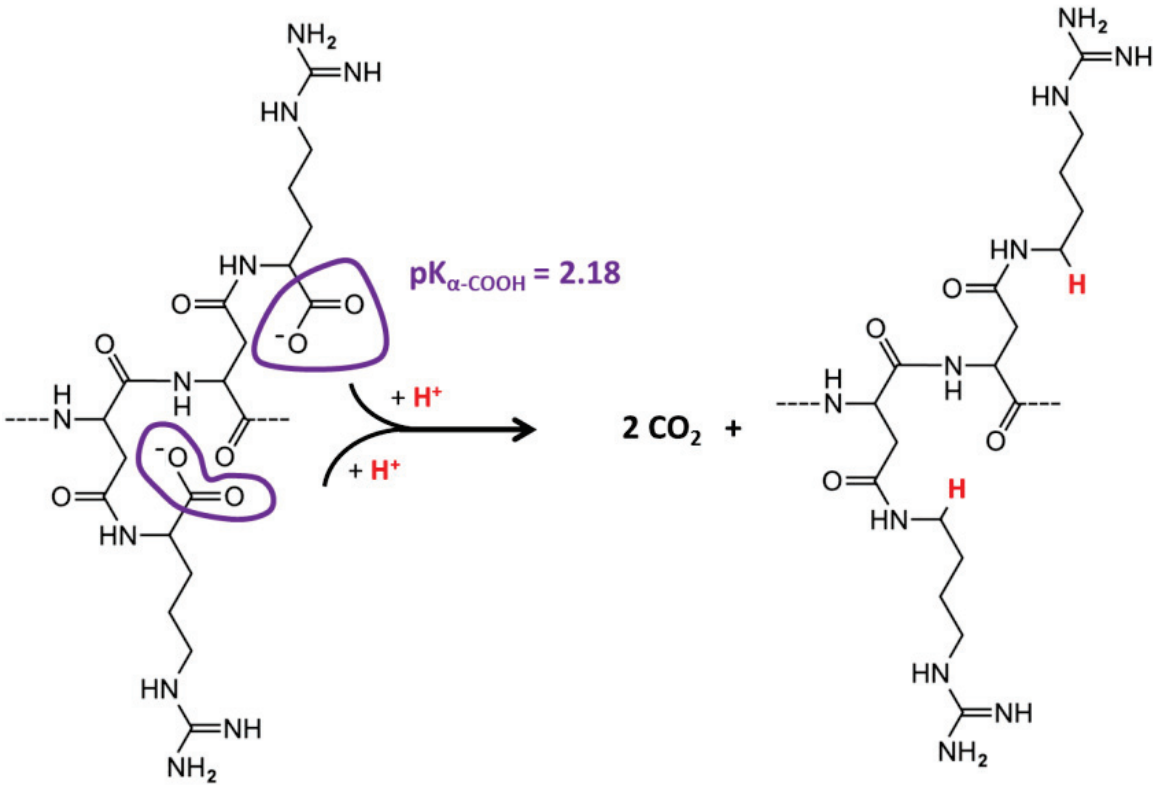


Fig. 32: Cyanophycin decarboxylation as a potential mechanism to maintain pH homeostasis. The enzymes required for cyanophycin synthesis (cyanophycin synthetase) and for argine decarboxylation (arginine decarboxylase) are encoded on two metagenome contigs that were assigned to *Sideroxydans* (contigs GALL_contig_000625 and GALL_contig_001871, respectively).

Other traits that are encoded by both the neutrophilic *S. lithotrophicus* strain ES-1 and the acidophilic metagenomic contigs assigned to *Sideroxydans*-like strains are likely to be also involved in the survival under high extracellular proton levels or in dealing with the consequences of increased proton flux into the cell. Examples of these include Na^+/H^+ antiporters (NhaP) to reduce intracellular proton levels, chaperones to maintain protein stability, exodeoxyribonuclease III and DNA mismatch repair protein (MutL/S) for DNA repair, superoxide dismutase and thiol:disulfide interchange protein DsbA/B to counter oxidative stress and cation efflux system protein CusB for the export of copper ions. Further and more detailed analyses will be conducted following additional sequencing efforts of the metagenomic DNA preparation by the G2L in order to increase sequence information, also on other members of the enrichment cultures.

4. On current and future developments in microbial ecology

Microbial ecologists have made astonishing discoveries over the last five decades which were partly driven by new technologies. For the period of the last 10-15 years, high-throughput DNA (and peptide) sequencing have had the most important impact. It appears likely that, in the absence of these new techniques, some of these discoveries would have not yet been made (or might even never be made) in hypothesis driven research projects. This raises the question as to the rationality of continuing to apply 'classical' approaches in microbial ecology and regarding current limitations in microbial ecology and means to overcome those. Clearly, since such discussions represent an almost boundless playground the Author needed to restrict himself to name but a few aspects. These are selected based on the relevance to the Author's own research.

Does the importance of classical approaches in microbial ecology fade?

Culture independent molecular tools to identify microbial taxa within environmental samples have been used since the hijacking of the PCR by microbial ecologists for the amplification and subsequent sequence analysis of 16S rRNA gene fragments (see section 3). A steady criticism of this methodological approach has been that these studies were often purely descriptive in nature (i.e. 'stamp collecting'). Nevertheless, probably also due to the invention of a variety of novel genetic fingerprint techniques and due to 'jumping on the technology driven band wagon' of the next generation sequencing techniques this line of investigation has survived for some time. An extension to this time frame has been achieved by, among others, successful attempts to combine diversity with function by either statistical means (i.e. correlation to data on environmental parameters) or by combination with other experimental approaches (e.g. stable isotope probing). Clearly, analyses of the microbial diversity within environmental samples will remain an important task, though does already now no longer represent the main 'punch line' of research articles in mainstream journals in microbial ecology.

The question raised above also points towards the believe of some colleagues of the Author that the time-consuming and mostly ineffective attempts to isolate and culture novel microorganisms from environmental samples is soon becoming pointless. Although the Author lacks the infamous crytall ball to envisage future developments he still believes to have several arguments in hand to argue for continued attempts to isolate microorganisms. In fact, the Author deems isolation of microorganisms an indispensable requirement in Microbiology and, hence, in microbial ecology. Culture independent approaches are still limited in several aspects. For example, genome and transcriptome analyses reveal

genes that are relevant for a particular cellular process, but do not provide quantitative information about this biochemical process. The same is valid for attempts that employ systems biology for *in silico* modelling of cellular processes. Again, these are, by and large, still restricted to those metabolic pathways for which kinetic data are available on all or, as a minimum, most of the enzymes involved. So, physiological and biochemical analyses of isolates are required to fill this gap.

Another reason for isolation of environmentally relevant microorganisms to remain a main objective of microbial ecologists for some time to come is that this process, in contrast to most analytical approaches, cannot be 'outsourced'. Growth requirements are specific for each individual ecotype of a species due to the individuality of the various microorganisms, and their isolation requires – among others – not only creativity but also persistence and repeated attempts, hence experience. Moreover, as the genome analyses of "*Ferrovum*" strains demonstrates (section 2.4.2), even in cases where clonal cultures can not be obtained the availability of cultures enriched in the target organism is still of great benefit.

So, pure culture remains to be the most powerful tool in Microbiology to study and exploit microorganisms and, therefore, "obtaining a pure culture,…. is the principal task of microbiologists." (Schlegel in his textbook "Allgemeine Mikrobiologie"; 7th edition, 1992; p. 205: "Eine Reinkultur herzustellen,....ist die vornehmlichste Aufgabe des Mikrobiologen."). Moreover, the availability of a large culture collection where each genome is cloned "in its original host" (a quote believed to derive from Stephen Giovannoni) together with the appropriate promoters, regulators and relevant cofactors etc. for each of its genes provides a huge potential for biodiscovery (Mühling *et al.* 2013).

Twenty years of research (i.e. since the publication of the genome of *Haemophilus influenzae*: Fleischmann *et al.* 1995) also mean that genome analysis may now likewise be considered a 'classical' approach in microbial ecology which loses its appeal at a continually increasing rate. This notion is further supported by the Author's experience to publish the results from the genome analysis of "*Ferrovum*" strain JA12 in leading scientific journals in the field of genomic and environmental microbial research. Overall, this trend will continue, though application to novel groups of environmentally relevant microorganisms (e.g. "*Ferrovum*") and in combination with high-throughput techniques in single cell genomics will provide a small but safe haven for microbial genomics for some time to come.

What limits further advances in microbial ecology?

Limitations in microbial ecology can be classified in terms of not yet available technology and data analysis. As for the former, the transition of microbial genomics to the genome analysis of individual cells provides a powerful technique in microbial ecology. However, apart from problems associated to, among others, microbial contamination of isolated cells the approach is also limited by problems caused by multiple displacement amplification (MDA) used to amplify the whole genomic DNA (Pinard *et al.* 2006; Yilmaz *et al.* 2010). MDA still appears to be the method of choice for whole genome amplification (e.g. Hou *et al.* 2015), though this may vary dependent on the particular cell type (e.g. microbial versus eukaryote, GC content) and application Bourcy *et al.* 2014). Improvements to whole genome amplification in order to achieve unbiased sequencing and improved genome coverage is, however, in any case still needed. In the context of single cell genomics, the targeted isolation of representative cells of, for example, candidate divisions of microorganisms might, in principle, be achievable using fluorescence activated cell sorting (FACS) from environmental samples. However, despite early reports (Sekar *et al.* 2004) flow cytometric sorting of prokaryotic cells labelled by FISH with specific 16S rRNA targeting probes has not been extensively used since. Reasons for this are likely to be connected to the low fluorescence signals of 'standard' FISH probes. Alternatively, application of CARD-FISH which results in improved fluorescence properties, requires protease treatment of the cell wall for the uptake of probes. This is, however, likely to also lead to increased fragility of the cells which limits flow cytometric sorting. This explanation is also supported by the successful application of the *R*ecognition of *IN*dividual *G*enes in a single bacterial cell by Fluorescence *In Situ* Hybridization (RING-FISH: Zwirglmaier *et al.* 2004) to the detection of toxin producing *Microcystis aeruginosa* (Dziallas *et al.* 2011). RING-FISH does not require protease treatment of the cells prior to probe hybridisation (Zwirglmaier *et al.* 2004). However, to the best of the Author's knowledge this approach has not yet been applied for the targeted isolation of, for example, candidate divisions of microorganisms for subsequent genome analysis following MDA.

Regarding data analysis biologists have to admit that they, in contrast to physicists (e.g. those working in the field of experimental particle physics or astronomy), have traditionally not been used to dealing with large datasets such as, for example, those that result from next generation sequencing approaches now routinely applied by many microbial ecologists. Although a variety of bioinformatic analysis tools has been developed over the last decade these still need to be transformed into more efficient and automated software suitable to combine and handle various sets of data. Additionally, bioinformatic training has to become a major part in curricula of under- and postgraduate course programmes in biological sciences. Online data bases represent treasure chests filled to overflow with data awaiting to be analysed which – at this tpoint in time – requires the help of bioinformaticians in most cases. Such

studies have the potential to reveal global patterns in microbial ecology from which general laws may be deduced.

In this context, a broader knowledge in the application of appropriate mathematical, in particular statistical tests is also required for data analysis in order to ‘see the forest for the trees’. That is, relevant information needs to be extracted from the various datasets typically collected by microbial ecologists. Moreover, consideration should already be given to statistical tests at the stage of designing sample collection or experimental setups in order to ensure that interpretation of the results withstands vigorous scrutiny.

References

A. Cited references

Abicht HK, Mancini S, Karnachuk OV, Solioz M (2011) Genome sequence of *Desulfosporosinus* sp. OT, an acidophilic sulfate-reducing bacterium from copper mining waste in Norilsk, Northern Siberia. *J Bacteriol* **193**: 6104–6105.

Adams MWW (1990) The metabolism of hydrogen by extremely thermophilic, sulfur-dependent bacteria. *FEMS Microbiol Rev* **75**: 219–237.

Allen M, Tait K, Mühling M, Weynberg K, Bradley C, Trivedi U *et al.* (2012) Genome sequence of *Stenotrophomonas maltophilia* PML168, which displays Baeyer-Villiger monooxygenase activity. *J Bacteriol* **194**: 4753–4754.

Amann RI, Springer N, Ludwig W, Görtz HD, Schleifer KH (1991) Identification in situ and phylogeny of uncultured bacterial endosymbionts. *Nature* **351**: 161–164.

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

Arnosti C, Grossart H-P, Mühling M, Joint I, Passow U (2011) Dynamics of extracellular enzyme activities under changed atmospheric pCO_2: a mesocosm investigation. *Aquatic Microb Ecol* **64**: 285–298.

Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.

Baker-Austin C, Dopson M (2007) Life in acid: pH homeostasis in acidophiles. *Trends Microbiol* **15**: 165–171.

Bass Becking LGM (1934) Geobiologie of Inleiding Tot de Milieukunde. Van Stockum & Zoon, The Hague.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000a) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Béjà O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP *et al.* (2000b) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.

Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK (2005) The contribution of species richness and composition to bacterial services. *Nature* **436**: 1157–1160.

Beijerinck MW (1913) De infusies en de ontdekking der backteriën. Jaarboek van de Koninklijke Akademie voor Wetenschappen. Müller, Amsterdam.

Bent SJ, Pierson JD, Forney LJ (2007) Measuring species richness based on microbial community fingerprints: the emperor has no clothes. *Appl environ Microbiol* **73**: 2399–2401.

Biegel E, Schmidt S, González JM, Müller V (2011) Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell Mol Life Sci* **68**: 613–634.

Biegel E, Müller V (2010) Bacterial Na^+-translocating ferredoxin:NAD^+ oxidoreductase. *Proc Natl Acad Sci USA* **107**: 18138–18142.

Bigham JM, Schwertmann U, Carlson L, Murad E (1990) A poorly crystallized oxyhydroxysulfate of iron formed by bacterial oxidation of Fe(II) in acidic mine waters. *Geochim Cosmochim Acta* **54**: 2743–2758.

Blom J, Albaum SP, Doppmeier D, Pühler A, Vorhölter F-J, Zakrzewski M, Goesmann A (2009) EDGAR: A software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* **10**: 154.

De Bourcy CFA, De Vlaminck I, Kanbar JN, Wang J, Gawad C, Quake SR (2014) A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* **9**: e105585.

Bowman B, Shin MY, Lee J-E, Cho Y-J, Chun J (2013) Analysis of full-length metagenomic 16S genes by SMRT® sequencing. American Society for Microbiology – Poster Presentation.

Bradford A, Handy RD, Readman WJ, Atfield A, Mühling M (2009) Impact of silver nanoparticle contamination on the genetic diversity of natural bacterial assemblages in estuarine sediments. *Environ Science Technol* **43**: 4530–4536.

Bull AT, Stach JE (2007) Marine actinobacteria: new opportunities for natural product search and discovery. *Trends Microbiol* **15**: 491–499.

Bull AT, Ward AC, Goodfellow M (2000) Search and discovery strategies for biodiscovery: the paradigm shift. *Microbiol Mol Biol Rev* **64**: 573–606.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Castanie-Cornet MP, Penfound TA, Smith D, Elliott JF, Foster JW (1999) Control of acid resistance in *Escherichia coli*. *J Bacteriol* **181**: 3525–3535.

Caumette P, Bertrand J-C, Normand P (2015) Some historical elements of microbial ecology. ***In:*** Environmental Microbiology: Fundamentals and Applications. Bertrand J-C, Caumette P, Lebaron P, Matheron R, Normand P, Sime-Ngando T (Eds.), Springer, Berlin, pp. 9-24.

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA (1988) A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone. *Nature* **334**: 340–343.

Cho JC, Giovannoni SJ (2004) Cultivation and growth characteristics of a diverse group of oligotrophic marine gammaproteobacteria. *Appl environ Microbiol* **70**: 432–440.

Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A (2005) Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol* **7**: 1104–1115.

Clarke, K.R., and Warwick, R.M. (2001) Change in marine communities: an approach to statistical analysis and interpretation. 2nd Edition. Plymouth, UK: PRIMER-E Ltd.

Cord-Ruwisch R (1985) A quick method for the determination of dissolved and precipitated sulfides in cultures of sulfate-reducing bacteria. *J Microbiol Methods* **4**: 33–36.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland (MA).

Darwin C (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Davey ME, Wood WA, Key R, Nakamura K, Stahl DA (1993) Isolation of three species of Geotoga and *Petrotoga*: two new genera, representing a new lineage in the bacterial line of descent distantly related to the '*Thermotogales*'. Syst appl Microbiol **16**: 191–200.

Drescher K, Nadell CD, Stone HA, Wingreen NS, Bassler BL (2014) Solutions to the public goods dilemma in bacterial biofilms. *Current Biol* **24**: 50–55.

DeLong EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89**: 5685–5689.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nature Rev Microbiol* **6**: 725–740.

Dziallas C, Pinnow S, Grossart H-P (2011) Quantification of toxic and toxin-producing cyanobacterial cells by RING-FISH in combination with flow cytometry. *Limnol Oceanogr Meth* **9**: 67–73.

Editorial (2006) Brute force in the clinic. *Nature* **441**: 667–668.

Eaton KA, Brooks CL, Morgan DR, Krakowka S (1991) Essential role of urease inpathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. *Infect Immun* **59**: 2470–2475.

Eberle M, Razem AC (1985) Effects of surface coal mining and reclamation on ground water in small watersheds in the Allegheny Plateau, Ohio. Water-Resources Investigations Report 85-4205, U.S. Geological Survey.

Ehinger S (2011) Microbial communities and microbial processes in a german gas reservoir and their relevance for CO_2 storage. PhD thesis, TU Bergakademie Freiberg, Germany.

Eilers H, Pernthaler J, Glöckner FO, Amann R (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl environ Microbiol* **66**: 3044–3051.

Engel H, Hanert H (1967) Isolierung von *Gallionella ferruginea* Ehrenberg. *Naturwissenschaften* **54**: 147.

Erler A (2009) Charakterisierung von *Petrotoga*-Isolaten aus Formationswasser eines norddeutschen Erdgasfeldes. Diploma thesis. TU Bergakademie Freiberg, Germany.

Eisen S, Poehlein A, Schlömann M, Johnson DB, Daniel R, Mühling M (2015a) Genome of the acidophilic ferrous iron oxidising isolate *Acidithrix ferrooxidans* strain Py-F3, the proposed type strain of the novel actinobacterial genus *Acidithrix*. GenomeA **3**: e00382-15.

Eisen S, Poehlein A, Schlömann M, Johnson DB, Daniel R, Mühling M (2015b) Genome of the acidophilic iron oxidizer *Ferrimicrobium acidiphilum* strain T23T. GenomeA **3**: e00383-15.

Emerson D, Floyd MM (2005) Enrichment and isolation of iron-oxidizing bacteria at neutral pH. *Methods Enzymol* **397**: 112–124.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.

Fiala G, Stetter KO (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100 °C. *Arch Microbiol* **145**: 56–61.

Farasin J, Andres J, Casiot C, Barbe V, Faerber J, Halter D *et al.* (2015) *Thiomonas* sp.CB2 is able to degrade urea and promote toxic metal precipitation in acid mine drainage waters supplemented with urea. *Front Microbiol* **6**: 993.

Fardeau ML, Cayol JL, Magot M, Ollivier B (1993) H_2 oxidation in the presence of thiosulfate, by a *Thermoanaerobacter* strain isolated from an oil producing well. *FEMS Microbiol Lett* **113**: 327–332.

Feingersch R, Béjà O (2009) Bias in assessments of marine SAR11 biodiversity in environmental fosmid and BAC libraries? *ISME J* **3**: 1117–1119.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.

Festschrift (1993) Einweihung des Wasserwerkes Kesselshain. Messedruck Leipzig GmbH.

Frommeyer M, Steinbüchel A (2013) Increased lysine content is the main characteristic of the soluble form of the polyamide cyanophycin synthesized by recombinant *Escherichia coli*. *Appl environ Microbiol* **79**: 4474–4483.

Fuhrman JA, McCallum K, Davis AA (1992) Novel major archaebacterial group from marine plankton. *Nature* **356**: 148–149.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ *et al.* (2005) Opinion: re-evaluating prokaryotic species. *Nature Rev. Microbiol*. **3**: 733–739.

Gilbert J, Mühling M, Joint I (2008) A rare SAR11 fosmid clone confirming genetic variability in the '*Candidatus* Pelagibacter ubique' genome. *ISME J* **2**: 790–793.

Gilbert JA, Field D, Swift P, Thomas S, Cummings D, Temperton B, Weynberg K, Huse S, Hughes M, Joint I, Somerfield PJ, Mühling M (2010a) The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. *PLoS ONE* **5**: e15545.

Gilbert JA, Meyer F, Schriml L, Joint IR, Mühling M, Field D (2010b) Metagenomes and metatranscriptomes from the L4 long-term coastal monitoring station in the Western English Channel. *Stand Genomic Sciences* **3**: 183–193.

Gilbert JA, Laverock B, Temperton B, Thomas S, Mühling M, Hughes M (2011) Metagenomics. ***In:*** *Methods in Molecular Biology*, Vol 733: *High-throughput Sequencing*. Kwon YM, Ricke SC (Eds.), Springer, New York, pp. 173–183.

Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B *et al.* (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60–63.

Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al.* (2005a) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82–85.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005b) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.

Griebler C, Stein H, Kellermann C, Berkhoff S, Brielmann H, Schmidt SI, Selesi D, Steube C, Fuchs A, Hahn HJ (2010) Ecological assessment of groundwater ecosystems – vision or illusion? *Ecol Eng* **36**: 1174–1190.

Guindon S, Delsuc F, Dufayard J-F, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods Mol BioL* **537**: 113–137.

Hallbeck L, Stahl F, Pedersen K (1993) Phylogeny and phenotypic characterization of the stalk-forming and iron oxidizing bacterium *Gallionella ferruginea*. *J Gen Microbiol* 1531–1535.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245–R249.

Harris JK, Kelley ST, Pace NR (2004) New perspective on uncultured bacterial phylogenetic division OP11. *Appl Environ Microbiol* **70**: 845–849.

Hedrich S, Schlömann M, Johnson DB (2011a) The iron-oxidizing proteobacteria. *Microbiology* **157**:1551–1564.

Hedrich S, Lünsdorf H, Kleeberg R, Heide G, Seifert J, Schlömann M (2011b) Schwertmannite formation adjacent to bacterial cells in a mine water treatment plant and in pure cultures of *Ferrovum myxofaciens*. *Environ Sci Technol* **45**: 7685–7692.

Heinzel E, Hedrich S, Janneck E, Glombitza F, Seifert J, Schlömann M (2009a) Bacterial diversity in a mine water treatment plant. *Appl environ Microbiol* **75**: 858–861.

Heinzel E, Janneck E, Glombitza F, Schlömann M, Seifert J (2009b) Population dynamics of iron-oxidizing communities in pilot plants for the treatment of acid mine waters. *Environ Sci Technol* **43**: 6138–6144.

Hofmann S (2012) Description of new putative species of the genus *Petrotoga*. Diploma thesis, TU Bergakademie Freiberg, Germany.

Hou Y, Wu K, Shi X, Li F, Song L, Wu H, *et al.* (2015) Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *GigaScience* **4**: 37.

Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* **180**: 366–376.

Ivars-Martínez E, D´Auria G, Rodríguez-Valera F, Sánchez-Porro C, Ventosa A, Joint I, Mühling M (2008) Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. *Mol Ecol* **17**: 4161–4175.

Jameson E, Joint I, Mann NH, Mühling M (2008) Application of a novel *rpoC1*-RFLP approach reveals that *Prochlorococcus* populations in the Atlantic gyres are composed of greater microdiversity than previously described. *Microbial Ecology* **55**: 141–151.

Jameson E, Joint I, Mann NH, Mühling M (2010) Detailed analysis of the microdiversity of *Prochlorococcus* populations along a North-South Atlantic Ocean transect. *Environmental Microbiology* **12**: 156–171.

Jameson E, Mann NH, Joint I, Sambles C, Mühling M (2011) The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. *ISME J* **5**: 1713–1721.

Janneck E, Arnold I., Koch T, Meyer J, Burghardt D, Ehinger S (2010) Microbial synthesis of schwertmannite from lignite mine water and its utilization for removal of arsenic from mine waters and for production of iron pigments. ***In:*** Mine water & innovative thinking. Wolkersdorfer C, Freund A (Eds), CBU Press, Sydney, Nova Scotia, pp. 131–134.

Joint I, Mühling M, Querellou J (2010) Culturing marine bacteria - an essential prerequisite for biodiscovery. *Microb Biotechnol* **3**: 564–575.

Johnson DB, McGinness S (1991) A highly efficient and universal solid medium for growing mesophilic and moderately thermophilic, iron-oxidizing, acidophilic bacteria. *J Microbiol Methods* **13**: 113–22.

Johnson DB, Hallberg KB (2007) Techniques for detecting and identifying acidophilic mineral-oxidizing microorganisms. ***In:*** Biomining. Rawlings DE, Johnson DB (Eds.), 1st edition, Springer, Berlin, pp. 237–261.

Johnson DB, Bacelar-Nicolau P, Okibe N, Thomas A, Hallberg KB (2009) *Ferrimicrobium acidiphilum* gen. nov., sp. nov. and *Ferrithrix thermotolerans* gen. nov., sp. nov.: heterotrophic, iron-oxidizing, extremely acidophilic actinobacteria. *Int J Syst Evol Microbiol* **59**: 1082–1089.

Johnson DB, Hallberg KB, Hedrich S (2014) Uncovering a microbial enigma: isolation and characterization of the streamer-generating, iron-oxidizing, acidophilic bacterium "*Ferrovum myxofaciens*". *Appl environ Microbiol* **80**: 672–680.

Joffroy B (2014) Etablierung einer real-time quantitativen PCR (qPCR) Methode zum Nachweis der Genexpression eines Na^+/H^+-Antiporters im acidophilen *Desulfosporosinus* sp. Stamm M1. Bachelor thesis, TU Bergakademie Freiberg, Germany.

Jones RM, Johnson DB (2015) *Acidithrix ferrooxidans* gen. nov., sp. nov.; a filamentous and obligately heterotrophic, acidophilic member of the *Actinobacteria* that catalyzes dissimilatory oxido-reduction of iron. *Res Microbiol* **166**: 111–120.

Jukes TH Cantor CR (1969) Evolution of protein molecules. ***In:*** Munro HN (Ed.) *Mammalian protein metabolism*. 1st edition, Academic Press, New York, pp. 21–132.

Kaeberlein T, Lewis K, Epstein SS (2002) Isolating ”uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.

Kamerbeek NM, Fraaije MW, Janssen DB (2004) Identifying determinants of NADPH specificity in Baeyer–Villiger monooxygenases. *Eur J Biochem* **271**: 2107–2116.

Kermer R, Hedrich S, Taubert M, Baumann S, Schlömann M, Johnson DB, Seifert J (2012) Elucidation of carbon transfer in a mixed culture of *Acidiphilium cryptum* and *Acidithiobacillus ferrooxidans* using protein-based stable isotope probing. *J Integr OMICS* **2**: 37–45

Klein R, Schlömann M, Zeng Y, Wacker B, Glombitza F, Janneck E, Mühling M (2013) Impact of the hydraulic retention time on the performance of a sulfidogenic bioreactor. *Adv Mat Res* **825**: 392–395.

Klein R, Tischler JS, Mühling M, Schlömann M (2014) Bioremediation of mine waters. ***In:*** Geobiotechnology I *Adv Biochem Eng Biotechnol,* Vol. 141, Schippers A, Glombitza F, Sand W (Eds.), Springer, Berlin, pp.109–172.

Klindworth A, Pruesse E, Schweer T, Peplles J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**: e1.

Kogure K, Simidu U, Taga N (1979) A tentative direct microscopic method for counting living marine bacteria. *Can J Microbiol* **25**: 415–420.

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.

Konstantinidis KT, Ramette A, Tiedje JM (2006) The bacterial species definition in the genomic era. *Phil Trans R Soc B* **361**, 1929–1940

Korbel KL, Hose GC (2011) A tiered framework for assessing groundwater ecosystem health. *Hydrobiologia* **661**: 329–349.

Krehenbrink M, Oppermann-Sanio FB, Steinbüchel A (2002) Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from *Acinetobacter* sp. strain DSM 587. *Arch Microbiol* **177**: 371–380.

Kucera S, Wolfe RS (1957) A selective enrichment method for *Gallionella ferruginea*. *J Bacteriol* **74**: 344–349.

Kumagai H, Fujiwara T, Matsubara H, Saeki K (1997) Membrane localization, topology, and mutual stabilization of the *rnfABC* gene products in *Rhodobacter capsulatus* and implications for a new family of energy-coupling NADH oxidoreductases. *Biochemistry* **36**: 5509–5521.

Lam KN, Charles TC (2015) Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. *Microbiome* **3**: 22.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lien, T., M. Madsen, F. A. Rainey, and N. K. Birkeland. 1998. *Petrotoga mobilis* sp. nov., from a North Sea oil-production well. *Int J Syst Bact* **48**: 1007–1013.

Lovely DR, Phillips EJP (1987) Rapid assay for microbially reduced ferric iron in aquatic sediments. *Appl environ Microbiol* **53**: 1536–1540.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Lupan I, Ianc MB, Kelemen BS, Carpa R, Rosca-Casian O, Chiriac MT, Popescu O (2014) New and old microbial communities colonizing a seventeenth-century wooden church. *Folia Microbiol* **59**: 45–51.

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* **95**: 3140–3145.

Malik B, Su W-w, Wald HL, Blumentals II, Kelly RM (1988) Growth and gas production by the hyperthermophilic archaebacterium, *Pyrococcus furiosus*. *Biotech Bioeng* **32**: 438–444.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–832.

Mary I, Heywood JL, Fuchs BM, Amann R, Tarran GA, Burkhill PH *et al.* (2006) SAR11 dominance among metabolically active low nucleic acid bacterioplankton in surface waters along an Atlantic Meridional Transect. *Aquatic Microb Ecol* **45**: 107–113.

Meehan CJ, Beiko RG (2014) A phylogenomic view of ecological specialization in the *Lachnospiraceae*, a family of digestive tract-associated bacteria. *Genome Biol Evol* **6**: 703–713.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**: 60.

Miranda-Tello E, Fardeau ML, Thomas P, Ramírez F, Casalot L, Cayol JL, Garcia JL, Ollivier B (2004) *Petrotoga mexicana* sp. nov., a novel thermophilic, anaerobic and xylanolytic bacterium isolated from an oil-producing well in the Gulf of Mexico. Int J Syst Evol Microbiol 54: 169–174.

Miranda-Tello E, Fardeau ML, Joulian C, Magot M, Thomas P, Tholozan JL, Ollivier B (2007) *Petrotoga halophila* sp. nov., a thermophilic, moderately halophilic, fermentative bacterium isolated from an offshore oil well in Congo. *Int J Syst Evol Microbiol* **57**(Pt 1): 40–44.

Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Morris JJ, Lenski RE, Zinser ER (2012) The Black Queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* **3**: e00036–12.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* **5**: 621–628.

Moya-Beltrán A, Cárdenas JP, Covarrubias PC, Issotta F, Ossandon FJ, Grail BM, Holmes DS, Quatrini R, Johnson DB (2014) Draft genome sequence of the nominated type strain of "*Ferrovum myxofaciens*", an acidophilic, iron-oxidizing betaproteobacterium. *GenomeA* **2**: e00834-14.

Mühling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, Post AF, Joint I, Mann NH (2005) Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ Microbiol* **7**: 499–508.

Mühling M, Fuller NJ, Somerfield PJ, Post AF, Wilson WH, Scanlan DJ, Joint I, Mann NH (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using *rpoC1*-based restriction fragment length polymorphism. *Aquatic Microb Ecol* **45**: 263–275.

Mühling M, Woolven-Allen JA, Murrell JC, Joint I (2008) Improved group-specific PCR primers for DGGE analysis of the genetic diversity of complex microbial communities. *ISME J* **2**: 379–392.

Mühling M (2012) On the culture-independent assessment of the diversity and distribution of *Prochlorococcus*. *Environ Microbiol* **14**: 567–579.

Mühling M, Joint I, Willetts AJ (2013) The biodiscovery potential of marine bacteria: an investigation of phylogeny and function. *Microb Biotechnol* **6**: 361–370.

Mühling M, Beier R, Müller P, Petzsch P, Drechsel A, Schlömann M, Labudde D (2016) OEZY: Optimising EnZYme selection for best performing terminal restriction fragment length polymorphism (TRFLP) analysis using ARB. Methods Ecol Evol **7**: 242-248.

Müller P (2011) Etablierung einer mathematischen Methode zur optimalen Auswahl von Restriktionsendo-nukleasen für die Terminale Restriktions-Fragmentlängen-Polymorphismus (TRFLP) Analyse von mikrobiellen Populationen. Bachelor thesis, TU Bergakademie Freiberg, Germany.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH *et al.* (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329.

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Pande S, Kaftan F, Lang S, Svatoš A, Germerodt S, Kost C (2016) Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. *ISME J* **10**: 1413–1423.

Pester M, Brambilla E, Alazard D, Rattei T, Weinmaier T, Han J *et al.* (2012) Complete genome sequences of *Desulfosporosinus orientis* DSM765^T, *Desulfosporosinus youngiae* DSM17734^T, *Desulfosporosinus meridiei* DSM13257^T, and *Desulfosporosinus acidiphilus* DSM22704^T. *J Bacteriol* **194**: 6300–6301.

Petzsch P, Poehlein A, Johnson DB, Daniel R, Schlömann M, Mühling M (2015a) Genome of the moderately acidophilic sulfate reducing firmicute *Desulfosporosinus acididurans*^{T} (strain M1). GenomeA **3**: e00881-15.

Petzsch P, Poehlein A, Johnson DB, Daniel R, Schlömann M, Mühling M (2015b) Genome sequence of the acidophilic sulfate-reducing *Peptococcaceae* strain CEB3. GenomeA **3**: e00886-15.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH (2006) Assessment of whole genome amplification-induced bias through highthroughput, massively parallel whole genome sequencing. *BMC Genomics* **7**: 216.

Pissoke A (2013) Untersuchungen zum Wachstum von azidophilen Eisen-oxidierenden Mikroorganismen auf verschiedenen Festmedien mit Fokus auf der Kultivierung von *Ferrovum myxofaciens*. Bachelor thesis, TU Bergakademie Freiberg, Germany.

Pollender A (2014) Genomanalyse eines neuen Isolats der Gattung *Petrotoga.* Postgraduate project thesis, TU Bergakademie Freiberg, Germany.

Pollender A (2015) Expressionsstudien an *Petrotoga* sp. Sch_Z2_3. Master thesis, TU Bergakademie Freiberg, Germany.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Raes J, Korbel JO, Lercher MJ, Mering von C, Bork P (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: R10.

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.

Ricke P, Kolb S, Braker G (2005) Application of a newly developed ARB software-integrated tool for in silico terminal restriction fragment length polymorphism analysis reveals the dominance of a novel *pmoA* cluster in a forest soil. *Appl environ Microbiol* **71**: 1671–1673.

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW (2009) Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* **4**: e6864.

Rösner S (2011a) Hydrogeologischer Ergebnisbericht Grundwassermonitoring Wasserfassung Kitzscher, Zeitraum 2009-2010. G.E.O.S. Freiberg Ingenieurgesellschaft mbH, (*unpublished*).

Rösner S (2011b) Hydrogeologischer Ergebnisbericht Grundwassermonitoring Bornaer Mulde, Wasserwerk Borna-Altstadt, Wasserwerk Borna-Kesselshain, Zeitraum 2009-2010. G.E.O.S. Freiberg Ingenieurgesellschaft mbH (*unpublished*).

Rokas A, Williams LB, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.

Roy A-S, Gibbons SM, Schunck H, Owens S, Caporaso JG, Sperling M, Nissimov JI, Romac S, Bittner L, Mühling M, Riebesell U, LaRoche J, Gilbert JA (2013) Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. *Biogeosciences* **10**: 555–566.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: 398–431.

Sachs G, Weeks DL, Wen Y, Marcus EA, Scott DR, Melchers K (2005) Acid acclimation by *Helicobacter pylori*. *Physiology* **20**: 429–438.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.

Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA (1986) Analysis of enzymatically amplified *β*-globin and HLA DQ*α* DNA with allele-specific oligonucleotide probes. *Nature* **324**: 163–166.

Sapra R, Bagramyan K, Adams MWW (2003) A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc Natl Acad Sci USA* **100**: 7545–7550.

Schober P (2011) Kultivierung eisenoxidierender Bakterien mit Hilfe modiizierter Festmedien. Bachelor thesis, TU Bergakademie Freiberg, 2011.

Schut GJ, Adams MWW (2009) The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**: 4451–4457.

Schut GJ, Boyd ES, Peters JW, Adams MWW (2013) The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic *Archaea* and their evolutionary implications. *FEMS Microbiol Rev* **37**: 182–203.

Sekar R, Fuchs BM, Amann R, Pernthaler J (2004) Flow sorting of marine bacterioplankton after fluorescence in situ hybridization. *Appl environ Microbiol* **70**: 6210–6219.

Simon RD, Weathers P (1976) Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. *Biochim Biophys Acta* **420**: 165–176.

Sirisena KA, Daughney CJ, Moreau-Fournier M, Ryan KG, Chambers GK (2013) National survey of molecular bacterial diversity of New Zealand groundwater: relationships between biodiversity, groundwater chemistry and aquifer characteristics. *FEMS Microbiol Ecol* **86**: 490–504.

Slater JH, Bull AT (1982) Environmental microbiology: biodegradation. *Philos Trans R Soc Lond B* **297**: 575–597.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al.* (2006) Microbial diversity in the deep sea and the underexplored ‘rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *Univ Kansas Science Bull* **38**: 1409–1438.

Sonnhammer ELL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175–182.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318**:1449–52.

Spratt BG, Staley JT, Fisher MC (2006) Introduction: species and speciation in micro-organisms. ***In:*** Species and speciation in micro-organisms. Fisher MC, Spratt BG, Staley JT (Eds.) *Phil Trans Roy Soc B* **361**: 1897–1898.

Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39**: 321–346.

Stein H, Kellermann C, Schmidt SI, Brielmann H, Steube C, Berkhoff SE, Fuchs A, Hahn HJ, Thulin B, Griebler C (2010) The potential use of fauna and bacteria as ecological indicators for the assessment of groundwater quality. *J Environ Monit* **12**: 242–254.

Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* **104**: 9052–9057.

Steube C, Richter S, Griebler C (2009) First attempts towards an integrative concept for the ecological assessment of groundwater ecosystems. *Hydrogeol J* **17**: 23–35.

Suttle CA, Chan AM (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl environ Microbiol* **60**: 3167–3174.

Suzuki MT, Preston CM, Béjà O, de la Torre JR, Steward GF, DeLong EF (2004) Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48**: 473–488.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tatum EL, Lederberg J (1947) Gene recombination in the bacterium *Escherichia coli*. *J Bact* **53**: 673–684.

Temperton B, Oliver A, Field D, Tiwari B, Mühling M, Joint I, Gilbert JA (2009) Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *ISME J* **3**: 792–796.

Touvinen OH, Kelly DP (1973) Studies on the growth of *Thiobacillus ferrooxidans*. I. Use of membrane filters and ferrous iron agar to determine viable numbers, and comparison with $^{14}CO_2$-fixation and iron oxidation as measures of growth. *Arch Mikrobiol* **88**: 285–98.

Tischler JS, Jwair RJ, Gelhaar N, Drechsel A, Skirl A, Wiacek C, Janneck E, Schlömann M (2013) New cultivation medium for "*Ferrovum*" and *Gallionella*-related strains. *J Microbiol Methods* **95**:138–144.

Ullrich SR, Poehlein A, Voget S, Hopert M, Daniel R, Leimbach A, Tischler JS, Schlömann M, Mühling M (2015) Permanent draft genome sequence of *Acidiphilium* sp. JA12-A1. *Stand Genomic Sciences* **10**: 56.

Ullrich SR, Poehlein A, Tischler JS, González C, Ossandon F, Daniel R, Holmes DS, Schlömann M, Mühling M (2016a) Genome analysis of the biotechnologically relevant acidophilic iron oxidising strain JA12 indicates phylogenetic and metabolic diversity within the novel genus "*Ferrovum*". *PLoS ONE* **11**: e0146832.

Ullrich SR, González C, Poehlein A, Tischler JS, Daniel R, Schlömann M, Holmes DS, Mühling M (2016b) Gene loss and horizontal gene transfer contributed to the genome evolution of the extreme acidophile "*Ferrovum*". *Front Microbiol* **7**: 797.

Vallon O (2000) New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure. *Proteins Struct Funct Genet* **38**: 95–114.

van Berkel WJH, Kamerbeek NM, Fraaije MW (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* **124**: 670–689.

Voitel M (2015) Phylogenetic and (meta)genomic characterisation of a microaerophilic enrichment culture in APPW medium (pH 3.5). Bachelor thesis, TU Bergakademie Freiberg, Germany.

von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.

Wagner-Döbler I, Beil W, Lang S, Meiners M, Laatsch H (2002) Integrated approach to explore the potential of marine microorganisms for the production of bioactive metabolites. *Adv Biochem Eng Biotechnol* **74**: 207–238.

Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979) Wide-spread occurrence of a unicellular, marine planktonic, cyanobacterium. *Nature* **277**: 293–294.

Ward DM, Weller R, Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**: 63–65.

Waterbury JB, Valois FW (1993) Resistance to cooccurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl environ Microbiol* **59**: 3393–3399.

Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI *et al.* (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**: 463–464.

Widdel F, Bak F (1992) Gram-negative mesophilic sulfatereducing bacteria. ***In:*** The Prokaryotes. Balows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (Eds.), Springer, New York, pp. 3352–3378.

Wiegand S, Dietrich S, Hertel R, Bongaerts J, Evers S, Volland S, Daniel R, Liesegang H (2013) RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation. *BMC Genomics* **14**: 667.

Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2**: 27.

Willetts A, Joint I, Gilbert JA, Trimble W, Mühling M (2012) Isolation and initial characterization of a novel type of Baeyer–Villiger monooxygenase activity from a marine microorganism. *Microb Biotechnol* **5**: 549–559.

Wilson WH, Carr NG, Mann NH (1996) The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. *J Phycol* **32**: 506–516.

Winkelmann N, Harder J (2009) An improved isolation method for attached-living *Planctomycetes* of the genus *Rhodopirellula*. *J Microbiol Methods* **77**: 276–284.

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**: 5088–5090.

Wolfe RS (1958) Cultivation, morphology, and classification of the iron bacteria. *J American Water Works Association* **50**: 1241–1249.

Wood JM (1983) Selected biochemical reactions of environmental significance. *Chemica Scripta* **21**: 155–160.

Yilmaz S, Allgaier M, Hugenholtz P (2010) Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nature Methods* **7**: 943–944.

Younger PL (2007) Groundwater in the environment: an introduction. Blackwell, London.

Zengler K, Toledo G, Rappé M, Elkins J, Mathur EJ, Short JM, Keller M (2002) Cultivating the uncultured. *Proc Natl Acad Sci USA* **99**: 15681–15686.

Zinder ND, Lederberg J (1952) Genetic exchange in *Salmonella*. *J Bacteriol* **64**: 679–699.

Zwirglmaier K, Ludwig W, Schleifer KH (2004) Recognition of individual genes in a single bacterial cell by fluorescence in situ hybridization – RING-FISH. *Mol Microbiol* **51**: 89–96.

B. Publications resulting from the Author's own research and related to the topic of the *Habilitation*

Please note that the successive numbers of the publications (i.e. 1 to 28) also refers to the order of the reprints as provided in the appendix.

Chapter 2: The extended 'classical' approach to the study of micro-organisms: isolation followed by phenotypic and genomic characterisation

2.1 Novel methods for the isolation of microorganisms

1. Joint I, **Mühling M**, Querellou J (2010) Culturing marine bacteria – an essential prerequisite for biodiscovery. *Microbial Biotechnology* **3**: 564–575.

2.2 Detailed phylogenetic classification of microbial strains

2. Ivars-Martínez E, D´Auria G, Rodríguez-Valera F, Sánchez-Porro C, Ventosa A, Joint I, **Mühling M** (2008) Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. *Molecular Ecology* **17**: 4161–4175.

   See also Mühling (2012; ref. 13) for further details on the evaluation of marker genes for the analysis of microbial diversity

2.3 Screening isolates for relevant enzymatic activities

3. Willetts A, Joint I, Gilbert JA, Trimble W, **Mühling M** (2012) Isolation and initial characterization of a novel type of Baeyer–Villiger monooxygenase activity from a marine microorganism. *Microbial Biotechnology* **5**: 549-559**.**

4. **Mühling M**, Joint I, Willetts AJ (2013) The biodiscovery potential of marine bacteria: an investigation of phylogeny and function. *Microbial Biotechnology* **6**: 361-370.

2.4 Genome analysis of microorganisms

2.4.1 A novel type of Baeyer–Villiger monooxygenase activity from a marine gammaproteobacterium

5. Allen M, Tait K, **Mühling M**, Weynberg K, Bradley C, Trivedi U, Gharbi K, Nissimov J, Mavromatis K, Jensen C, Grogan G, Ali S (2012) Genome sequence of *Stenotrophomonas maltophilia* PML168, which displays Baeyer-Villiger monooxygenase activity. *Journal of Bacteriology* **194**: 4753-4754.

2.4.2 Genome analysis of acid mine drainage related acidophilic bacteria

6. Ullrich SR, Poehlein A, Voget S, Hopert M, Daniel R, Leimbach A, Tischler JS, Schlömann M, **Mühling M** (2015) Permanent draft genome sequence of *Acidiphilium* sp. JA12-A1. *Standards in Genomic Sciences* **10**: 56.

7. Ullrich SR, Poehlein A, Tischler JS, González C, Ossandon F, Daniel R, Holmes DS, Schlömann M, **Mühling M** (2016a) Genome analysis of the biotechnologically relevant acidophilic iron oxidising strain JA12 indicates phylogenetic and metabolic diversity within the novel genus "*Ferrovum*". *PLoS ONE* **11**: e0146832.

8. Ullrich SR, González C, Poehlein A, Tischler JS, Daniel R, Schlömann M, Holmes DS, **Mühling M** (2016b) Gene loss and horizontal gene transfer contributed to the genome evolution of the extreme acidophile "*Ferrovum*". *Frontiers in Microbiology* **7**: 797.

9. Petzsch P, Poehlein A, Johnson DB, Daniel R, Schlömann M, **Mühling M** (2015a) Genome of the moderately acidophilic sulfate reducing firmicute *Desulfosporosinus acididurans*^{T} (strain M1) *GenomeA* **3**: e00881-15.

10. Petzsch P, Poehlein A, Johnson DB, Daniel R, Schlömann M, **Mühling M** (2015b) Genome sequence of the acidophilic sulfate-reducing *Peptococcaceae* strain CEB3 *GenomeA* **3**: e00886-15.

11. Eisen S, Poehlein A, Schlömann M, Johnson DB, Daniel R, **Mühling M** (2015) Genome of the acidophilic ferrous iron oxidising isolate *Acidithrix ferrooxidans* strain Py-F3, the proposed type strain of the novel actinobacterial genus *Acidithrix*. *GenomeA* **3**: e00382-15.

12. Eisen S, Poehlein A, Schlömann M, Johnson DB, Daniel R, **Mühling M** (2015) Genome of the acidophilic iron oxidizer *Ferrimicrobium acidiphilum* strain T23^{T}. *GenomeA* **3**: e00383-15.

2.4.3 Genome analysis and taxonomic characterisation of *Petrotoga* strains isolated from gas producing wells

See report on the metagenome analysis of a microaerophilic enrichment culture.

Chapter 3: Culture independent molecular diversity and metagenomics

3.1 Method development for genetic fingerprint techniques

3.1.1 Evaluating marker genes for potential horizontal gene transfer events and genetic resolution

13. **Mühling M** (2012) On the culture-independent assessment of the diversity and distribution of *Prochlorococcus*. *Environmental Microbiology* **14**: 567–579.

    See also Ivars-Martínez *et al.* (2008: ref. 2)

3.1.2 Development of PCR primers specific for particular phylogenetic groups

14. **Mühling M,** Woolven-Allen JA, Murrell JC, Joint I (2008) Improved group-specific PCR primers for DGGE analysis of the genetic diversity of complex microbial communities. *The ISME Journal* **2**: 379–392.

15. **Mühling M**, Fuller NJ, Somerfield PJ, Post AF, Wilson WH, Scanlan DJ, Joint I, Mann NH (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using *rpoC1*-based restriction fragment length polymorphism. *Aquatic Microbial Ecology* **45**: 263-275.

16. Jameson E, Joint I, Mann NH, **Mühling M** (2008) Application of a novel *rpoC1*-RFLP approach reveals that *Prochlorococcus* populations in the Atlantic gyres are composed of greater microdiversity than previously described. *Microbial Ecology* **55**: 141–151.

3.1.4 OEZY – a new software tool for the selection of the most suitable restriction endonucleases for TRFLP analyses

17. **Mühling M**, Beier R, Müller P, Petzsch P, Schlömann M, Labudde D (2016) OEZY: Optimising EnZYme selection for best performing terminal restriction fragment length polymorphism (TRFLP) analysis using ARB. *Methods in Ecology and Evolution* **7**: 242-248.

3.2 Applications of genetic fingerprint techniques

Analysis of the microdiversity of abundant groups of marine cyanobacteria and the potential impact of viral control on those

18. **Mühling M**, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, Post AF, Joint I, Mann NH (2005) Genetic diversity of marine *Synechococcus* and co-occurring

cyanophage communities: evidence for viral control of phytoplankton. *Environmental Microbiology* **7**: 499-508.

19. Jameson E, Joint I, Mann NH, **Mühling M** (2010) Detailed analysis of the microdiversity of *Prochlorococcus* populations along a North-South Atlantic Ocean transect. *Environmental Microbiology* **12**: 156–171.

20. Jameson E, Mann NH, Joint I, Sambles C und **Mühling M** (2011) The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. *The ISME Journal* **5**: 1713–1721.

Analysis of the potential danger of environmental pollution by anthropogenically produced nanoparticles

21. Bradford A, Handy RD, Readman WJ, Atfield A, **Mühling M** (2009) Impact of silver nanoparticle contamination on the genetic diversity of natural bacterial assemblages in estuarine sediments. *Environmental Science & Technology* **43**: 4530–4536.

Analysis of the impact of increasing atmospheric pCO_2 on the microbial diversity in the marine environment

22. Arnosti C, Grossart H-P, **Mühling M**, Joint I, Passow U (2011) Dynamics of extracellular enzyme activities under changed atmospheric pCO_2: a mesocosm investigation. *Aquatic Microbial Ecology* **64**: 285–298.

    See also Roy *et al.* (2013: ref. 24) in which samples from a further mesocosm experiment were analysed, but with 16S-tag Illumina sequence analysis.

Analysis of the microbial population dynamics in sulfidogenic bioreactors used for bioremediation of AMD waters

23. Klein R, Schlömann M, Zeng Y, Wacker B, Glombitza F, Janneck E, **Mühling M** (2013) Impact of the hydraulic retention time on the performance of a sulfidogenic bioreactor. *Advanced Materials Research* **825**: 392-395.

The 'next generation fingerprint techniques': use of high-throughput sequencing methods

24. Roy A-S, Gibbons SM, Schunck H, Owens S, Caporaso JG, Sperling M, Nissimov JI, Romac S, Bittner L, **Mühling M**, Riebesell U, LaRoche J, Gilbert JA (2013) Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. *Biogeosciences* **10:** 555-566.

    See also Gilbert *et al.* (2010a: ref. 27) and the report on the results from the BMBF-funded project BRUNNTEC in section 3.2.

3.3 Metagenomic approaches

3.3.1 Use of high capacity cloning vectors

25. Gilbert J, **Mühling M**, Joint I (2008) A rare SAR11 fosmid clone confirming genetic variability in the '*Candidatus* Pelagibacter ubique' genome. *The ISME Journal* **2**: 790–793.

26. Temperton B, Oliver A, Field D, Tiwari B, **Mühling M**, Joint I, Gilbert JA (2009) Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal* **3**: 792–796.

3.3.2 Use of high-throughput sequencing techniques

27. Gilbert JA, Field D, Swift P, Thomas S, Cummings D, Temperton B, Weynberg K, Huse S, Hughes M, Joint I, Somerfield PJ, **Mühling M** (2010a) The taxonomic and functional diversity of microbes at a temperate coastal site: a 'multi-omic' study of seasonal and diel temporal variation. *PLoS ONE* **5**: e15545.

28. Gilbert JA, Meyer F, Schriml L, Joint IR, **Mühling M**, Field D (2010b) Metagenomes and metatranscriptomes from the L4 long-term coastal monitoring station in the Western English Channel. *Standards in Genomic Sciences* **3**: 183–193.

29. Gilbert JA, Laverock B, Temperton B, Thomas S, **Mühling M**, Hughes M (2011) Metagenomics. ***In:*** Kwon YM, Ricke SC (Eds.) *Methods in Molecular Biology*, Vol 733, pp. 173-183. Springer.

    See also report on the metagenome analysis of a microaerophilic enrichment culture in section 3.3.2.