

Nikolaos Strepis

BIOPROSPECTING OF *TRICHOCOCCUS* SPECIES



BIOPROSPECTING OF *TRICHOCOCCUS* SPECIES

Nikolaos Strepis

THESIS COMMITTEE

Promotor

Prof. Dr A.J.M. Stams
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotors

Dr D.Z. Machado de Sousa
Associate professor, Laboratory of Microbiology
Wageningen University & Research

Dr P.J. Schaap
Associate professor, Systems and Synthetic Biology
Wageningen University & Research

Other members

Prof. Dr M.H. Zwietering, Wageningen University & Research
Dr M.A. Pereira, University of Minho, Portugal
Prof. Dr B. Teusink, VU, Amsterdam
Dr S. Lücker, Radboud University, Nijmegen

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE).

BIOPROSPECTING OF *TRICHOCOCCUS* SPECIES

Nikolaos Strepis

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 16 January 2019
at 4 p.m. in the Aula.

Nikolaos Strepis

Bioprospecting of *Trichococcus* species,
226 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2019)

With references, with summary in English

DOI: <https://doi.org/10.18174/465966>

ISBN: 978-94-6343-388-4

**“LET THE FUTURE TELL THE TRUTH, AND EVALUATE
EACH ONE ACCORDING TO HIS WORK AND
ACCOMPLISHMENTS. THE PRESENT IS THEIRS; THE
FUTURE, FOR WHICH I HAVE REALLY WORKED, IS MINE.”**

Nikola Tesla

SUMMARY

Nowadays a vast diversity of cultured microorganisms have their genomes sequenced and made accessible to the public. Genome databases provide an ideal platform for discovering new physiological properties among these organisms. A particular application is the mining of these databases for finding new catalysts for the production of value-added compounds from waste materials. Further functional analysis of these microorganisms can provide valuable insights into defining strategies for optimizing product yields and titres. In this thesis, the *Trichococcus* genus was analysed using (functional) genomics approaches, and some of the species studied for their capacity to produce 1,3-propanediol (1,3-PDO) – a monomer used in the manufacturing of polytrimethylene terephthalate (major application), polyurethane, cosmetics, personal care and cleaning products. Based on the genomic patterns associated with this feature in the analyzed *Trichococcus* species, a global search for 1,3-PDO producing-strains was performed among 84,329 species (for which genomic data is publicly available). This approach, of mining microbial genomes for a specific phenotypic-desired trait is referred to within this thesis as bioprospecting.

Two novel members of *Trichococcus* genus were characterized within the frame of this PhD research. *Trichococcus ilyis* and the psychrotolerant *Trichococcus shcherbakoviae*. As seen among other *Trichococcus* species, their 16S rRNA gene is highly conserved and the application of genomic index values was necessary for the delineation of the new species. During the implementation of phylogenomics, genotype mapping was performed for the complete genus. Two major outcomes were derived from the data; *T. pasteurii* is able to ferment glycerol into the biotechnological attractive compound 1,3-PDO and *T. shcherbakoviae* can produce extracellular polymeric substances (EPS). This EPS is likely to be beneficial during growth at low temperatures by functioning as a cryoprotectant.

Genome and domain analyses of the nine members of *Trichococcus* genus were performed. Twenty-two closely related species the genera *Carnobacterium* and *Aerococcus* were also included in the analyses for comparison. Traits like 1,3-PDO production, arabinan (plant wall

component) degradation, alginate (biofilm component) degradation, EPS formation, the potential to grow at low temperature and resistance to high salinity were analysed in detail. To validate these findings, in vitro trials were also conducted.

The functional analysis in the genus *Trichococcus* uncovered a genomic synteny of 17 genes dedicated to 1,3-PDO in *T. pasteurii* and *T. flocculiformis* strain ES5. Inspired by this discovery, we applied a functional analysis to all the bacterial genomes in the public repositories to gain insights into bacteria that can produce 1,3-PDO from glycerol, and the pathways involved in doing so. In a top-down semantic systems biology approach, we could analyse 84,329 bacterial genomes for the presence of key functional components. This led to the identification of potential natural producers of 1,3-PDO and a few were selected for laboratory testing and validation. Furthermore, until now *Clostridium butyricum* was the only species reported to produce 1,3-PDO in a vitamin B12-independent way and in this study, many more potentially B12-independent producers were identified.

Considering all the metabolic knowledge revealed for the trait of 1,3-PDO, physiological and proteomics analysis was applied on *T. pasteurii* and *C. butyricum*. With a low concentration of glycerol (5 g/L) and yeast extract (1 g/L), *T. pasteurii* was observed to have a higher yield than *C. butyricum* (70 % versus 58 % respectively). In continuous bioreactors, a similar yield was obtained for *T. pasteurii* under well-controlled conditions. The metabolic outcome was altered at a higher glycerol concentration (45 g/L), where 1,3-PDO production was significantly lower. This 1,3-PDO production inhibition is caused by the formation 3-hydroxypropionaldehyde (3-HPA). Proteomic analysis showed a different fold change of enzymes catalyzing the conversion of 3-HPA when a high glycerol concentration was provided. The large B12-depedent operon in *T. pasteurii* was present in the proteome signifying its importance for 1,3-PDO production.

EPS formation was studied in *T. patagoniensis*, which was isolated from a low temperature environment (Patagonia). *T. patagoniensis* can grow at -5 °C in anoxic conditions excreting EPS. This EPS was analysed, and inulin detected as the main component. Only a limited number of bacteria are known to produce inulin, and none of the psychrotoler-

ant species. The proteomics analysis revealed further cold adaptation mechanisms in *T. patagoniensis*. In freezing temperature, the bacterium is equipped with multiple enzymes ensuring the safe operation of the translation machinery.

In summary, this thesis provides evidence of the potential of (functional) omics approaches to study the physiology of microorganisms and reveal unknown properties. The extended metabolic profiling of 1,3-PDO and the discovery of inulin production by *Trichococcus* species are examples of this potential. Other exceptional phenotypic properties were reported, setting the landscape for future bioprospecting.

TABLE OF CONTENTS

CHAPTER 1	015
General introduction and thesis outline	
CHAPTER 2	035
Description of <i>Trichococcus ilyis</i> sp. nov. by combined physiological and in silico genome hybridization analyses	
CHAPTER 3	051
Genome-guided identification of novel physiological traits of <i>Trichococcus</i> species	
CHAPTER 4	075
Systematic function-based genome prospecting for industrial traits applied to 1,3-propanediol production	
CHAPTER 5	097
Physiological and proteomics analysis of 1,3-propanediol metabolism in <i>Trichococcus pasteurii</i> and <i>Clostridium butyricum</i>	
CHAPTER 6	121
Proteomics of the psychrotolerant <i>Trichococcus patagoniensis</i> and identification of the produced EPS as inulin	
CHAPTER 7	145
<i>Trichococcus shcherbakoviae</i> sp. nov. isolated from a laboratory anaerobic expanded granular sludge bed bioreactor	
CHAPTER 8	159
General discussion and future recommendations	
REFERENCES	175
APPENDICES	211

CHAPTER 1

GENERAL INTRODUCTION AND THESIS OUTLINE

BIOPROSPECTING FOR MICROBIAL ADDED-VALUE PRODUCTS

An unknown microbial world

Bacteria are ubiquitous and are the main drivers of geochemical cycles on Earth¹. Microbial diversity is estimated to consist of more than 1 trillion microbial species², but the majority remains uncultivable and therefore unexploited¹. Discovering novel bacteria may reveal new phenotypes, adding to current knowledge on microbial metabolism. In the quest for novel microbes, frequently targeted environments are soil, the human gut, freshwater and marine waters³. A remarkable example that illustrates microbial diversity is a recent study conducted in Colorado, where metagenomes of aquifer sediments and groundwater samples were analysed, and 47 new bacterial phyla identified solely in this environment⁴. Habitats with more extreme conditions, such as low or high temperature and pH, are also good sources for finding novel microorganisms with new physiological properties^{1,5,6,7,8}. A study focused on the freezing environment of Antarctic lakes resulted in the isolation of 36 novel strains⁵, thereby increasing the number of known psychrotolerant/ psychrophilic species. In another conducted in Armenia, novel bacteria were recently isolated from geothermal springs, a high temperature ecosystem (40 and 44 °C)⁶. Furthermore, these novel thermophiles were said to produce exopolysaccharides (EPS), a biotechnologically interesting compound. There are a vast number of diverse microbes in many heterogeneous ecosystems leading to remarkable physiologies. In order to benefit from such phenotypes, a sufficient knowledge of their diverse metabolisms should be obtained.

Understanding microbial diversity

Addressing all the microbial diversity requires good taxonomical markers. For the past few decades, the 16S rRNA gene was used as the golden standard for microbial phylogenetic classification⁹. The similarity threshold for differentiation at species level is set at 98,7

% of the 16S rRNA gene sequence identity¹⁰. Due to the highly conserved 16S rRNA genes of several bacteria, especially within the *Mycobacterium* genus, determining species level only through 16S rRNA gene analysis may be laborious¹⁰. When the 16S rRNA gene identity of a novel strain and that of its closest relative are identical by higher than 98.7%, genome based methods are necessary to accurately assign organisms to an existing or new taxon¹¹. Traditionally experimental DNA-DNA hybridization was performed, however nowadays, with the possibility of genome sequencing, in-silico DNA-DNA comparisons are constantly applied (e.g. Average Nucleotide Identity (ANI) and digital DNA-DNA hybridisation (dDDH)). An additional advantage of genome sequencing is the possibility to further analyse the information in the genome and identify potential metabolic traits of the microorganism.

Average nucleotide identity (ANI) is based on high identity among conserved genes present in two genomes¹². A complete genome sequence is not required for the comparison. The ANI threshold for species delineation is 94-96 % sequence identity¹². ANI has been used for taxonomical distribution of species in multiple studies^{13,14} and the identification of new additions to microbial genera¹⁵.

Digital DNA-DNA hybridisation (dDDH) is an *in silico* adaption of experimental DNA-DNA hybridisation with the advantage of minimising the time of the analysis and experimental errors. The dDDH divides the genomes into fragments and compares their identities using BLAST¹⁶. High identity pairs are valued as high score segments pairs (HSP) for generating the hybridization value between two genomes¹⁷. The threshold of dDDH for defining new species is set at 70 %¹¹. As taxonomical method, dDDH contributed to the classification of various microorganisms including the collective phylogenetic analysis of 1,003 genomes¹⁸.

Functional domain comparison is based on conserved protein functional domains in the genomes, the conserved sequences of protein structures¹⁹. The comparison of a genome against the pan-domains or core domains of the genus can determine the variation of the species within the group. Whole genome domain class content correlates with and complements the 16S rRNA metric. Protein domains are based

on profile Hidden Markov Models that lend more weight to signifying amino acids²⁰. The strongest advantage of functional domains analysis is that it may help in inferring physiological properties of the microorganisms. The functional landscape links to its physiology and ecology.

Application of microorganisms in biotechnology

Addressing the microbial functional diversity facilitates the search for properties that are of potential interest for industrial application. The ambition to exploit microbes for identifying products with potential biotechnological value is here defined as bioprospecting. Bioprospecting is not a new idea and, since the discovery of penicillin in 1928, numerous microbes have been screened for their capacity to produce valuable compounds²¹. Nowadays, about 50 % of all commercial drugs either directly originate from microbes or are synthesized by mimicking their metabolites²². The difference nowadays is that bioprospecting may benefit from omics studies, considering the high number of sequences of organisms and metagenomes available in public databases. Bioprospecting may be extended to include the search for potential microbial candidates that produce biofuels, agrochemicals, enzymes, personal care products, flavours and fragrances.

Anaerobic microorganisms are frequently used in biotechnology due to their metabolic capabilities (e.g. production of reduced compounds, low biomass yields) and their capacity to thrive under anoxic conditions (avoids the need for oxygenation of the bioreactor). From the early times, anaerobic organisms were applied in food processing, including in the production of yoghurt, beer and cheese. More recently, application have been extended to the biofuels industry, where anaerobic bacteria are used to produce ethanol, butanol and hydrogen^{23,24}. As we shift towards a bio-based society, the generation of energy-rich compounds from cheap substrates or even waste streams is increasingly seen as a valuable attribute of anaerobes²⁵.

By searching the genomes of unexplored microorganisms, it is possible to find novel candidates to produce compounds of interest. Omics studies can provide insights into the biosynthetic power, regulative factors and metabolic profile of microorganisms that can be

extensively studied. In addition, functional genomics can contribute with information on favourable environmental conditions that can be assessed and is important for optimizing the production of a certain metabolite (Figure 1.1). An interdisciplinary approach of physiological characterisation and omics analyses can reveal the effective metabolic potential of microbes. Previously, the use of comparative genomics and transcriptomics guided the physiological exploitation of *Propionibacterium acidipropionici*²⁶. Using omics studies for identifying stress resistance mechanisms and describing the secondary metabolites pathways, the propionic acid production yield could be increased²⁶. Even though more than 190 products are derived from the *Cyanobacteria* species, a recent study using genomics was able to identify even more unique metabolites with increased biotechnological potential that were produced from four *Cyanobacterial* strains²⁷.

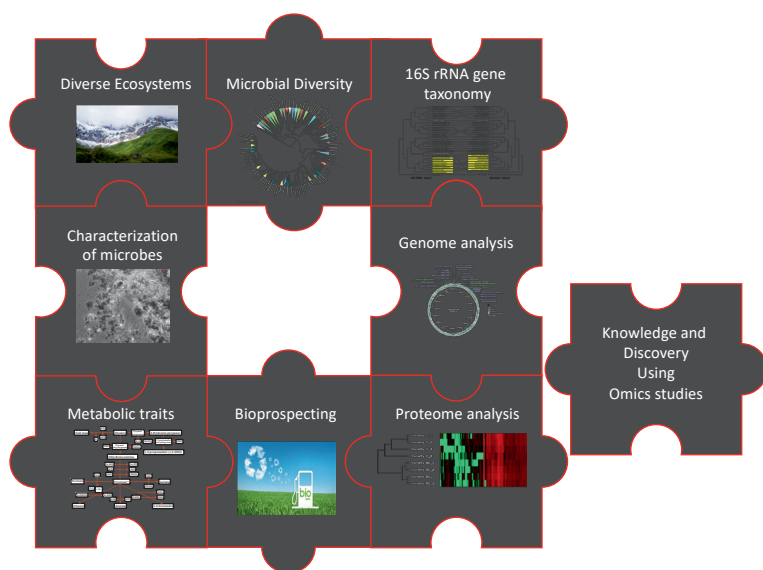


Figure 1.1 | The missing piece of the puzzle is the omics studies, which can be used for identifying biotechnological compounds and improving their production. This can benefit bioprospecting by focusing on specific microbial traits.

OMICS AGE – NEW PERSPECTIVES FOR BIOPROSPECTING

Genomics: the cornerstone of omics studies

Omics studies can enhance bioprospecting by identifying novel species, essential genes in large databases and key pathways for the synthesis of metabolites²⁸. Since the first genome was sequenced (*Haemophilus influenza*) in 1995, genomics has become a powerful tool. Due to the technological innovations in next generation sequencing (NGS), genomics have expanded even further. The cost of sequencing has dropped substantially for Illumina platforms (second-generation sequencing)²⁹, as well as for the latest technologies, PacBio (third-generation sequencing)³⁰. Thus far, more than 149,000 microbial genomes exist in the prokaryotic database of NCBI ("www.ncbi.nlm.nih.gov"), which are grouped in 50 different bacterial phyla and 11 different archaeal phyla³¹. Approximately 90 % of the bacterial genomes accessible the public are classified as draft genomes (*i.e.* genomic sequence not constructed as a chromosome). A combination of second-generation and third-generation sequencing technologies can produce complete genomes. However, an evaluation study revealed that there is an insignificant loss of genetic information in draft genomes when compared to complete genomes³². The largest sequenced bacterial genome is from *Sorangium cellulosum* with a size of 14 Mb that includes 12,000 genes, while the smallest genome is from *Candidatus 'Nasuia deltocephalinicola'* with size of 112 kb and 137 genes³³. Most submitted genomes (draft and complete) are from six phyla *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Firmicutes*, *Proteobacteria* and *Spirochaetes* (Figure 1.2). The number of species as well as the number of available genomes will drastically increase with metagenome sequencing and the pioneer computational algorithms for binning microbial genomes in the datasets.

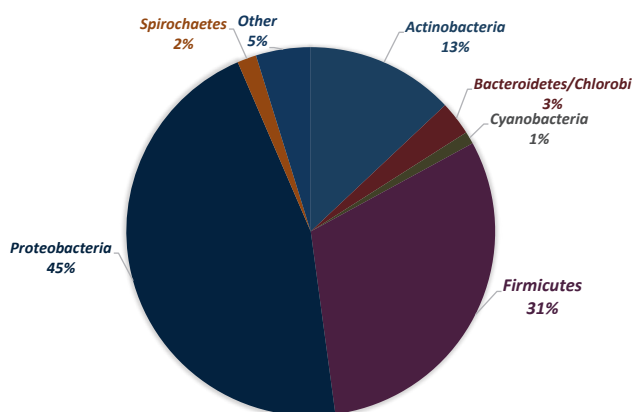


Figure 1.2 | Percentage of sequenced genomes of bacterial species from different phyla.

The genetic content of the bacterial communities contains key-factors for complex phenotypes, including metabolic traits, and can guide physiological analysis^{34,35}. Classic targets for genomic screening are enzymes with novel or improved activity, such as esterases, decarboxylases and phosphatases, which may have industrial applications³⁶. Metabolic cascades may be clustered in genomic synteny³⁷ and genomic comparison of these can strengthen the conclusions for individual and common phenotypes in the species. Additionally, comparative genomics can indicate the heterogeneity and diversity in genotypes accumulated over numerous years of adaptation³⁸. Essential criteria should be fulfilled for a precise comparison: an accurate functional annotation of genes and usage of available physiological information. In recent years, genomic annotation has mainly been based on homology using search algorithms such as BLAST¹⁶. However, annotated genes in the databases may be mischaracterized, thereby reducing the accuracy of newly described genes. To obtain a robust annotation, a combination of two different annotation methods should be conducted. A possible combination can be an annotation method based on SEED database (RAST platform)³⁹ and an annotation method based on functional domains. Comparative genomics can additionally contribute to phylogenomics and identifying conserved

genes. A recent study using comparative genomics of all available bacterial genomes defined 500 conserved genes in the bacterial species and conducted a phylogenomics clustering⁴⁰.

Functional omics

To achieve full characterisation of complex biological systems, multiple layers of omics studies should be implemented. Genomics offers a complete overview of the metabolic potential of each organism but provides no insights into the cell dynamics. To have a real snapshot of the presence of metabolic enzymes functional omics studies – proteomics and transcriptomics – must be applied. Proteomics is the analysis of the complete available repository of proteins in the cell under specified conditions or in a specific time frame. Comparative proteomics analysis is conducted by peptide identification, measuring the peptide abundances, modifications and variations under multiple conditions⁴¹. Current research studies tend to use shotgun MS-based methods, which are suitable for studying complete proteomes⁴². Proteomics can identify enzymes connected to valuable metabolic properties of species. For instance, by analysing the proteome of *Sphingomonas wittichii* RW1 bacterium, 56 ring-hydroxylating dioxygenases were identified involved in the degradation of chlorinated congeners of dibenzofuran, an ordinary dispersed pollutant⁴³. Recently, the use of comparative proteomics discovered non-annotated cold shock proteins in *E. coli*, which actively contribute to the regulation of metabolism at low temperatures⁴⁴.

Transcriptomics and metabolomics complement other omics studies. Transcriptomics is the quantification of the expression of all transcribed genes at a specific time point⁴⁵. In comparison, transcriptomics provides a better overview than proteomics regarding expression quantification as many genes do not translate into enzymes and many proteins are difficult to quantify e.g. due to post-translation modifications and membrane bounds⁴⁶. Metabolomics directly investigates directly all intercellular and extracellular metabolites generated by the microorganism⁴⁷. Depending on the complexity of the metabolite, metabolomics can be based on mass spectrometry or nuclear

magnetic resonance. Due to the excessive costs, a quantification by high performance liquid chromatography or gas chromatography is frequently sufficient for specific pathways.

GENUS *TRICHOCOCCUS*

For the purpose of this PhD research, *Trichococcus* species were considered an equitable choice for potential bioprospecting. *Trichococcus* species derive from diverse habitats, including even environments with sub-zero temperatures. Within the genus *Trichococcus*, two species were associated with having a potential biotechnological value: *Trichococcus* strain ES5⁴⁸, able to produce 1,3-propanediol, and *T. patagoniensis*, able to produce a mucoid substance during growth in freezing temperatures⁴⁹. Studying *Trichococcus* species using multiple omics approaches may answer questions regarding their physiology and their potential for biotechnological exploitation.

The name *Trichococcus* (*trixa*: hair, *kókkos*: coccus) derives from the shape of members of this genus, which is cocci with aggregations creating structures of 'hair-like' flocci⁵⁰ (Figure 1.3). All members are facultative anaerobic Gram-positive bacteria and they are not observed to form spores⁵⁰.

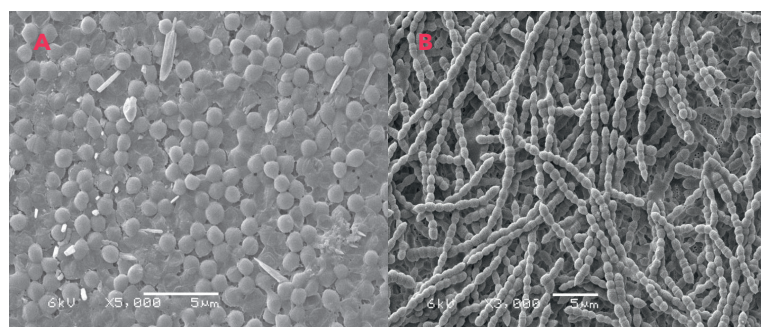


Figure 1.3 | SEM microscopy of: A. *Trichococcus* strain ES5 B. *Trichococcus pasteurii*. In *Trichococcus* strain ES5 hair-like filamentous structures are observed like the ones observed in the type strain of the genus (*T. flocculiformis*). These aggregations inspired the naming of that species.

1,3-propanediol production by specific *Trichococcus* species

Due to the depletion of fossil fuels, sustainable energy generation has gained extensive focus during the last few decades. Biodiesel is one the major contributors of renewable energy. Glycerol is a by-product of the biodiesel production process and its valorisation would contribute to the overall economic feasibility of biodiesel production. *Trichococcus* strain ES5 is biotechnologically interesting as it can ferment glycerol to 1,3-propanediol (1,3-PDO). Amongst other useful products, 1,3-PDO is used as monomer for composing the polyester polytrimethylene terephthalate (PTT)⁵¹. PTT is completely biodegradable, and it has a good washfastness⁵², characteristics that render it as a commercially valuable polyester, especially for the manufacture of fabric fibres and textiles⁵².

Microbial glycerol fermentation starts with glycerol dehydratase converting glycerol to 3-hydroxypropionaldehyde (3-HPA). The 3-HPA intermediate is usually toxic for bacterial cells and thus is used as an antimicrobial compound with the commercial name reuterin (from *Lactobacillus reuteri*)⁵³. Despite 3-HPA toxicity, the bacteria are able to reduce 3-HPA to 1,3-propanediol by 1,3-propanediol dehydrogenase (Figure 1.4)⁵². Glycerol dehydratase can be vitamin B12-dependent or B12-independent, depending on the host organism⁵⁴. Most of the known microorganisms encode the B₁₂-dependent form. Thus far, only *C. butyricum* has been observed to use B₁₂-independent glycerol dehydratase. The variant forms of glycerol dehydratase are activated by different glycerol dehydratase activation enzymes. In *Trichococcus* strain ES5 and *T. pasteurii*, B12-dependent glycerol dehydratase is identified in the genome. *Trichococcus* strain ES5 has a production yield of around 65% ($\text{mol}_{1,3\text{-PDO produced}} / \text{mol}_{\text{glycerol consumed}}$) 1,3-PDO, which is comparable with other well-studied natural producers such as *Klebsiella pneumonia*, *Clostridium butyricum* and *Clostridium pasteurianum*^{55,56,57}.

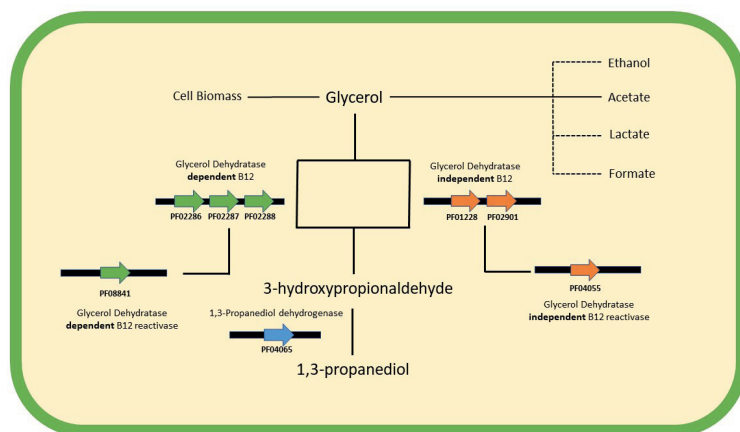


Figure 1.4 | Metabolic pathway of glycerol fermentation to 1,3-PDO and related genes/domains.

Investigating the 1,3-PDO metabolic trait in *Trichococcus* species can provide information regarding the metabolic trait composition and the physiological conditions necessary for improving 1,3-PDO production. This investigation can be expanded to the thousands of genomic sequences stored in the public genomic repositories enriching the knowledge of the 1,3-PDO trait. By screening the numerous microbial genomes, novel candidates for 1,3-PDO production may be identified. It could be expected that by analysing all available genomes more bacteria with B12-independent glycerol dehydratase will be identified. The use of omics studies in databases can exploit the metabolic trait and indicate the conserved genes that are essential to produce 1,3-PDO.

***T. patagoniensis*, a bacterium with “extraterrestrial” properties**

T. patagoniensis was isolated by scientists at National Aeronautics and Space Administration (NASA) in a project for studying microbes on Earth that represent life on other planets. *T. patagoniensis* is a psychrotolerant anaerobic bacterium, isolated from a solid guano collected from a lake in Chilean Patagonia⁴⁹. Survival mechanisms that provide tolerance to high salinity (6,5 % NaCl in the medium - double

of the sea water salinity) and low temperature (growth at minus 5 °C) can qualify *T. patagoniensis* as a bacterial model for studying cold adaptation mechanisms and the production of metabolites at low temperatures (Figure 1.5). So far, there are 46 published and 37 drafts sequenced genomes of psychrophiles and psychrotolerant species⁵⁸. These microorganisms can produce compounds tailored for resisting water freezing temperatures, and therefore have applications in bioindustry. Survival at low temperatures is usually characterised by the existence of cold shock domains (CSD)⁵⁹. Other mechanisms exist for protecting the cell from freezing and for functioning at sub-zero temperatures, such as proteome adjustments, optimized transcription system, reinforced membrane wall and protective compounds. In general, for cold adaptation, the complete proteome must be adjusted with high flexibility, decreased thermostability and improved activity⁵⁹. The proper folding of RNA molecules is impaired at lower temperatures, hence specific RNA-protein interactions aim to facilitate proper RNA folding. DEAD-box proteins serve to unwind RNA duplexes and displace proteins from ribonucleoprotein complexes⁶⁰. The cell wall must maintain membrane fluidity and structural integrity in the freezing environment⁶¹. Psychrophilic adaptations include an increased amount of polyunsaturated to saturated fatty acids ratio in the membrane phospholipids and alterations in lipid class compositions^{62,63}. Exopolysaccharides are another form of protection against cold. EPS are exocellular polysaccharides that contribute to biofilm formation. At a low temperature, psychrophilic bacteria excrete excessive EPS amounts which surround the cell and alter the physiochemical environment, creating a protective layer against mechanical pressure and cold^{64,65}. A mucoid substance has already been observed to be formed by *T. patagoniensis*⁴⁹. This substance is surrounding *T. patagoniensis* cells forming a protecting layer at zero and sub-zero temperatures. The antifreezing capabilities of EPS may be of interest for biotechnological applications. Frequently, antifreezing compounds are used in numerous applications like in food bioindustry, agriculture (biofertilizer for increasing resistance of plants against the cold) and medicine (cryopreservation of cells)^{65,66}.

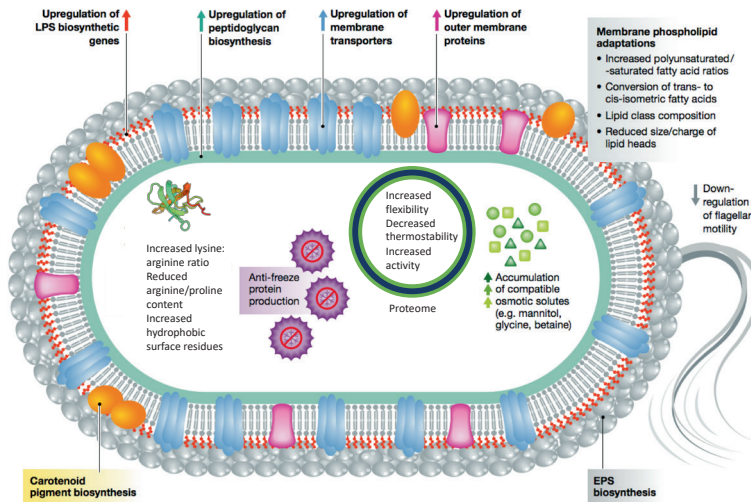


Figure 1.5 | Cold adaptation mechanisms in psychrophilic and psychrotolerant species (picture was adapted⁵⁹ and modified-license number 4278750226876).

***Trichococcus* taxonomy requires alternative taxonomical markers**

At the beginning of this PhD research, there were just five *Trichococcus* species described; *Trichococcus flocculiformis*⁶⁷, *Trichococcus collinsi*⁵⁰, *Trichococcus palustris*⁵⁰, *Trichococcus patagoniensis*⁵⁸ and *Trichococcus pasteurii*⁵⁰. Since the creation of the genus in 1984 with *T. flocculiformis*, other older *Trichococcus* members were re-classified as species from the genera *Lactosphaera* and *Ruminococcus*. Two independent research groups in Germany, Scheff *et al.* and Schink *et al.*, described *T. flocculiformis* and *T. pasteurii*, respectively, but the first was set in the genus *Trichococcus*, whereas *T. pasteurii* was classified as *Ruminococcus pasteurii* and ten years later was described as *Lactosphaera pasteurii*^{50,68}. Before 16S rRNA gene sequencing, laboratories used their experience and criteria for determining species as mentioned in the Bergey's manual for determinative bacteriology (solely based on morphological and physiological properties)⁶⁹. Apparently, *T. pasteurii* was a challenging bacterium and the bacterial systematics were misleading. The classification problem of the *Trichococcus* species was not solved even after the analysis of the 16S rRNA gene taxonomical marker.

The revision of taxonomy for previous *Trichococcus* species signifies the need to establish a better classification approach for the genus. Members of the *Trichococcus* genus have a highly conserved 16S rRNA gene sequence (99-100%). Therefore, 16S rRNA gene classification cannot be the only applied taxonomical marker, especially for the additions of new species. A complete genome utilizing approach can be a more suitable strategy. In the last year, two additional species were added by using genomic information as reported here; *Trichococcus paludicola* and *Trichococcus alkaliphilus*⁷⁰.

THESIS OUTLINE

In this PhD research, we focused on bioprospecting, using *Trichococcus* species as model case. Interest in *Trichococcus* research was generated by the fact that several strains can produce added-value metabolites, such as 1,3-PDO, and anti-freezing compounds. We used omics studies to approach this unexplored genus and investigated their genotype and phenotype. Proteomics revealed further metabolic properties for cold adaptation mechanisms.

In **Chapter 1**, an introduction to the research challenge, topic and approach of the research outlined. The importance of bioprospecting for microbial added-value products by looking to uninvestigated microorganisms, specifically *Trichococcus* species, is described, and how omics studies can be used to aid in bioprospecting is further explained.

In **Chapter 2**, a novel *Trichococcus* species, *Trichococcus ilyis* (strain R210) is characterized using a genome-guided approach. *Trichococcus* strain R210 was previously isolated from a sulfate-reducing bioreactor, but taxonomic classification was unresolved, because 16S rRNA gene identity with close relatives was high. Within this thesis, the genomes of all *Trichococcus* isolates, including the genome of *Trichococcus* R210, were sequenced. This made it possible to apply dDDH and ascertain the classification of *Trichococcus* strain R210 as a new species of the genus *Trichococcus*, *Trichococcus ilyis*.

In **Chapter 3**, glycerol conversion to 1,3-PDO by *Trichococcus* strain ES5 and *T. pasteurii* was explored, together with a general genome comparison of all *Trichococcus* species genomes. dDDH and functional protein domains were used for the taxonomical and functional clustering of the different species within this genus. Peculiar or unique metabolic properties, and characteristics such as 1,3-PDO production, survival of the cold and high salinity tolerance, were researched in all the genomes. A complete operon dedicated to 1,3-PDO production was observed; this operon included 17 genes with multiple glycerol transporters and a B12-dependent glycerol dehydratase activator. The operon was identical in *Trichococcus* strain ES5 and *T. pasteurii*, both able to ferment glycerol to 1,3-PDO. Only four species could grow at 0

°C, although all seven-species contained a high number of cold shock domains. High salinity survival may be correlated with the genomic evidence of osmoregulation and all species could tolerate high NaCl concentration except *T. pasteurii*, which that lacked the related genes.

In **Chapter 4**, all bacteria for which a sequenced genome is available in the public domain were researched for genes involved in 1,3-PDO production. The analysis was based on the model of the complete glycerol operon previously identified in *Trichococcus* species in Chapter 3. Bacterial genomes were mined for the trait and an unbiased, collective and accurate database was generated following the “Knowledge and Discover in Databases” (KDD) approach. This easily queried semantic database provides information for 1,3-PDO production including essential and accessory genes as well as physiological characteristics, such as anaerobicity, pathogenicity, B12 synthesis and salinity tolerance. All bacterial genomes in the database were screened for B12-dependent and B12-independent glycerol dehydratases and 1,3-PDO dehydrogenases. Until then, *C. butyricum* had been the only bacterium with a vitamin B12-independent glycerol dehydratase, but other potential candidates were identified: e.g. *Clostridium diolis*, *Clostridium beijerinckii*, *Raoultella planticola*. In addition, 187 candidates were identified that have a B12-dependent pathway for glycerol fermentation. Bacteria predicted to produce 1,3-PDO were experimentally validated for this trait.

In **Chapter 5**, glycerol fermentation was investigated to gain insight into the biotechnological potential of *Trichococcus* and comparisons were made with industrially applied strains. Both, *Trichococcus* strain ES5 and *T. pasteurii*, had a similar 1,3-PDO yield to *Clostridium butyricum* VPI 3266. Proteomics was further used to study the effect of high glycerol concentration, specifically considering the accumulation of 3-hydroxypropionaldehyde (3-HPA) with increasing substrate concentrations. 3-HPA production negatively influenced 1,3-PDO production, being the bottleneck of the process. To achieve a reproducible and a steady yield of 1,3-PDO production from the *Trichococcus* strains, we established continuous anaerobic bioreactors with initial glycerol concentrations of 1,8 g/L, 3,7 g/L and 23 g/L.

In **Chapter 6**, the psychrotolerant *T. patagoniensis* was studied to understand the adaptive mechanisms involved in cold resistance. Comparative proteomics was conducted in cultures grown at 30 °C and 0 °C. A significant upfold change in the abundance of cold shock proteins was observed at 0 °C. Interestingly, proteins related with translation mechanisms, such as ribosomal proteins, were highly abundant in the proteome of cultures grown at 0 °C and DEAD-box protein was observed only at the low temperature. We further investigated the EPS content of *T. patagoniensis* cultures grown at 30 °C and 0 °C by Fourier-transform infrared spectroscopy (FTIR). The EPS was identified as the fructose polymer inulin. This fatty acid is known to protect the cells from freezing by covering them while preventing the freezing of water molecules. Finally, we observed a higher amount of unsaturated cell lipids at 0 °C, which supports the flexibility of the membrane wall.

In **Chapter 7**, we described a novel *Trichococcus* species, *Trichococcus shcherbakoviae* strain Art1. *Trichococcus shcherbakoviae* was isolated from an anaerobic bioreactor operated at low temperature (3-8 °C). This new member of the *Trichococcus* genus had a conserved 16S rRNA gene sequence like the rest of the species in the genus. The closest species based on sole use of the 16S rRNA gene was *T. pasteurii*. We implemented genomic based methods, dDDH and ANI for an accurate phylogenomics. *Trichococcus* strain Art1 comparison with other *Trichococcus* strains resulted in dDDH values lower than 70 % and of ANI values lower than 94 %. Hence it was classified as a new strain of the genus *Trichococcus*.

In **Chapter 8**, the research described in the previous chapters is discussed in a broader context, and the main overall conclusions are presented. The aim and the scope of this thesis are put in perspective and compared with similar approaches in the literature. In addition, suggestions for improving omics studies approaches are mentioned and the future of bioprospecting is considered.

CHAPTER 2

DESCRIPTION OF *TRICHOCOCCUS ILYIS* SP. NOV. BY COMBINED PHYSIOLOGICAL AND *IN SILICO* GENOME HYBRIDISATION ANALYSES

Nikolaos Strepis^{1,2}, Irene Sánchez-Andrea¹, Antonie H van Gelder¹, Henri van Kruistum¹, Nicole Shapiro³, Nikos Kyrpides³, Markus Göker⁴, Hans-Peter Klenk^{4,5}, Peter Schaap², Alfons JM Stams^{1,6}, Diana Z Sousa¹

1. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

2. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands

3. DOE Joint Genome Institute, Walnut Creek, CA, USA

4. Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

5. School of Biology, Newcastle University, Newcastle, UK

6. Centre of Biological Engineering, University of Minho, Braga, Portugal

Chapter adapted from publication:

International Journal of Systematic and
Evolutionary
Microbiology (2016)
doi: 10.1099/ijsem.0.001294

ABSTRACT

Trichococcus species share high similarity of their 16S rRNA gene sequences (> 99 %). Digital DNA-DNA hybridisation values (dDDH) among all described *Trichococcus* species (*T. flocculiformis* DSM 2094^T, *T. pasteurii* DSM 2381^T, *T. collinsii* DSM 14526^T, *T. palustris* DSM 9172^T, and *T. patagoniensis* DSM 18806^T) indicated that *Trichococcus* sp. strain R210^T represents a novel species of the genus *Trichococcus*. The dDDH values showed a low DNA relatedness between strain R210^T and all other *Trichococcus* spp. (23–32%). Cells of strain R210^T were motile, slightly curved rods with sizes 0.63–1.40 x 0.48–0.90 µm and stained Gram-positive. Growth was optimal at pH 7.8 and at temperature of 30 °C. Strain R210^T could utilise several carbohydrates, and the main products from glucose fermentation are lactate, acetate, formate, and ethanol. The genomic DNA G+C content of strain R210^T was 47.9 %. Based on morphological, physiological, biochemical characteristics along with measured dDDH values for all *Trichococcus* spp., it is suggested that strain R210^T represents a novel species within the genus *Trichococcus*, for which the name *Trichococcus ilyis* is proposed. The type strain of *Trichococcus ilyis* is R210^T (=DSM 22150^T =JCM 31247^T).

Trichococcus sp. strain R210^T was previously isolated from a sulfate-reducing bioreactor treating municipal wastewater mixed with an industrial stream containing citrate⁷¹. Isolation was performed in basal medium supplemented with 10 mM citrate as sole energy and carbon source (without sulfate). The main fermentation products of citrate were acetate and formate, whereas the main products of glucose fermentation were lactate, ethanol, acetate and formate. This product profile is typically observed in heterolactic fermenters⁷².

To date, five *Trichococcus* species have been validly named, namely *T. flocculiformis* (DSM 2094^T)⁷³, *T. pasteurii* (DSM 2381^T)⁷⁴, *T. collinsii* (DSM 14526^T)⁶⁷, *T. palustris* (DSM 9172^T)⁷⁵, and *T. patagoniensis* (DSM 18806^T)⁴⁹. All these *Trichococcus* species can use a broad range of sugars and polyols for growth. A peculiar feature of *Trichococcus* strains is the ability to grow over a wide range of temperatures, from close to 0 °C to over 45 °C. This makes these strains particularly interesting for biotechnological applications at mildly low temperature. They are also tolerant to oxygen, which is advantageous because it eliminates the need of maintaining bioreactors completely oxygen-free.

Morphologically, members of the genus *Trichococcus* are pleomorphic (regular cocci, olive-like rods) and may have several cell arrangements (cells can grow individually, in pairs, in long chains, or in grape-like conglomerates)⁶⁷. The similarity of the 16S rRNA gene sequence of described *Trichococcus* species (99-100 %) is above the threshold for the definition of a new species¹³. However, DNA-DNA hybridisation (DDH) values among them are lower than 70 %, the threshold for the delineation of new species⁴⁹. The presence of highly conserved 16S rRNA genes in different species is not so uncommon. Another example can be found in *Bacillus*, where some species can have a 16S rRNA gene similarity of 97-100 %⁷⁶. For these cases, DDH is accepted as the gold standard for complete taxonomic classification and proposal of new species. Experimental DDH assays are based on optical measurements and prone to experimental errors⁷⁷. The rapid progress in genome sequencing allows utilisation of computational methods for genome comparison and taxonomic differentiation⁷⁸. Genome Blast Distance Phylogeny (GBDP) approaches have been optimised for calculation of digital equivalents for DDH values (dDDH), which are analogous to DNA-DNA hybridisation⁷⁹.

To clarify the phylogenetic assignment of strain R210^T, the genome was sequenced along with the genomes of all five previously described *Trichococcus* species and calculated dDDH values between these species using GBDP. Consequently, genome-guided physiological characterisation was performed.

T. flocculiformis (DSM 2094^T), *T. pasteurii* (DSM 2381^T), *T. collinsii* (DSM 14526^T), *T. palustris* (DSM 9172^T), and *T. patagoniensis* (DSM 18806^T) were obtained from the German Collection for Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany) and cultivated on DSMZ recommended media. Strain R210^T was previously isolated by Stams *et al.*⁷¹, and since then was preserved in our laboratory. Strain R210^T was grown at 30 °C on an anaerobic bicarbonate-buffered medium⁸⁰ supplemented with 0.1 g L⁻¹ yeast extract and 20 mM D-glucose.

Genomic DNA of all the strains was extracted using MasterPure Gram positive DNA purification Kit (Epicenter, Madison, WI) according to manufacturers' instructions. The genomes of strain R210^T, *T. pasteurii*, *T. flocculiformis*, *T. collinsii*, and *T. palustris* were sequenced at the JGI (Walnut Creek, CA) using a Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA) within the Genomic Encyclopaedia of archaeal and bacterial type strains, phase II⁸¹. The genome of *T. patagoniensis* was sequenced at Baseclear (Leiden, The Netherlands) using an Illumina HiSeq2500 platform (Illumina Inc., San Diego, CA).

The genomes of strain R210^T (9,304,840 reads and 151 bp read length), *T. pasteurii* (10,101,940 reads and 151 bp read length), *T. flocculiformis* (10,454,276 reads and 151 bp read length), *T. collinsii* (9,948,768 reads and 151 bp read length), *T. palustris* (11,006,414 reads and 151 bp read length), and *T. patagoniensis* (5,740,601 reads and 126 bp read length) were assembled using a pipeline comprising: Ray⁸², to generate an initial assembly, followed by Opera⁸³, for genome scaffolding, and CAP3⁸⁴, for assembling optimisation. For Ray assembler, the optimal kmer size was calculated with KmerGenie⁸⁵. In addition, the genomes were assembled by JGI (Walnut Creek, CA) using ALLPATHS-LG⁸⁶ and Velvet⁸⁷. Both assemblies were merged and improved with SSPACE⁸⁸.

The dDDH values (including confidence intervals, CIs) between strain R210^T and the established *Trichococcus* type strains were measured using the Genome-to-Genome Distance Calculator available at <http://ggdc.dsmz.de>¹⁷.

Genome annotation of strain R210^T was performed using an in-house pipeline based on published tools and databases: Prodigal⁸⁹, InterProScan⁹⁰, tRNAscan-SE⁹¹, RNAmmer⁹², UniRef50⁹³, Swiss-Prot database⁹⁴, and combined with RAST³⁹. The merged results from the in-house annotation pipeline and RAST were used to curate the annotation of the genome by completing the full description of the encoding regions. The genomic sequences, as well the coding sequences, were used as input for BRIG (Blast Ring Image Generator)⁹⁵. Sequences were assigned with KEGG orthology (KO) identifiers using KEGG Automatic Annotation Server (KAAS) and further linked to metabolic pathways⁹⁶.

16S rRNA gene sequence analysis was performed with ARB v6.0⁹⁷ using the non-redundant SILVA database⁹⁸. 16S rRNA gene sequences of strain R210^T and all the currently isolated *Trichococcus* spp. were implemented in the all-species living tree project⁹⁹. Three phylogenetic trees were generated based on maximum-likelihood (RAxML algorithm), neighbour-joining (Jukes-Cantor correction) and maximum parsimony, and a consensus tree was generated with the tool integrated into ARB v6.0.

Cell morphology of strain R210^T was observed using a phase-contrast microscope (Leica DM2000 microscope, Leica Microsystems, The Netherlands) on glucose-grown cultures. Cell sizes were measured using the software Fiji-ImageJ¹⁰⁰. Growth was estimated by measurements of optical density (at 600 nm) with a spectrophotometer (Hitachi U-1500, Labstuff, The Netherlands).

The cellular fatty acids of strain R210^T were analysed at DSMZ (Braunschweig, Germany). The analysis was done with the system of Sherlock MIS (MIDI Inc, Newark, NJ) using previously described methods^{101,102}. Cells for lipids analysis were cultivated in pre-reduced tryptic soy broth (BD, Franklin Lakes, NJ) as previously used for other *Trichococcus* strains⁶⁷.

The growth of strain R210^T was tested for a temperature range from 0 °C to 45 °C and a pH range from 6.0 to 9.6. The ability of strain R210^T to utilise a set of substrates had been previously tested⁴⁸. In this study, additional substrates were tested (at a concentration of 20 mM, except mentioned otherwise): L-lactose, sucrose, D-galactose, D-fructose, D-glucose, D-arabinose, formate, D-galacturonate, D-glucuronate, D-mannose, L-rhamnose, raffinose, L-sorbose, D-gluconate, D-tagatose, D-xylose, starch, glycerol, methanol, ethanol, D-sorbitol, inositol, D-dulcitol and H₂/CO₂ [80/20 (v/v), 1.5 x 10⁵ Pa]. Substrates for testing were selected based on genomic information, i.e. the presence genes potentially encoding for the degradation of those substrates. Soluble substrates and intermediates (sugars and volatile fatty acids) were measured with a Thermo Electron HPLC system equipped with an Agilent Metacarb 67H column (300 mm x 6.5 mm) (Thermo, Waltham, MA) and a refractive index detector. The mobile phase used was sulfuric acid (0.01 N) at a flow rate of 0.8 mL min⁻¹. The column temperature was set at 45 °C. Hydrogen and CO₂ were measured with a Shimadzu GC-2014 gas chromatograph equipped with a Molsieve 13X column (2 m x 3 mm) (Shimadzu, Kyoto, Japan) and a thermal conductivity detector (TCD). Argon was used as carrier gas at a flow rate of 50 mL min⁻¹, and temperatures in the injector, column and detector were 80, 100 and 130 °C respectively.

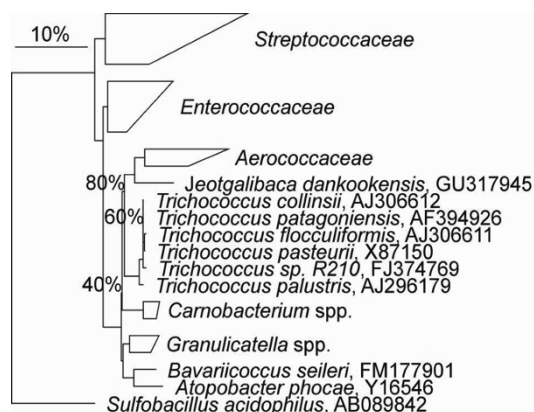


Figure 2.1 | Consensus tree of 16S ribosomal RNA gene sequences showing the phylogenetic affiliation of strain R210^T.

Phylogenetic analysis of *Trichococcus* spp. based on 16S rRNA gene sequence indicates a high similarity (> 99 %) among all the species of the genus (Figure 2.1, Table 2.1). The closest relative of strain R210^T is *T. pasteurii* with a 16S rRNA gene sequence similarity of 99.5 % (Table 2.1). The 16S rRNA gene sequence of *T. pasteurii* is 99.9 % similar to the ones of *T. collinsii* and *T. patagoniensis*, while *T. patagoniensis* and *T. collinsii* have an identical 16S rRNA gene sequence (100 % similarity). This indicates that the 16S rRNA gene of *Trichococcus* spp. has not evolved and taxonomic differentiation of species within this genus needs to be based on DDH. In this work we calculated dDDH values of strain R210^T with other *Trichococcus* type strains and values were all lower than 70 % (Table 2.1), which is accepted as the threshold value for the definition of new species¹⁷. Based on this, strain R210^T can certainly be classified as a new species of the genus *Trichococcus*.

Table 2.1 | dDDH values in percentage with confidence intervals between type strains of *Trichococcus* species (calculated using the GGDC tool, <http://ggdc.dsmz.de/distcalc2.php>). Percentage of 16S rRNA gene sequence similarity between *Trichococcus* spp. is shown in parenthesis.

Strain	Strain R210 ^T	<i>T. pasteurii</i> DSM 2381 ^T	<i>T. floccu- liformis</i> DSM 2094 ^T	<i>T. collinsii</i> DSM 14526 ^T	<i>T. palustris</i> DSM 9172 ^T	<i>T. patago- niensis</i> DSM 18806 ^T
R210 ^T	100±0.0 (100)	32±3.0 (99.5)	30±3.0 (99.5)	30±3.0 (99.7)	23±2.8 (99.1)	31±2.5 (99.7)
<i>T. pasteurii</i> DSM 2381 ^T	32±3.0 (99.5)	100±0.0 (100)	48±3.3 (99.8)	40±3.1 (99.9)	23±2.9 (99.1)	39±2.5 (99.9)
<i>T. flocculiformis</i> DSM 2094 ^T	30±3.0 (99.5)	42±3.1 (99.8)	100±0.0 (100)	36±3.0 (99.7)	23±2.8 (99.0)	36±2.4 (99.7)
<i>T. collinsii</i> DSM 14526 ^T	30±3.0 (99.7)	40±3.2 (99.9)	36±3.0 (99.7)	100±0.0 (100)	23±2.9 (99.3)	41±2.5 (100)
<i>T. palustris</i> DSM 9172 ^T	23±2.8 (99.1)	23±2.9 (99.1)	23±2.8 (99.0)	23±2.0 (99.3)	100±0.0 (100)	23±2.5 (99.3)
<i>T. patagoniensis</i> DSM 18806 ^T	31±2.5 (99.7)	39±2.5 (99.9)	36±2.4 (99.7)	41±2.5 (100)	23±2.5 (99.3)	100±0.0 (100)

Table 2.2 | Morphological and physiological characteristics of described species of the genus *Trichococcus*.

Characteristic	1	2	3	4	5	6
Source	Sulfate-reducing bioreactor* ^a	Septic pit ^b	Activated sludge ^b	Hydrocarbon-spill site ^b	Swamp ^b	Guano penguin ^c
Gram stain	+	+ ^b	+ ^b	+ ^b	+ ^b	Variable ^c
Cell length (µm)	0.63-1.40	1.0-1.5 ^b	1.0-2.5*	1.0-2.5 ^b	1.0-2.5 ^b	1.3-2.0 ^c
Cell shape	Cocci	Short chain of cocci ^b	Long chain of cocci ^b	Rods ^b	Cocci ^b	Cocci ^c
pH range (optimum)	6.0-9.6 (7.8)	5.5-9.0 ^b	5.8-9.0 ^b	6.0-9.0 ^b (7.5)	6.2-8.4 ^b (7.5)	6.0-10 ^c (8.5)
Temperature range (optimum) (°C)	4-40 (30)	0-42 ^b (25-30)	4-40 ^b (25-30)	7-36 ^b (23-30)	0-33 ^b	-5-35 ^c (28-30)
DNA G+C content (mol%)						
<i>In vitro</i>	ND	45.0 ^b	48.0 ^b	47.0 ^b	48.0 ^b	45.8 ^c
<i>In silico</i>	47.9	45.7	44.3	44.3	45.9 ^x	46.9
Substrate utilization						
Formate	-	-	ND	ND	ND	- ^c
L-Lactate	- ^a	- ^b	- ^a	- ^c	+ ^c	- ^c
L-Malate	+ ^a	+ ^b	- ^a	+ ^b	- ^b	+ ^c
d-Gluconate	-	+	ND	ND	ND	ND
Glycerol	-	+	-	-	-	- ^c
d-Arabinose	+ ^a	+	+ ^a	ND	ND	+ ^c
d-Xylose	-	+	+ ^a	ND	ND	ND
d-Galactose	+	+	ND	ND	ND	ND
L-Rhamnose	+	+	ND	ND	ND	ND
d-Fructose	+ ^a	+	+ ^a	ND	ND	+ ^c
d-Lactose	+	+ ^b	+ ^b	- ^b	+ ^b	- ^c
Maltose	+ ^a	+ ^b	+ ^a	- ^b	+ ^b	+ ^c
d-Mannitol	+ ^a	+ ^b	- ^a	+ ^b	+ ^b	+ ^c
Sucrose	+ ^a	+	+ ^a	ND	ND	+ ^c
Methanol	-	-	ND	ND	ND	- ^c
Ethanol	-	-	ND	ND	ND	- ^c
d-Sorbitol	-	+	- ^a	+ ^b	+ ^b	ND

Strains: 1, R210^T; 2, *T. pasteurii* DSM 2381^T; 3, *T. flocculiformis* DSM 2094^T; 4, *T. col-linsii* DSM 14526^T; 5, *T. palustris* DSM 9172^T; 6, *T. patagoniensis* DSM 18806^T. Data are from this study unless otherwise indicated. All strains can grow on glucose and pyruvate, citrate (ND for *T. palustris*) and cellobiose (ND for *T. patagoniensis*). All strains did not grow with H₂ /CO₂. Strain R210^T and *T. pasteurii* can grow on D-glucuronate, D-mannose and raffinose and cannot grow on D-galacturonate, starch, D-tagatose, L-sorbose, inositol and dulcitol. +, Positive; -, negative; ND, Not determined. *Data from: a, van Gelder *et al* ⁴⁸; b, Liu *et al*.⁶⁷; c, Pikuta *et al*.⁴⁹.

Substrate utilisation of strain R210^T, and comparison with other *Trichococcus* species, is presented in Table 2.2. Strain R210^T can use several carbohydrates, which is corroborated by the presence of numerous putative genes from the Embden-Meyerhof-Parnas (EMP) pathway, pentose phosphate pathway, and phosphoketolase pathway. The main products of glucose fermentation by strain R210^T were acetate, formate, lactate and ethanol. An operon with genes coding for rhamnulokinase (EC 2.7.1.6), L-rhamnose isomerase (EC 5.3.1.14), rhamnulose-1-phosphate aldolase (EC 4.1.2.19), lactaldehyde dehydrogenase (EC 1.2.1.22) was identified in strain R210^T (TR210_1063-1066). Lactaldehyde dehydrogenase is an essential enzyme for the conversion of rhamnose to 1,2-propanediol¹⁰³, which was observed as a main product in the experimental growth tests of strain R210^T on rhamnose (together with acetate and formate). An additional operon potentially linked to rhamnose metabolism (TR210_973-982) contained genes encoding for L-rhamnose mutarotase, which may play a role in the first steps of rhamnose degradation¹⁰³, together with tripartite ATP-independent periplasmic (TRAP) transporters, which may be involved in substrate import to the cell¹⁰⁴. A mannose-specific IIC component gene (EC 2.7.1.69, TR210_1375) from the phosphotransferase system (PTS) is also present in the genome of strain R210^T. The PTS system is responsible for the uptake of carbohydrates and includes the mannose-specific IIC component, which can catalyse mannose to mannose 6-phosphate, during uptake¹⁰⁵. Sucrose was also metabolised by strain R210^T; genes encoding for beta-glucuronidase (EC 3.2.1.31, TR210_689 and TR210_756), maltose-6'-phosphate glucosidase (EC 3.2.1.122, TR210_416) and glycogen operon protein (EC 3.2.1.-, TR210_2418) were identified (Figure 2.2).

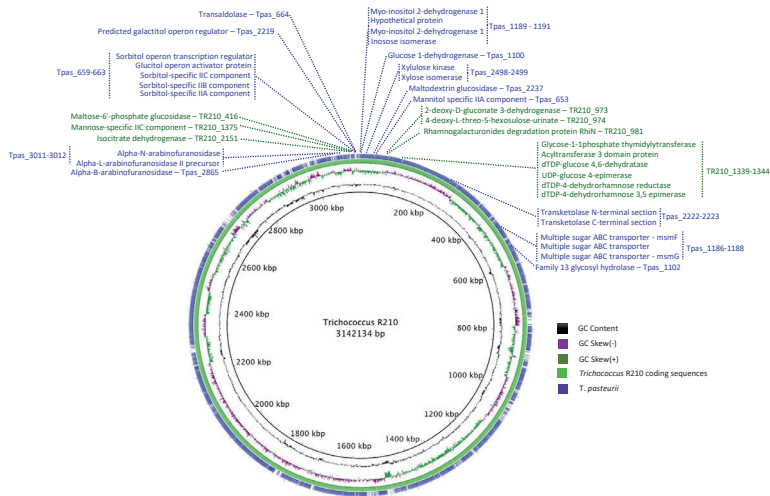


Figure 2.2 | BRIGG image showing the genomic differences and regions of strain R210^T and *T. pasteurii*. The genes present only in the genome of strain R210^T are shown in green, while the genes that exist only in *T. pasteurii* are shown in blue. Genomic synteny is observed for carbohydrates degrading genes.

Strain R210^T is a citrate fermenting bacterium and its genome contains some, but not all, genes of the tricarboxylic acid cycle (TCA). The gene encoding for isocitrate dehydrogenase (EC 1.1.1.42), involved in the conversion of isocitrate (via oxalosuccinate) to 2-oxoglutarate, is present in the genome (TR210_2151), but also a NAD⁺-dependent isocitrate dehydrogenase (EC 1.1.1.41, TR210_901). The latter enzyme is capable of direct converting of isocitrate to 2-oxoglutarate.

T. pasteurii is the closest relative of strain R210^T, and it can use some polyalcohols, such as glycerol and sorbitol. Strain R210^T does not grow with these substrates, it lacks the genes encoding 1,3-propanediol dehydrogenase (EC 1.1.1.202) and glycerol dehydratase (EC 4.2.1.30) required for the conversion of polyalcohols¹⁰⁶.

Cells of strain R210^T were slightly curved rods of 0.5-0.9 µm in diameter and 0.6-1.4 µm in length, were motile and occurred singly or in pairs. Growth was observed between 4 °C and 40 °C with an optimum at 30 °C. The optimal pH for growth was 7.8 with range between 6.0 and 9.6. Major fatty acid methyl esters (FAMES) of strain R210^T were:

$C_{14:0}$ (16.1 %), $C_{16:1\omega9c}$ (35.5 %), $C_{16:0}$ (19.7 %) and $C_{18:1\omega9c}$ (19.7 %). These were also the major fatty acids in *T. pasteurii* and *T. palustris* (Table 2.3). Membranes of *T. flocculiformis* and *T. patagoniensis* lacked the unsaturated fatty acid $C_{16:1\omega9c}$ (the most abundant membrane fatty acid in strain R210^T and *T. pasteurii*). Unsaturated fatty acids are normally associated to increased membrane fluidity and their relative abundance can vary with environmental factors, e.g. temperature¹⁰⁷.

Table 2.3 | Cellular fatty acid compositions of members of the genus *Trichococcus*.

Fatty acid	1	2	3	4	5	6
$C_{12:0}$	0.7	0	1	12	0	0.7
$C_{14:0}$	16.1	14	28	57	21	11.2
$C_{15:0}$	0.6	0	0	0	0	0
$C_{16:1\omega9c}$	35.5	46	0	18	20	0
$C_{16:0}$	19.7	15	16	14	15	16.5
$C_{18:1\omega9c}$	19.7	18	6	0	22	21.8
$C_{18:1\omega7c}$	0.6	0	0	0	0	0
$C_{18:0}$	3.0	2	2	0	4	3.3
$C_{20:1\omega9c}$	0.8	0	0	0	0	0
$C_{20:1\omega7c}$	0.8	0	0	0	0	0

Strains: 1, R210^T (data from this study); 2, *T. pasteurii* DSM 2381⁶⁷; 3, *T. flocculiformis* DSM 2094⁶⁷; 4, *T. collinsii* DSM 14526⁶⁷; 5, *T. palustris* DSM 9172⁶⁷; 6, *T. patagoniensis* DSM 18806⁴⁹. Values are presented as percentages of total fatty acids.

The data of dDDH analysis conducted for all *Trichococcus* spp. indicated that strain R210^T considerably differs from other *Trichococcus* spp.. Therefore, based on morphological, physiological, biochemical characteristics along with measured dDDH values for all *Trichococcus* spp., it is suggested that strain R210^T represents a novel species within the genus *Trichococcus*, for which the name *Trichococcus ilyis* is proposed. The type strain of *Trichococcus ilyis* is R210^T (=DSM 22150^T =JCM 31247^T).

DESCRIPTION OF *TRICHOCOCCUS ILYIS* SP. NOV.

T. ilyis (i'ly.is. Gr. fem. n. *ilyis*, *ilyos* mud, sludge; latinised Gr. gen. n. *ilyis* of sludge referring to the source of isolation).

Cells are motile, slightly curved rods with sizes 0.63-1.40 x 0.48-0.90 µm. Cells grow in a temperature range of 4 °C to 40 °C and the optimum at 30 °C. pH for growth ranges is 6.0-9.6 and the optimum at pH 7.8. Strain R210^T can utilise a broad range of carbon sources for growth, such as pyruvate, L-malate, D-arabinose, citrate, D-fructose, D-galactose, D-glucuronate, D-glucose, D-mannitol, D-mannose, L-rhamnose, D-cellobiose, D-lactose, D-maltose, sucrose and raffinose. No growth was observed with formate, L-lactate D-galacturonate, D-gluconate, L-sorbose, D-tagatose, D-xylose, starch, glycerol, ethanol, methanol, inositol, D-dulcitol and D-sorbitol. Strain R210^T did not grow autotrophically on H₂/CO₂.

The type strain is *Trichococcus ilyis* (=DSM 22150^T = JCM 31247), isolated from sludge of a sulfate-reducing bioreactor with a citrate-containing waste stream. The genomic G+C content of the type strain is 47.9 %.

ACKNOWLEDGEMENTS

This research was supported by the European Research Council under the European Union's Seventh Framework Program (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science. The work conducted by the U.S. Department of Energy Joint Genome Institute (DOE-JGI), a DOE Office of Science User Facility, was supported by the Office of Science of the DOE under Contract No. DE-AC02-05CH11231. The authors gratefully acknowledge M. Huntemann, A. Clum, M. Pillay, K. Palaniappan, N. Varghese, N. Mikhailova, D. Stamatidis, T.B.K. Reddy, C.Y. Ngan, C. Daum, V. Markowitz, N. Ivanova, and T. Woyke from JGI for their technical assistance.

GENOME ACCESSION NUMBER IN EUROPEAN NUCLEOTIDE ARCHIVE

Strain R210^T <http://www.ebi.ac.uk/ena/data/view/PRJEB12709>

Trichococcus collinsii <http://www.ebi.ac.uk/ena/data/view/PRJEB12704>

Trichococcus flocculiformis <http://www.ebi.ac.uk/ena/data/view/PRJEB12705>

Trichococcus palustris <http://www.ebi.ac.uk/ena/data/view/PRJEB12707>

Trichococcus pasteurii <http://www.ebi.ac.uk/ena/data/view/PRJEB12708>

Trichococcus patagoniensis <http://www.ebi.ac.uk/ena/data/view/PRJEB13128>

The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of *Trichococcus ilyis* strain R210^T is FJ374769.

CHAPTER 3

GENOME-GUIDED IDENTIFICATION OF NOVEL PHYSIOLOGICAL TRAITS OF *TRICHOCOCCUS* SPECIES

Nikolaos Strepis^{1,2}, Henry D Naranjo¹, Jan Meier-Kolthoff³, Markus Göker³, Nicole Shapiro⁴, Nikos Kyrpides⁴, Hans-Peter Klenk^{3,5}, Peter J Schaap², Alfons JM Stams^{1,6}, Diana Z Sousa¹

1. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

2. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands

3. Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

4. DOE Joint Genome Institute, Walnut Creek, CA, USA

5. School of Biology, Newcastle University, Newcastle, UK

6. Centre of Biological Engineering, University of Minho, Braga, Portugal

ABSTRACT

Trichococcus species can degrade a wide range of carbohydrates and some species can convert glycerol to 1,3-propanediol (1,3-PDO). Genomes of *Trichococcus* species are available, providing the opportunity to further explore the physiological potential of this group. In this study, we conducted a functional analysis of all described *Trichococcus* species for genotype-phenotype mapping. Among the different species, we were able to identify unique genes involved in e.g. the utilization of specific substrates (glycerol, arabinan and alginate), antibiotic resistance, tolerance to low temperatures and osmoregulation. *Trichococcus* strain ES5 and *T. pasteurii* are known 1,3-PDO producers. A large operon containing 17 genes was only present in these two species and contained essential genes for 1,3-PDO production from glycerol. In the genomes of all species, six cold shock domains were identified and four of the *Trichococcus* species could grow at 0 °C. Osmoregulation protein domains were detected in all *Trichococcus* species, except in *T. palustris*, which had a lower resistance to salinity when compared to other *Trichococcus* strains. This study shows the value of using genomic information to guide phenotypic and metabolic characterisation of microorganisms and to investigate traits that are not usually assessed in a more conventional species characterisation.

INTRODUCTION

Trichococcus species were isolated from diverse and geographically spread ecosystems. Various species derive from waste treatment systems or contaminated sites: *T. flocculiformis* (activated sludge)⁷³, *T. pasteurii* (septic pit)⁶⁷, *T. collinsii* (hydrocarbon split soil)⁶⁷, *T. ilyis* (sulfate reducing bioreactor)¹⁰⁸, *Trichococcus* strain ES5 (anaerobic sludge)⁴⁸; while others were isolated from natural environments: *T. patagoniensis* (guano from penguin, Patagonia)⁴⁹, *T. palustris* (swamp)⁶⁷, and *T. paludicola* and *T. alkaliphilus* (high elevation wetland, Tibet)⁶⁸ (Table S3.1).

Trichococcus species share a very high 16S rRNA gene sequence similarity, in the range of 98-100 %⁶⁷. This often impairs the taxonomic classification of new strains within this genus on the basis of 16S rRNA gene sequence similarity, and whole genome comparison needs to be performed (e.g. DNA-DNA hybridisation). More recently, due to the availability of genomic data for members of this genus, digital DNA-DNA hybridisation (dDDH) was used to effectively classify *Trichococcus* species^{17,68}. The availability of genomic information also provides also the opportunity for comparing and analysing gene/function diversity among the different species. Functional analysis based on protein domains allows species clustering based on their potential metabolic functions, thereby connecting genotype and physiology^{20,109}.

All members of the *Trichococcus* genus can grow on glucose, cellobiose, D-mannose, fructose and sucrose^{48,67}. *Trichococcus* species belong to the lactic acid bacteria (LAB) and are phylogenetically related to the genera *Carnobacterium* and *Aerococcus*⁵⁰. However, individual members of the genus *Trichococcus* show unique phenotypes. *Trichococcus* sp. strain ES5 and *T. pasteurii* are capable of converting glycerol mainly to 1,3-PDO with product yields comparable to those of other known 1,3-PDO producers, such as *Clostridium butyricum* and *Klebsiella pneumoniae*^{110,111}. As 1,3-PDO is used for the synthesis of biodegradable polyesters, this specific phenotype is interesting for biotechnological applications^{51,112,113}. In general, *Trichococcus* species grow at a broad temperature range of 4 °C to 40 °C. *T. patagoniensis* can grow at -5 °C and tolerates a salinity of up to 5 % (w/v) NaCl⁴⁹,

which is also the case for several related *Carnobacterium* species, such as *C. funditum*, *C. alterfunditum* and *C. pleistocenium*^{114,115}.

The objective of this study was to use functional analysis based on protein domains to identify novel metabolic traits in *Trichococcus* species. This genome-guided in vitro study also contributed as well to the verification of specific phenotypes in different *Trichococcus* species. To complete the phenotypic assignment, the members of *Trichococcus* genus were compared with other closely related LAB.

MATERIALS AND METHODS

Source of genomes

The genome of *Trichococcus* strain ES5 was sequenced at the Joint Genome Institute from the US Department of Energy (JGI-DOE) (Walnut Creek, CA) using an Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA). This genome (11,259,926 reads and 151 bp read length) was assembled and annotated as described previously¹⁰⁸. All the publicly available genome sequences of *Trichococcus* species, *i.e.* *T. pasteurii*, *T. flocculiformis*, *T. palustris*, *T. collinsii*, *T. patagoniensis*, *T. alkaliphilus*, *T. paludicola* and *T. ilyis* and the genomic sequences of LAB and *Bacillus subtilis* were obtained from the NCBI database¹¹⁶.

Functional analysis and genome annotation

All the genomes from *Trichococcus* and LAB species, and *B. subtilis* (32 species in total) were re-annotated using the pipeline of Semantic Annotation Platform with Provenance (SAPP) that includes Prodigal (2.6) for predicting coding gene sequences^{89,117}. *T. paludicola* and *T. alkaliphilus* locus tags were based on Prodigal v2.6.2018 prediction (*T. paludicola*: Ga019, *T. alkaliphilus*: PXZT) for comparison purposes. Functional analysis was based on protein Hidden Markov Model domains (HMM) and the significance of gapped alignment scores can be model (profile HMM generated by InterProScan (version 5.17-56.0)^{90,118,119}. An InterPro protein domains matrix was generated for all

the *Trichococcus*, selected LAB, and *B. subtilis*. *B. subtilis* was used as an outlier for the study and was not included in the core and unique protein domain analysis. Core protein domains (present in all genera) and unique protein domains (present in only one of the analysed species) were identified. The presence/absence matrix of protein domains from all species was converted to distances by using the dice coefficient and hierarchical clustering from the UPGMA method. Statistical analysis (generation of a hierarchical clustering from protein domains) was carried out in R¹²⁰. For functional protein domain clustering, the analysis was performed in DARwin 6.0.0.14¹²¹. Besides this, 16S rRNA gene sequences were extracted from the genomes and aligned using the software CLC Main Workbench 8.0 (CLC Bio, Aarhus, Denmark). A 16S rRNA gene-based clustering tree was constructed using the UPGMA method and the general time reversible model.

Whole-genome based analyses

All pairs of strains were compared using the Genome-to-Genome Distance Calculator 2.1 (GGDC; <https://ggdc.dsmz.de>) under recommended settings¹⁷ and pairwise digital DNA-DNA hybridisation values (dDDH) were inferred accordingly. Afterwards, the distance matrix was subjected to a clustering using established thresholds for delineating species¹⁷ as well as subspecies¹²². Clustering was done using the OPTSIL clustering program¹²³ as previously explained⁷⁶.

A genome sequence-based phylogenetic analysis based on the coding regions was conducted using the latest version of the Genome-BLAST Distance Phylogeny (GBDP) method as previously described¹²⁴. Briefly, BLAST+¹²⁵ was used as a local alignment tool and distance calculations were done under recommended settings (greedy-with-trimming algorithm, formula d_g , e-value filter 10^{-8}). 100 pseudo-bootstrap replicates were assessed under the same settings for each. Next, a balanced minimum evolution tree was inferred using FastME v2.1.4 with SPR postprocessing¹²⁶. Replicate trees were reconstructed in the same way and branch support was subsequently mapped onto the tree¹²⁴. Finally, exchanged genomic syntenies were defined with Sibelia¹²⁷ using default parameters, and visualised in a circular graph by Circos¹²⁸.

Microbial growth tests

Growth experiments were conducted with anaerobic basal medium prepared as previously described⁸⁰. 45 mL of medium were dispensed in 120 mL serum bottles, which were sealed with rubber stoppers and aluminium caps. The bottles' headspace was flushed with N₂/CO₂ (80/20 v/v) to a final pressure of 1.5 bar. After autoclaving, and before inoculation, the medium was supplemented with 0.5 mL of salts solution and 2.5 mL of bicarbonate solution⁷¹. Yeast extract was added to the medium at a concentration of 0.1 g/L. Substrates were added to the medium from sterile stock solutions. Glycerol was tested in concentration of 20 mM. Alginate was tested at a concentration of 5 mM and arabinan (sugar beet, Ara:Gal:Rha:GalUA=88:3:2:7) at 0.4 % (v/v). Incubations were conducted in the dark, without stirring and at 30 °C (unless stated otherwise).

Antibiotic resistance tests

Antibiotic resistance tests for tetracycline were performed in plates with rich *Clostridium* medium (Fisher Scientific, PA) and 1 % agar. Minimum inhibitory concentration (MIC) tetracycline test stripes were used with a test range of 0.016-256 mg/L (Liofilchem, Roseto degli Abruzzi, Italy). The incubation was conducted inside anaerobic containers at 30 °C.

Psychrotolerance and salinity test

Temperature and salinity tests were performed using 20 mM of glucose as substrate. The growth of all members of *Trichococcus* genus was tested at 0 °C and monitored for 45 days. For salinity tolerance experiments, sodium chloride was used at concentrations of 2, 4, 6, 8, 10 % (w/v). The growth of *Trichococcus* species at different salinities was monitored for ten days.

Analytical measurements

Growth was quantified by optical density (OD 600 nm), measured in a

spectrometer (Hitachi U-1500, Labstuff, The Netherlands). Soluble metabolites, such as glucose, glycerol, 1,3-PDO, lactate, ethanol, acetate and formate were measured using the Thermo Electron HPLC system equipped with an Agilent Metacarb 67H column (Thermo, Waltham, MA), which had as mobile phase sulphuric acid (5 mM) at a flow rate of 0.8 mL min⁻¹ and temperature at 45 °C.

RESULTS AND DISCUSSION

Genome and protein domains comparison

The SAPP annotation of *Trichococcus* species resulted in a matrix that included 3,540 proteins domains, including occurrence frequency. The assembled genomes of *Trichococcus* species all have similar sizes (3.0-3.3 Mbp) and the G+C content ranges from 44-48 % (Table S3.1). There are 2,431 core protein domains and 3,540 pan protein domains, meaning that multiple protein domains are conserved in the genomes of these species.

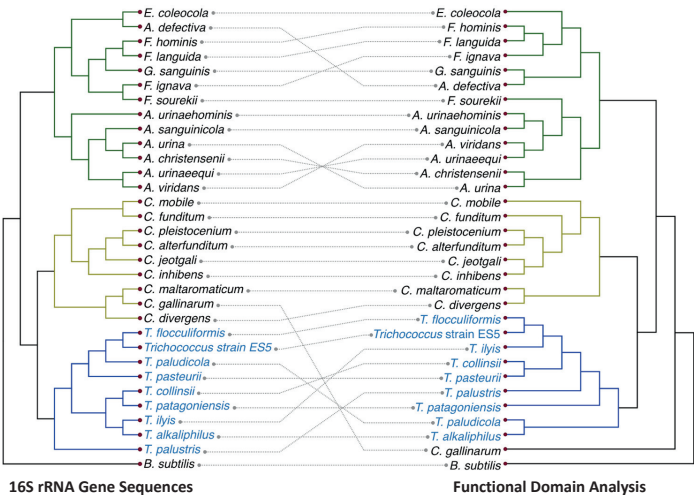


Figure 3.1 | Comparison of the hierarchical clustering based on the 16S rRNA gene sequence with a hierarchical clustering based on presence or absence of functional domains for *Trichococcus* and *Lactobacillales* species. Independent clusters are represented by distinct colours.

Functional clustering of protein domains shows that *Trichococcus* species belong to an independent unit among other *Lactobacillales* species (Figure 3.1). The genomes of *Trichococcus* sp. strain ES5, *T. patagoniensis*, *T. collinsii*, *T. flocculiformis* and *T. ilyis* have large conserved genomic synteny with multiple protein domains (Figure 3.2). *T. palustris* is separated from the cluster as was noted previously¹⁰⁸.

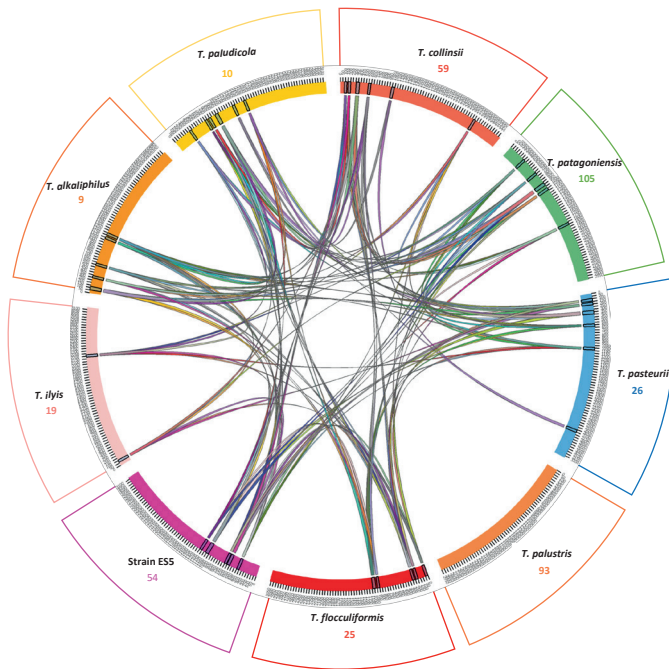


Figure 3.2 | Conserved synteny between the genomes of *Trichococcus* species. Each colour represents a *Trichococcus* genome. Synthetic regions larger than 40 kb are indicated. Numbers represent unique species domains and coloured lines indicate genome rearrangements. Note that *T. palustris* has no shared syntenic regions larger than 40 kb with other *Trichococcus* species.

This matrix was studied for unique metabolic traits in *Trichococcus* species. A summary of the metabolic traits found in *Trichococcus* species, which were further explored in this work, is shown in table 3.1. Protein domains of enzymes involved in the conversion of glycerol, alginate and arabinan were identified, and subsequent in vitro

trials were conducted to confirm the utilisation of these compounds. Additionally, cold adaptation and osmoregulation proteins domains are present in the genomes indicating that several species may be well adapted to survive at low temperature and high salinity. Finally, protein domains for CRISPR and the production/tolerance of antimicrobial compounds were predicted for four of the species. For conserved protein domains, the analysis was expanded to LAB species to improve the functional clustering of *Trichococcus* species and obtain further insights into protein domain specificities. Functional analysis indicated close clustering among *Trichococcus* members. Moreover, the GGDC-based subspecies clustering (Figure 3.3) revealed that *Trichococcus* strain ES5 is a subspecies of *T. flocculiformis* with a digital DDH value of 71 % indicating distinct subspecies.

Table 3.1 | Unique and interesting genes that were discovered by functional analysis of *Trichococcus* species. Strains: 1. *Trichococcus* sp. strain ES5; 2. *T. pasteurii*; 3. *T. collinsii*; 4. *T. palustris*; 5. *T. patagoniensis*; 6. *T. ilyis*; 7. *T. flocculiformis*; 8. *T. alkaliphilus*; 9. *T. paludicola*.

Organism	Functional annotation	Locus tag
Substrates		
1,3-PDO production		
1 2	Glycerol kinase	TES5_2082 Tpas2911
1 2	Dihydroxyacetone activator	TES5_2083 Tpas_2912
1 2	Glycerol dehydrogenase	TES5_2084 Tpas_2913
1 2	Phosphoenolpyruvate phosphotransferase	TES5_2085-2087 Tpas_2914-2916
1 2	1,3-propanediol dehydrogenase	TES5_2088 Tpas_2917
1 2	Dihydroxyacetone	TES5_2089-2090 Tpas_2918-2919
1 2	Glycerol dehydratase	TES5_2091-2093 Tpas_2920-2922
1 2	Glycerol dehydratase activator	TES5_2094-2095 Tpas_2923-2924
1 2	Cobalamin adenosyltransferase	TES5_2096 Tpas_2925
1 2	Hypothetical protein	TES5_2097,2099 Tpas_2926, 2928
1 2	Glycerol uptake facilitator	TES5_2098 Tpas_2927
Alginate utilisation		
3	Alginate lyase	Tcol_1369,1377,1704
Arabinan utilisation		
5	Glycosyl hydrolase	Tpat_54,101,590,610,948-949,136,1167,1171,1259,2028,2033,2527-2528,2577,2585,2682
5	Metal-dependent hydrolase	Tpat_57,88,320,321,954,1043,1060,1227,1247,1391, 1392,2241

5	Extracellular endo-alpha-(1->5)-L-arabinanase	Tpat_1197,1296
Psychrotolerance		
All	Cold-shock protein	TR210_741 Tpas_88 Tpal_285 TES5_627 Tflo_313 Tcol_65 T pat_494 PXZT01000007.1_99 Ga0192377_1015_35
All	Cold shock protein signature	Tpat_494, 1801, 1802, 1901, 1923 Tcol_65, 532, 554, 1698, 1699 Tflo_313, 455, 458, 688, 980 TES5_627, 827, 849, 1357,135 9 Tpal_285,869,1036,1801,1820,1 877 Tpas_88,599,1471,1472,2297, 2758 TR210_741,1024,1470, 1709, 1819, 1842 PXZT01000016.1_53, 1.1_301, 5.1_152, 4.1_46, 5.1_150 Ga0192377_1004_168, 1008_10-11, 1002_82, 1004_145
Salinity tolerance		
6	Glycine betaine transporter OpuD	TR210_1348
2 3 5 6 8 9	Betaine binding ABC transporter protein	Tpas_2814-2815 Tcol_1997 Tpa t_1468 TR210_2767-2768,2770 Ga0192364_3215_54-57 PXZT01000008_23-26
1 3 6 7 8 9	Osmotically activated choline ABC transporter	TES5_1206-1209 Tcol_773- 776 TR210_342-345 Tflo_1131- 1134 Ga0192364_2415_1215 PXZT01000003.2_54-57
1 2 5 6	Choline binding protein A	TES5_1355 Tpas_1469 Tpat_1570 TR 210_2363,1711,2104
1 2 3 5 7 8 9	Glycine betaine ABC transport system	TES5_1660-1662 Tcol_1041- 1043 Tpas_2619- 21 Tpat_203-05 Tflo_1599-01 Ga0192356_1653_23-25 PXZT01000006.2_75-77
Bacterial defence		
5	SNARE associated Golgi protein	Tpat_1693,1825
4	Tetracycline resistance	Tpal_1098,1664,1687
2	Toxin antidote HigA	Tpas_511
2	Plasmid system killer	Tpas_512
6	Bacteriocin class Iib	Tflo_874,878-879
5	Cas9	Tpat_1430
1 5 7	Cas1 6	TES5_196 Tpat_1431 Tflo_184
5	Cas2	TES5_195 Tpat_1432
1 2 6 7	Cas3	TES5_201 Tpas_1155 TR210_680 Tflo_179
1 2 6 7	Cas5d	TES5_200 Tpas_1156 TR210_679 Tflo_180
1 2 6 7	Casd1	TES5_199 Tpas_1157 TR210_675 Tflo_181
1 2 6 7	Csd2	TES5_198 Tpas_1158 TR210_677 Tflo_182
1 2 6 7	Csd4	Tpas_1159 TR210_676 Tflo_183

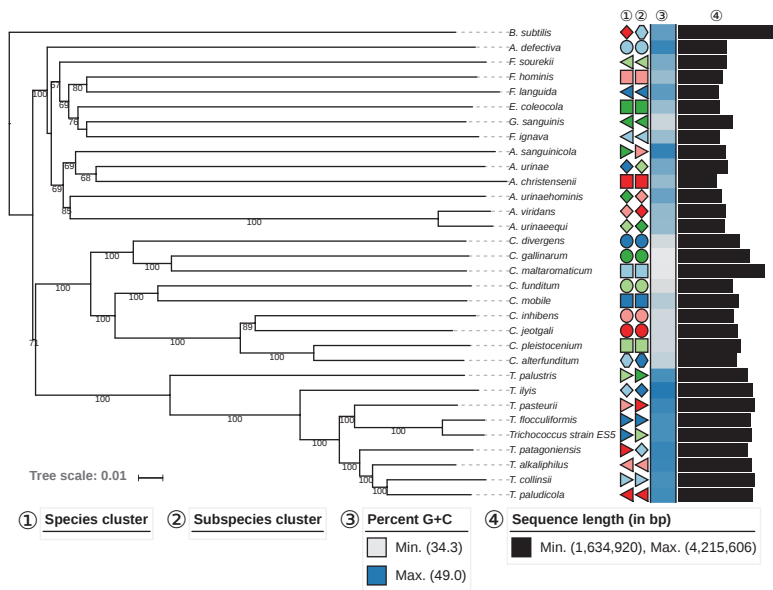


Figure 3.3 | Genome-based phylogenomic analysis restricted to coding regions. Tree inferred with FastME 2.1.4 from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications, with an average branch support of 88 %. Leaf labels are further annotated by their affiliation to species (1) and subspecies (2) clusters as well as their genomic G+C content (3) and their overall genome sequence length (4).

Metabolic properties of *Trichococcus* species

In general, *Trichococcus* species can utilise substrates such as cellobiose, sucrose, maltose, and glucose, but their metabolic ability was not previously studied at genomic level. *In silico* analysis indicated that *Trichococcus* species possess multiple genes regarding carbohydrate utilisation. Genes codifying for the Embden-Meyerhof-Parnas (EMP) pathway and pentose phosphate pathway (PPP) were found in all the genomes of *Trichococcus* species. All *Trichococcus* species have genes for the fermentation of pyruvate to ethanol, acetate and lactate. We determined growth and product formation on glucose for all available *Trichococcus* species (Table 3.2). Glucose fermentation products are lactate, formate, acetate and ethanol, but the product

ratio is species-dependent. Lactate is the main product for all six species, except for *T. patagoniensis*, in which formate is the main product. This is supported by the presence of a pyruvate formate-lyase in the genome of *T. patagoniensis* (Tpat_2317). Ethanol yield in cultures of *T. patagoniensis* and *T. collinsii* was 15 % and 11 % ($\text{mol}_{\text{ethanol}}/\text{mol}_{\text{glucose}}^{\text{consumed}}$), respectively, which is almost double that of the observed yield for other *Trichococcus* species.

Table 3.2 | Glucose and glycerol fermentation products of *Trichococcus* species. Glucose consumption generates acetate, ethanol, lactate and formate, while glycerol fermentation has 1,3-PDO, acetate, formate and lactate. Glucose and glycerol were added in concentration of 20 mM.

Strain	Glucose (mM)	Lactate (mM)	Formate (mM)	Acetate (mM)	Ethanol (mM)	Carbon recovery (%)
<i>T. patagoniensis</i> DSM 18806T	0.67 ± 0.15	12.83 ± 0.46	19.00 ± 0.52	6.97 ± 0.21	8.97 ± 0.31	80.8
<i>T. palustris</i> DSM 9172T	0.00 ± 0	17.17 ± 0.81	13.70 ± 0.17	5.10 ± 0.20	6.63 ± 0.21	81.8
<i>T. collinsii</i> DSM 14526T	6.47 ± 0.49	22.17 ± 0.21	3.27 ± 0.25	0.63 ± 0.15	1.07 ± 0.12	85.4
<i>T. pasteurii</i> DSM 2381T	1.23 ± 0.65	25.33 ± 0.87	5.50 ± 0.17	1.50 ± 0.00	1.93 ± 0.72	81.2
<i>T. flocculiformis</i> DSM 2094T	0.70 ± 0.53	23.37 ± 2.06	7.53 ± 0.15	2.73 ± 0.15	3.87 ± 0.23	80.0
<i>T. ilyis</i> DSM 22150T	0.00 ± 0.00	21.63 ± 0.86	9.93 ± 0.42	3.43 ± 0.23	4.63 ± 0.25	80.1
<i>Trichococcus</i> strain ES5 DSM 23957	0.00 ± 0.00	23.60 ± 0.66	7.70 ± 0.36	2.17 ± 0.12	4.10 ± 0.20	80.5
	Glycerol (mM)	Lactate (mM)	Formate (mM)	Acetate (mM)	1,3-PDO (mM)	Carbon recovery (%)
<i>T. pasteurii</i> DSM 2381T	0.00 ± 0.00	0.50 ± 0.10	0.90 ± 0.46	3.60 ± 0.65	13.80 ± 0.21	85
<i>Trichococcus</i> strain ES5 DSM 23957	0.00 ± 0.00	0.50 ± 0.00	2.27 ± 0.21	4.45 ± 0.20	12.25 ± 0.10	82.3

T. pasteurii and *Trichococcus* strain ES5 are the only *Trichococcus* species that can degrade glycerol to 1,3-PDO. The genome of these species encodes for a complete genomic synteny (17 genes) potentially

involved in glycerol conversion (Table 3.1), which is absent in other *Trichococcus* species. Two of the genes in this synteny are essential for glycerol conversion to 1,3-PDO: glycerol dehydratase (alpha, beta and gamma subunits) and 1,3-propanediol dehydrogenase. Additional genes encoding for glycerol kinase and glycerol dehydrogenase are also located in the operon. Genes with a regulatory function are glycerol dehydratase activator, which activates glycerol dehydratase, and cobalamin adenosyltransferase, which is involved in the conversion of cobalamin (vitamin B12) to its coenzyme form, adenosylcobalamin. Glycerol dehydratase requires vitamin B12 as a binding co-factor¹²⁹. In the operon, a gene for a glycerol uptake facilitator was identified. Glycerol uptake facilitator is responsible for the uptake of glycerol in the cell¹³⁰.

T. collinsii has unique domains related to alginate utilisation and encodes three alginate lyases (Table 3.1). Alginate is an anionic polysaccharide, which participates in the formation of exopolymer saccharides (EPS) and therefore in the formation of biofilms¹³¹. *In vitro* testing confirmed that *T. collinsii* utilises alginate ($OD_{722}=0.19 \pm 0.00$). This result suggests that the alginate lyases enzymes in the genome might contribute to the degradation of EPS.

A high number of short chain dehydrogenase/reductase (SDR) genes were identified in the genomes of *T. paludicola* (25 genes) and *T. alkalicophilus* (21 genes). These genes were not found in other *Trichococcus* species. The SDR family contains multiple enzymes involved in the metabolism of amino acid, lipid, carbohydrate and cofactors¹³². The SDR protein family is well distributed in all domains of life including bacteria and archaea, which signifies their important role especially for the function of lipid metabolism (lipid droplets and transportation)¹³³.

In the genome of *T. patagoniensis*, 17 homologous domains of glycosyl hydrolases were identified, but they all belong to genes with hypothetical proteins (Table 3.1). In addition, 12 homologous genes of metal-dependent hydrolase were identified. The glycoside family-1 of hydrolases includes e.g. glucosidases, galactosidases and hydrolases¹³⁴. In addition, two genes of extracellular endo-alpha-(1->5)-L-arabinanase exist in the genome. Extracellular endo-alpha-(1->5)-L-arabinanase is involved in the degradation of arabinan and it

is a main enzyme in the degradation of the plant cell wall. *B. subtilis* uses the enzyme to dissolve plant biomass in the soil¹³⁵. In addition, *B. subtilis* synthesises several glycosyl hydrolases for depolymerisation of plant cell wall polysaccharides¹³⁵. To confirm the protein domains prediction, the growth of *T. patagoniensis* on arabinan was tested in vitro. *T. patagoniensis* could utilise and grow on arabinan ($\text{OD}_{76} = 0.25 \pm 0.02$). *T. patagoniensis* was isolated from penguin guano and could have a function in the degradation of sugar materials in this ecosystem. Arabinan-degrading enzymes are important for the food industry (wine, sweeteners, etc.) and for biofuels production (degradation of lignocellulosic biomass)¹³⁶.

Psychrophilic characteristics

Six cold shock domains (CSD) (IPR011129) were identified in all *Trichococcus* genomes (Table 3.1). One additional CSD was identified in the genomes of *T. palustris* and *T. ilyis*. These conserved CSDs in *Trichococcus* species belong to genes annotated as cold shock protein signatures. In such genes three additional domains were observed, the cold-shock DNA-binding (IPR002059), nucleic acid-binding OB-fold (IPR012340) and cold-shock conserved site (IPR019844). Each member within the *Trichococcus* genus has a complete cold-shock protein with seven domains including CDS (IPR011129) in the genome (Table 3.1). This cold shock protein is composed of additional domains of: ATPase F1 nucleotide-binding (IPR000194), AAA+ ATPase (IPR003593), transcription termination factor Rho (IPR004665), rho termination factor N-terminal (IPR011112), rho termination factor RNA-binding domain (IPR011113), nucleic acid-binding OB-fold domain (IPR012340) and P-loop containing nucleoside triphosphate hydrolase domain (IPR027417). Only four of the *Trichococcus* species can grow at 0 °C, i.e. *T. pasteurii*, *Trichococcus* strain ES5, *T. collinsii* and *T. patagoniensis*. The lag phase for *T. patagoniensis* and *T. palustris* was eight days, whereas the *Trichococcus* strain ES5 and *T. collinsii* grew after 23 days (Figure S3.1). *T. paludicola* and *T. alkaliphilus* were previously characterised as psychrotolerant and were tested in temperatures as low as 4 °C⁶⁸.

As cold shock domains are conserved among members of *Trichococcus* genus, other genomes of LAB species were analysed. Functional analysis supports that the cold shock protein is identified in the genomes of all studied LAB species. However, only seven of the 22 species contain six to eight additional CDS, *Carnobacterium mobile*¹¹⁴, *C. pleistocenium*¹¹⁵, *C. jeotgali*¹³⁷, *C. inhibens*¹³⁸, *C. funditum* strain pf3¹¹⁴, *C. maltaromaticum*¹³⁹, *C. alterfunditum*¹¹⁴ (Table S3.1).

High salinity resistance

Functional analysis indicated protein domains related to osmoregulation in *Trichococcus* species (Table 3.1). Multiple domains encoded in genes related to osmoregulation mechanisms, such as glycine and betaine transportation, were identified in the genomes of all *Trichococcus* species. During osmotic pressure bacterial cells can survive by increasing in the cytoplasm the concentration of osmoprotectants like betaine and choline¹⁴⁰. Osmoprotectants without charge can raise osmotic pressure in the cell. Either cellular synthesis or the uptake of osmoprotectants can increase resistance to osmotic stress^{141,142}. Each of the *Trichococcus* genomes contains multiple osmoregulation genes (Table 3.1), except for *T. palustris*.

In vitro analysis of salinity tolerance confirmed that *Trichococcus* species can survive high osmotic stress. Only *T. palustris* was sensitive to salinity as growth was inhibited at 2 % NaCl (Figure S3.2). All the other bacteria can grow in media with a NaCl concentration of 2 %. At 4 % salinity and after 6 days, growth was observed for only four of the tested bacteria, *T. pasteurii*, *Trichococcus* strain ES5, *T. flocculiformis*, *T. patagoniensis*. After ten days, weak growth was observed at 6 % NaCl for *T. patagoniensis*, *T. pasteurii* and *Trichococcus* strain ES5 (Figure S3.2). During seven days incubation, *T. paludicola* and *T. alkaliphilus* were previously observed to tolerate NaCl concentrations up to 4.5 %⁶⁸. Possibly, the encoded osmoregulation mechanisms provide salinity tolerance to *Trichococcus* species and enhance their ability to survive in high salinity concentrations of up to 6 % (w/v) NaCl.

Defence systems of *Trichococcus*

Recent studies support the effective defence of the CRISPR system in bacteria against viral threats¹⁴³. The CRISPR system contains Cas genes which target and destroy foreign DNA in the cells. We identified Cas genes in *T. patagoniensis*, *T. flocculiformis*, *T. pasteurii*, *T. ilyis*, and *Trichococcus* strain ES5 (Table 3.1). The CRISPR system in *T. patagoniensis* can be classified as Cas2, type II-C, while other *Trichococcus* species encode a class 1 type I-C CRISPR system¹⁷. Several spacer sequences (*i.e.* foreign nucleic acid sequences merged in the genome by CRISPR systems) were found in the analysed *Trichococcus* species: *T. fluccoliformis* (27), *T. patagoniensis* (88), *T. pasteurii* (115), *T. ilyis* (80), *Trichococcus* strain ES5 (82). The alignment of the spacers sequences from all *Trichococcus* species resulted in low similarity and further analysis showed that they do not contain common foreign DNA. Interestingly, some common spacers that occur in several halophilic and methanogenic archaea were identified, which deserves further analysis.

In *Trichococcus* species, alternative potential defensive mechanisms were found (Table 3.1). The domain of SNARE associated Golgi protein was identified in three hypothetical genes in *T. patagoniensis* genome. This type of SNARE proteins in bacteria can be used for promoting or blocking membrane fusion especially against eukaryotic cells¹⁴⁴.

Antibiotic resistance may be enhanced in *T. palustris* due to unique domains. The *T. palustris* genome possesses genes encoding for tetracycline resistance proteins (Table 3.1). Tetracycline is a common antibiotic for Gram positive and Gram negative bacteria¹⁴⁵. Only in the *T. palustris* genome was the genetic repository for potential resistance to tetracycline observed. *In vitro*, it was confirmed that *T. palustris* can grow in tetracycline concentrations until 4 µg/ml. *T. ilyis* and *T. palustris* were tested and were not observed to tolerate equivalent tetracycline concentrations.

The genome of *T. pasteurii* contains the gene of the toxin antidote protein HigA and the gene of a plasmid system killer (Table 3.1). The two genes are associated with bacterial toxin-antitoxin (TA) proteins and regulate the tolerance of the cells under environmental and chemical stress¹⁴⁵. The antitoxins main target is the binding of complementary

toxins by blocking substrate recognition. The *T. flocculiformis* genome contains three homologous genes for the domain bacteriocin class IIb, which inhibits the growth of other microorganisms¹⁴⁶.

CONCLUSION

The genome-guided characterisation of *Trichococcus* species resulted in the discovery of novel functional traits within this genus. This approach revealed a large operon that is responsible for the production of 1,3-PDO in *Trichococcus* species and the genes required for degrading complex molecules like alginate and arabinan. These metabolic traits of *Trichococcus* species may enhance the biotechnological value of the species. Their robust phenotype at low temperature and high salinity, which was confirmed by protein domain analysis and *in vitro* trials, may further benefit their application. The CRISPR system and the unique defence mechanisms provide a leverage against viral attacks for *Trichococcus* species. In conclusion, functional clustering of the genus *Trichococcus* identified interesting genotypes that unveil phenotypic characteristics specific to the species.

ACKNOWLEDGEMENTS

This research was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science. The work conducted by the U.S. Department of Energy Joint Genome Institute (DOE-JGI), a DOE Office of Science User Facility, was supported by the Office of Science of the DOE under Contract No. DE-AC02-05CH11231.

SUPPLEMENTARY DATA

Table S3.1 | All species used for *in silico* analysis. The *Trichococcus* and *Lactobacillales* species have a different environment isolation.

Bacterium	DSMZ ID	Scaffolds	Genome size	% GC	Number of Domains	Environment isolation	Pathogenic
<i>Trichococcus pasteurii</i>	2381	21	3,288,204	45	3,108	Septic pit	No
<i>Trichococcus palustris</i>	9172	17	3,004,925	46	2,845	Swamp	No
<i>Trichococcus ilyis</i>	22150	44	3,194,513	48	2,977	Sulphate reducing bioreactor	No
<i>Trichococcus patagoniensis</i>	18806	38	3,017,565	47	2,712	Guano from penguin Patagonia	No
<i>Trichococcus collinsii</i>	14526	6	3,343,507	44	3,155	Hydrocarbon spilt site	No
<i>Trichococcus flocculiformis</i>	2094	91	3,123,895	44	3,054	Activated sludge	No
<i>Trichococcus strain ES5</i>	23957	50	3,194,513	44	3,045	Methanogenic sludge granules	No
<i>Trichococcus paludicola</i>	104691	5	1,935,222	46	1,561	High elevation wetland, Tibetan Plateau	No
<i>Trichococcus alkaliphilus</i>	104692	25	2,964,741	46	1,522	High elevation wetland, Tibetan Plateau	No
<i>Aerococcus sanguinicola</i>	14282	1	2,033,849	47	3,181	Human blood	Yes
<i>Abiotrophia defectiva</i>	9849	4	2,068,984	46	4,621	Human intestine	Yes
<i>Aerococcus christensenii</i>	15819	46	1,655,357	39	3,826	Human genitals	Yes
<i>Aerococcus urinae</i>	7446	1	1,974,262	42	2,312	Urinary tract, infection	Yes
<i>Aerococcus urinaeequi</i>	20341	9	2,017,038	39	4,921	Human peritoneal-related ascites	Yes
<i>Aerococcus urinaehominis</i>	15634	1	1,854,293	42	4,318	Human urinary tract	Yes
<i>Aerococcus viridans</i>	20340	31	2,018,981	39	4,883	Air and dust	Yes
<i>Carnobacterium alterfunditum</i>	5972	1	2,506,249	36	5,986	Ace Lake Antarctica	No
<i>Carnobacterium divergens</i>	20623	1	2,651,963	35	6,180	Gut Atlantic salmon	No
<i>Carnobacterium funditum</i>	5971	1	2,362,574	35	5,529	Ace Lake, Antarctica	No
<i>Carnobacterium gallinarum</i>	4847	1	3,081,008	34	6,089	Hindgut chambers, Atlantic cod	No
<i>Carnobacterium inhibens</i>	13024	1	2,381,355	35	5,747	Gut Atlantic salmon	No

<i>Carnobacterium jeotgali</i>	23388	12	2,549,728	35	6,029	Korean traditional fermented food	No
<i>Carnobacterium maltaromaticum</i>	20344	1	3,696,047	34	7,976	Fresh pork	Yes
<i>Carnobacterium mobile</i>	4848	6	2,632,867	37	2,697	Irradiated chicken meat	No
<i>Carnobacterium pleistocenium</i>	17715	1	2,695,901	35	6,334	Permafrost of the Fox Tunnel Alaska	No
<i>Eremococcus coleocola</i>	15696	21	1,795,927	38	4,350	Reproductive tract of horse.	Yes
<i>Facklamia hominis</i>	-	1	1,921,344	38	4,300	Human abscess on buttocks	Yes
<i>Facklamia ignava</i>	-	6	1,788,753	38	4,166	Human clinical specimens	Yes
<i>Facklamia languida</i>	-	1	1,739,240	43	3,926	Human clinical specimens	Yes
<i>Facklamia sourekii</i>	104271	48	2,092,896	41	4,917	Human sources	Yes
<i>Globicatella sanguinis</i>	7447	256	2,356,976	35	4,597	Cerebrospinal fluid	Yes
<i>Bacillus subtilis</i>	347	1	4,215,606	44	4,331	Soil	No

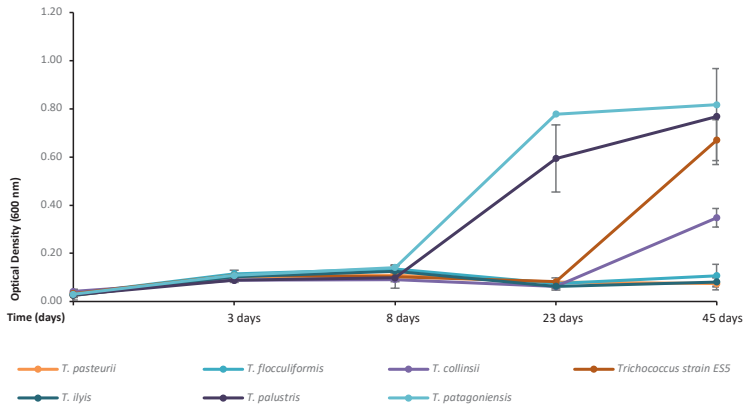


Figure S3.1 | Growth in 0 °C for *Trichococcus* species. For *T. patagoniensis* and *T. collinsii* growth was observed in 8 days, while *Trichococcus* strain ES5 and *T. palustris* growth was observed at 45 days. The rest of the species could not grow in 0 °C.

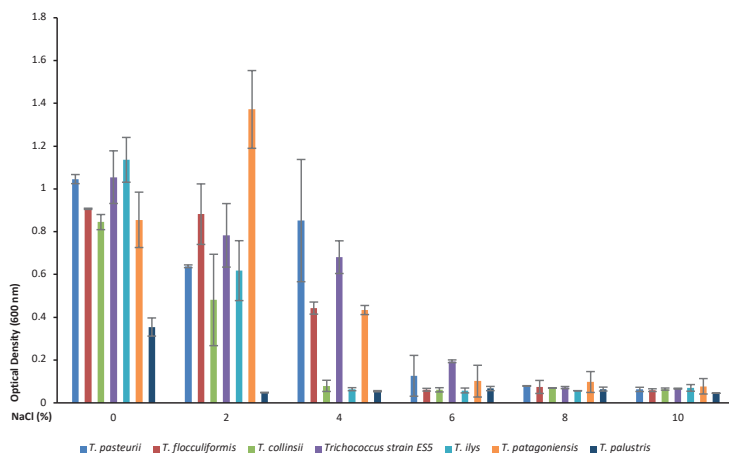


Figure S3.2 | Growth in high salinity for *Trichococcus* species. *Trichococcus* species could resist until 2 % NaCl. *T. flocculiformis* could tolerate until 4 % NaCl (w/v). For the high salinity of 6 % NaCl, three species were observed with growth, *T. patagoniensis*, *T. pasteurii* and *Trichococcus* strain ES5. The lag phase for 6 % NaCl was counted to 10 days. *T. palustris* was not able to tolerate salinity.

CHAPTER 4

SYSTEMATIC FUNCTION-BASED GENOME PROSPECTING FOR INDUSTRIAL TRAITS APPLIED TO 1,3-PROPANEDIOL PRODUCTION

Nikolaos Strepis^{1,2,#}, Jasper J Koehorst^{1,#}, Alfons JM Stams^{2,3}, Diana Z Sousa², Peter J Schaap¹

1. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands

2. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

3. Centre of Biological Engineering, University of Minho, Braga, Portugal

#. Equal contribution

ABSTRACT

Due to the success of next-generation sequencing, there has been a vast build-up of sequenced microbial genomes in the public repositories. To facilitate bioprospecting from this huge genomic potential, new efficient and flexible methods need to be developed. In this study, Semantic Web techniques are applied to develop a function-based genome mining approach following a knowledge and discovery in database protocol. Focusing on the industrially important trait of 1,3-propanediol (1,3-PDO) production, 178 new candidate species were identified. Furthermore, the genetic architecture of the trait was resolved, and essential domains were identified. Three newly identified non-pathogenic strains were successfully tested for 1,3-PDO production.

INTRODUCTION

Next generation sequencing technologies (NGS) have turned the publicly available genome repositories into data-rich scientific resources that currently contain structural and functional genome information of more than one hundred thousand bacterial genomes¹¹⁶. For biotechnological benefits, these repositories are excellent resources to mine for new and alternative cell factories, industriophilic traits and enzymes.

Biotechnology has embraced omics technologies and their impact on biotech innovation¹⁴⁷. However, setting up large scale functional screenings in genomics resources for industriophilic traits is challenging. NGS data generation has caused biocuration to be outpaced rapidly and currently more than 99% of the functional predictions in UniProt are based on automated computational predictions¹⁴⁸. The quality of computationally inferred functional genome annotations may vary due to the lack of data, the element-wise provenance, the use of different annotation pipelines and the non-updated reference databases. The acceptance thresholds and naming of protein functions are inconsistent and non-standardised defined. All these factors culminate in a low degree of interoperability. In the databases, the structural annotations, the predicted genes and protein sequences, underlying the functional annotations are presented in a highly-standardised format. For this reason, there is a much higher degree of interoperability and 'bottom-up' approaches can normally be used. These approaches start with a gene or protein sequence to find a corresponding function in genome(s) of interest. However, over larger phylogenetic distances, gene sequence-based clustering algorithms are hampered by lateral gene transfer, gene fusion/fission events and domain expansions¹⁴⁹. Furthermore, at larger scales, they suffer from high computational cost as time and memory requirements scale quadratically with the number of genome sequences to be compared.

A systematically robust function-based genome screening procedure requires a high degree of semantic interoperability. This means that functional information can be directly compared based on a pre-established syntactic interoperable genome annotation and computational predictions are linked to their provenance. To accomplish this,

we recently developed SAPP, a semantic annotation infrastructure supporting FAIR computational genomics¹¹⁷. SAPP uses the GBOL ontology as syntax¹⁵⁰ and automatically predicts, tracks and stores structural and functional genome predictions, associated datasets and element-wise provenance in a Linked Data format. For a systematic presentation of protein functions, protein domain architectures are used as proxy¹⁵¹. Demonstrating a high level of scalability, this set up was successfully used in an integrated analysis of the functional landscape of 432 *Pseudomonas* strains¹⁰⁹.

In this study, interoperable genome annotations are used in a systematic function-based *in silico* screening for bacterial species that can convert the bio-refinery by-product glycerol into the industrial high-value monomer 1,3-propanediol (1,3-PDO)^{52,152}. 1,3-PDO is an important precursor of biomaterials. It is currently mainly used as a monomer for novel polyester and biodegradable plastics, such as polytrimethylene terephthalate⁵². 1,3-PDO is a typical product of glycerol fermentation and currently very few species, mostly *Enterobacteria*, are known to form it. The underlying metabolic trait, a two-step reductive conversion of glycerol to 1,3-PDO, regenerates NAD⁺ that is required for the oxidative conversion of glycerol to dihydroxyacetone (Figure 4.1). In the first step, dehydration of glycerol to 3-hydroxypropionaldehyde (3-HPA) is mediated by a vitamin B12-dependent glycerol dehydratase. However, an oxygen sensitive B12-independent alternative enzyme was reported. The second step reduces 3-HPA to 1,3-PDO using the (NADH)+H⁺ dependent 1,3-propanediol-oxydoreductase (PDOR) regenerating NAD⁺.

For genome prospecting, a collection of 84,329 publicly available bacterial genomes were loaded in the SAPP semantic framework, structurally and functionally annotated, and mined for candidates to produce 1,3-PDO using a Knowledge Discovery in Databases approach (KDD)¹⁵³. Overall, the systematic analysis increased our knowledge on the genetic architecture of this metabolic trait, in terms of the overall domain composition, distribution and essentiality. The approach suggested that compared to 29 producers in literature (Table 4.1), at least 187 genome sequenced species potentially have the trait for 1,3-PDO production. Newly identified non-pathogenic species

such as *Clostridium magnum* were in vitro validated for 1,3-PDO production and *Trichococcus pasteurii* and *Trichococcus* strain ES5 were confirmed in the study as internal validation. When grown in glycerol, all produced 1,3-PDO as main product.

Table 4.1 | Overview of known 1,3-PDO producers.

Organism	Genome assembly ID	B12-vitamin (in silico prediction)	1,3-PDO (mol/mol)	Acetate (mol/mol)	Reference
<i>Clostridium perfringens</i> strain 13	GCA_000009685.1	Dependent	N.A.	N.A.	Sun et al., 2003154
<i>Klebsiella michiganensis</i> M5a1 / <i>K. oxytoca</i>	GCA_000308735.2	Dependent	0.41	0.03	Yang et al., 2007155
<i>Citrobacter freundii</i> ATCC 8090	GCA_000312465.1 - GCA_000734905	Dependent	0.65	0.22	Barbirato et al., 1998156
<i>Clostridium pasteurianum</i> DSM 525	GCA_000330945.1	Dependent	0.50	0.12	Taconi et al., 2009157
<i>Halanaerobium saccharolyticum</i> DSM 6643	GCA_000350165.1	Dependent	0.41	0.26	Kivistö et al., 2012158
<i>Clostridium diolis</i> DSM 15410	GCA_000409695.1	N.A.	0.58	0.12	Otte et al., 2009159
<i>Klebsiella pneumoniae</i> ATCC 25955	GCA_000409715.1	Dependent	0.65	0.22	Barbirato et al., 1998156
<i>Clostridium butyricum</i> DSM 10702	GCA_000409755.1	Independent	0.60	0.01	Gonzalez-Pajuelo et al., 2004160
<i>Raoultella planticola</i> DSM 3069	GCA_000735435.1	N.A.	N.D.	0.15	Jarvis et al., 1997161
<i>Klebsiella pneumoniae</i> DSM 2026	GCA_000949515.1	Dependent	0.62	0.09	Mu et al., 2006162
<i>Clostridium beijerinckii</i> NRRL B-593	GCA_002006325.1	N.A.	0.59	0.00	Gungormusler et al., 2011163
<i>Caloramator viterbensis</i> DSM 13723	N.A.	N.A.	0.69	0.21	Seyfried et al., 2002164
<i>Clostridium butyricum</i> AKR102a	N.A.	N.A.	0.52	N.D.	Wilkens et al., 2012165
<i>Clostridium butyricum</i> CNCM 1211	N.A.	N.A.	0.64	0.11	Barbirato et al., 1998
<i>Clostridium butyricum</i> DSM 5431	N.A.	N.A.	0.51	N.A.	Rehman et al., 2008166
<i>Clostridium butyricum</i> E5	N.A.	N.A.	0.54	0.08	Petitdemange et al., 1995167
<i>Clostridium butyricum</i> F2b	N.A.	N.A.	0.63	0.13	Papanikolaou et al., 2004168
<i>Clostridium butyricum</i> VPI 1718	N.A.	N.A.	0.68	0.04	Chatzifragkou et al., 2011110

<i>Clostridium multifementans</i> DSM 5431	N.A.	N.A.	0.40	0.01	Biebl et al., 1999169
<i>Klebsiella pneumoniae</i> DSM 4799	N.A.	N.A.	0.45	N.D.	Jun et al., 2010170
<i>Klebsiella pneumoniae</i> HR526	N.A.	N.A.	0.40	N.D.	Xu et al., 2009171
<i>Klebsiella pneumoniae</i> XJ-Li	N.A.	N.A.	0.61	0.05	Zhang et al., 2007172
<i>Lactobacillus brevis</i> N1E9.3.3	N.A.	N.A.	0.46	0.07	Vivek et al., 2018173
<i>Lactobacillus buchneri</i> B 190	N.A.	N.A.	0.27	0.18	Schutz et al., 1984174
<i>Lactobacillus panis</i> PM1	N.A.	N.A.	N.A.	N.A.	Kang et al., 2014175
<i>Pantoea agglomerans</i> CNCM 1210	N.A.	N.A.	0.61	0.16	Barbিরato et al., 1998156
<i>Shimwellia blattae</i> ATCC 33430	N.A.	N.A.	0.45	N.A.	Rodriguez et al., 2016176
<i>Clostridium beijerinckii</i> DSM 791	N.A.	N.A.	0.55	N.A.	Pradima et al., 2017177
<i>Shimwellia blattae</i> ATC 33430	N.A.	N.A.	0.45	N.A.	Pradima et al., 2017177

MATERIALS AND METHODS

Data annotation and mining

Bacterial genomes were downloaded from the ENA database using the enaBrowserTools¹⁷⁸. All downloaded microorganisms were converted into a semantic repository using the annotation platform based on functional analysis (SAPP)¹¹⁷. All genomes were de-novo structurally re-annotated using Prodigal⁸⁹ and functionally annotated using Pfam from InterProScan. Each genome was stored as an individual database and was queried using the SAPP HDTQuery module which is a combination of Apache SPARK and SPARQL.

Oxidative branch identification

To identify strains containing the oxidative branch, as shown in figure 4.1, a filtering for DAK gene was applied. Genes containing DAK1 (PF02733) or DAK1_2 (PF13684) in combination with DAK2, (PF02734), 95 % gene confidence according to Prodigal and neighbour linking

was applied to identify strains capable of growth on glycerol.

Reductive branch identification

Strains that contained the oxidative branch were further investigated for the reductive branch using domains according to the B12-dependent or -independent glycerol dehydratase pathway (Figure 4.1). Region identification was performed using PF02287 and 20.000 bp up and downstream. Regions containing all domains of interest were further investigated for domain occurrence and synteny.

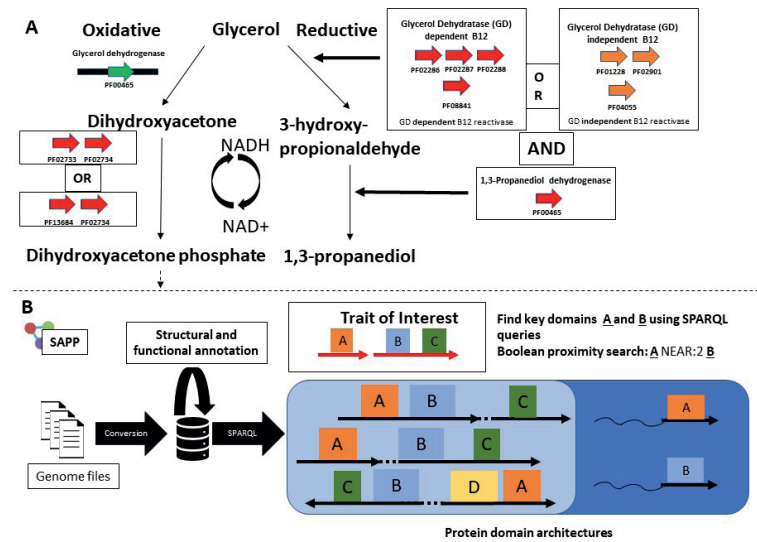


Figure 4.1 | Search strategy for 1,3-propanediol candidate species. **A.** Key domains in the main pathway for B12-dependent bioconversion of glycerol to 1,3-propanediol are indicated in red. Key domains for the alternative B12-independent reductive pathway are indicated in orange. Note the generic iron alcohol domain is used in both the oxidative and reductive branch but not included in the search strategy for the oxidative branch (indicated in green). **B.** Generalised functional based search strategy for traits using SAPP. Genome sequences in standard format are converted to an RDF database and complemented with structural and functional annotation. SPARQL, search strategies are deployed to identify domains of interest (dark blue) and complemented with proximity searches (light blue) to find key domain enriched regions.

Phenotypic information

Phylogenetic information was complemented with phenotypic information through BacDive. The BacDive resource¹⁷⁹ was parsed and each entry record was transformed into Resource Description Framework (RDF) allowing the integration of genotypic and phenotypic information. SPARQL queries were used to retrieve information with regards to pathogenicity and temperature.

Physiological analysis

The *Clostridium magnum* DSMZ 2767¹⁸⁰ was obtained from German Collection for Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). *Trichococcus* strain ES5 and *T. pasteurii* were stored locally. The strains of *Acetobacterium woodii* DSMZ 1030¹⁸¹ and *Acetobacterim wieringae* DSMZ 1911¹⁸⁰ were previously locally stored. The anaerobic growth media for *C. magnum*, *A. wieringae* and *A. woodii* growth experiments was prepared as previously described⁸⁰ with the addition of 1 g/L yeast extract. The medium was added as 45 mL in 120 mL serum bottles and sealed with rubber stoppers and aluminium caps. Vials of the medium were flushed with N₂/CO₂ (80/20 v/v) to a final pressure of 1.5 bar. Before inoculation, the medium was supplemented with 0.5 mL of slates solution and 2.5 of bicarbonate solution. *C. magnum* required 0.5 g/L of cysteine as additional reducing agent. The inoculation of all strains was 5 % of a total serum bottle. *T. pasteurii*, *Trichococcus* strain ES5 *C. magnum*, *A. wieringae* and *A. woodii* were grown on glycerol in a concentration of 20 mM. The yield was measured as 1,3-PDO (mol) divided by utilised glycerol (mol). Growth measurements were based on optical density (OD) using a spectrometer (Hitachi U-1500, Labstuff, Capelle aan den IJssel, The Netherlands). All soluble substrates and intermediates were measured with an Agilent HPLC system equipped with Agilent Metacarb 67H column (3,006.5 mm) (Thermo Fisher Scientific, MA) and a refractive index detector. The OD and HPLC measurements were conducted in time points of 0, 24 and 48 hours. All experiments were performed in triplicate.

RESULTS

Development of the genome mining workflow

To cope with the huge amount of biological input data, we followed a KDD process (Figure 4.2). In this systematic multistep process, the actual 'pattern searching' step is preceded by the equally important steps of 'data preparation' and 'incorporation of prior knowledge'¹⁸². For metabolic trait discovery, the KDD process was adjusted to the specific needs in bioprospecting. Data silo and transformation enabled an important level of interoperability, pattern searching, pattern validation and modification using a training data set and data-mining.

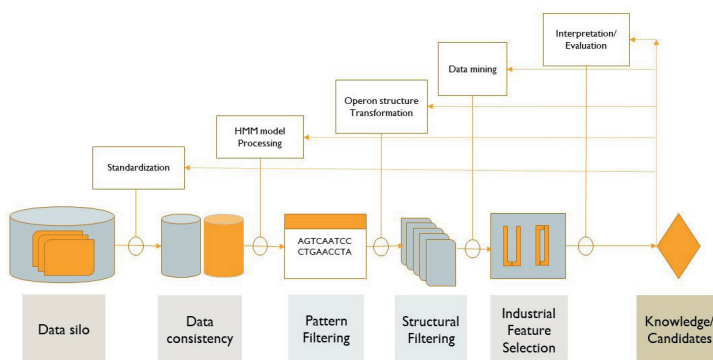


Figure 4.2 | The KDD process is hallmarked by several steps and iterations.

Incorporation of prior knowledge

In 1,3-PDO production, two parallel pathways are used for dissimilation of glycerol. In the oxidative pathway glycerol is dehydrogenated to dihydroxyacetone by a NAD^+ -linked glycerol dehydrogenase, then to dihydroxyacetone phosphate by an ATP-dependent dihydroxyacetone kinase¹⁸³. Additionally, two alternative reductive pathways, either dependent on vitamin B12 or not, exist for 1,3-PDO production. To

validate of the search strategy, and to gain further genetic insights into the domains involved in 1,3-PDO formation, a list of known 1,3-PDO producing strains was obtained. From the literature, 29 strains have been reported to produce 1,3-PDO (Table 4.1). A training dataset consisting of genome sequences from eleven different strains that had the genomes available in the EBI-ENA database¹⁸⁴ was used as genome sequences are not available for all of them. For *Citrobacter freundii* ATCC8090, two draft genome sequences of comparable quality could be obtained and initially both were kept increasing the number of genomes to twelve. For four strains in this set the 1,3-PDO operon was molecularly characterised. In addition, *Trichococcus* strain ES5 and *Trichococcus pasteurii* were shown to produce 1,3-PDO and have a large operon including 17 genes for the trait (Chapter 3). As the two species had been studied extensively, we used them in this study for cross-validation purposes.

Data transformation

To obtain a high degree of interoperability, and for the inclusion of data provenance, the nine genome sequences were *de novo* structurally and functionally annotated through the SAPP framework using the modules Prodigal⁸⁹ for gene prediction and InterProScan⁹⁰ for protein annotation. Genome annotations were exported as Linked Data in graph database using the RDF as a data-metadata model¹¹⁷. Protein domain architectures were used as proxy for protein function. As no differences were observed between the functional annotations of two *Citrobacter freundii* ATCC8090 draft genome sequences they were treated as one.

Development of a function-based search strategy for the oxidative branch

The first two enzymes essential for the oxidative pathway method of 1,3-PDO production are glycerol dehydrogenase which is dependent on NADH and produces dihydroxyacetone (DHA), and DHA kinase which phosphorylates DHA. The four Pfam domains describing these

key reactions are Fe-ADH (PF00465) for glycerol dehydrogenase and DAK1 (PF02733) or DAK1_2 (PF13684) both capturing the kinase domain of the dihydroxyacetone kinase family in combination with DAK2 (PF02734), capturing phosphatase domain of the dihydroxyacetone kinase family (Figure 4.3). A SPARQL query for DAK2 domain neighbours in the training set database followed by a Boolean DAK1/DAK2 or DAK1_2/DAK2 proximity search in the search results, showed that the genomes of all 1,3-PDO producing species in the training data set encoded at least one dihydroxyacetone kinase family protein with a DAK1/DAK2 or DAK1_2/DAK2 protein architecture; five strains coded for DAK1/DAK2 and four strains DAK1_2/DAK2. Three *Clostridia* strains, *C. perfringens* str. 13, *C. pasteurianum* DSM 525 and *C. diolis* DSM 15410 encode both configurations (Table 4.1). Another observation is the high persistency and copy number of the Pfam domain family of iron-containing alcohol dehydrogenase/aldehyde reductases (PF00465) with on average 16.6 copies per genome. Due to the inherent low discriminative power of this domain, it was decided to only use a DAK1/DAK2 OR DAK1_2/DAK2 proximity and not to include PF00465 in the large-scale functional screening as preselection for the oxidative branch.

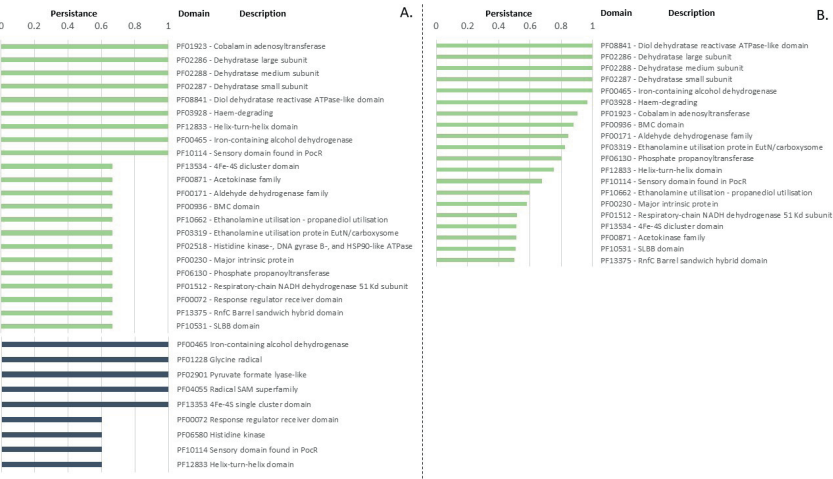


Figure 4.3 | Overall domain compositions. A. Overall domain composition of the B12-dependent and B12-independent 1,3-PDO operon derived from the training data set B-12 dependent domains are with green and B12-independent domains with grey. Input strains are shown in Table 4.1. B. Overall domain composition of

B12-dependent candidates after screening. Beside the four specified domains the core was expanded by the '4Fe-4S single cluster domain' (PF13353). Other domains highly persistent in B12-dependent operon such as the 'sensory domain found in PocR' and the 'helix-turn-helix domain' are also enriched in some of the putative B12-independent operonic structures except in *Clostridium butyricum* DSM 10702.

Development of a function-based search strategy for the reductive branch

For 1,3-PDO production two alternative pathways are known, a well-studied B12-dependent pathway and a lesser studied oxygen sensitive B12-independent pathway (Figure 4.1). A series of SPARQL queries were used to obtain the persistency and copy number of key domains in the two alternative pathways among the 12 genomes of the training set (Table 4.2).

Table 4.2 | Properties of key domains involved in glycerol dissimilation in 1,3-PDO producers.

* Conditional mean excluding zero values.

Domain	Mean Copy number*	Proximity Search query (compounds and distance)
Oxidative pathway		
PF02733 (DAK1)	2.2 1.6 2.7	DAK1 AND DAK2 OR DAK1_2 AND DAK2 (immediately adjacent)
PF13684 (DAK1_2)		
PF02734 (DAK2)		
B12-dependent reductive Pathway		
PF00465	16,6	All, within 20.000 up or downstream of the B12 dependent dehydratase domains
PF02286	1.7	
PF02287	1.7	
PF02288	3.7	
PF08841	1.7	
B12-independent reductive Pathway		
PF01228	13.8	All, within 20.000 up or downstream of the B12 independent dehydratase domains
PF02901	14.4	
PF04055	32.4	

The B12-independent glycerol dehydratase is reported to be involved in 1,3-PDO production¹⁸⁵ and consists of a glycine radical (PF01228) and a pyruvate formate lyase-like domain (PF02901). Both domains are present in all twelve genomes of the training data set with a relatively high copy number of six. The persistency of the radical SAM superfamily domain, a key part of the B12-independent glycerol dehydratase reactivase enzyme⁵⁴, was also with an average copy number of 32 indicating that these three domains are promiscuous and normally also used to support other functionalities (Table 4.1).

As key domains of both the B12-dependent and independent pathways were either too generic indicated by a high copy number or were associated with a low level of persistency, a Boolean multi-compound proximity search was developed for the reductive branch. In this approach, the physical co-localisation of key domains in the respective genomes was included in the search. From the list of natural producers with a molecularly characterised 1,3-PDO operon, the size of 1,3-PDO operon was estimated to encompass approximately 18,000 nucleotides. Furthermore, signifying domains can be found on both strands due to the use of internal promoters (Figure 4.4). Including this, the search criteria were set up so that the region of interest for B12-dependent pathway should contain at least five signature domains (PF00465, PF02286, PF02287, PF02288, PF08841) in a window of 40,000 bp extending 20,000 up and downstream of the dehydratase domains. Based on the single characterised B12-independent strain, four domains were specified (PF00465, PF01228, PF02901, PF04055) extending 20,000 up and downstream of the dehydratase domains. Furthermore, taking internal promoters into account, domains could be present on both strands.

Applying these criteria to the training data set resulted in the identification of three strains containing B12-dependent operon signatures; *Citrobacter freundii* ATCC8090, *Klebsiella pneumoniae* DSM2026, and *Clostridium perfringens*; four strains containing the B12-dependent as well as B12-independent operon signatures; *Clostridium pasteurianum* DSM525, *Halanaerobium saccharolyticum* DSM6643, *Klebsiella michiganensis*, and *K. pneumoniae* ATCC25955) (Figure 4.3); and four strains solely containing the independent operon signatures;

Clostridium beijerinckii NRRL B-593, the *Clostridium butyricum* E4, *Clostridium diolis* DSM15410 and *Raoultella planticola* DR3.

Data silo

To mine publicly available genomes for the presence of the vitamin B12-dependent 1,3-PDO operon 84,329 bacterial genomes containing, 51 phyla, 64 classes, 145 orders, 335 families, 1,126 genera and 2,661 species were obtained from the EBI-ENA database. The SAPP semantic annotation framework was subsequently used for a *de novo* structural and functional annotation of these genomes resulting in a semantic database of 365,920,933 predicted protein encoding genes linked to the corresponding protein sequences, predicted protein domain architectures and structural and functional prediction provenance. Further analysis was performed on predicted protein encoding genes with a Prodigal confidence score of at least 95 %. At this threshold 95.1 % of the proteins remain pertaining 98.3 % of the assigned protein domains.

Data mining for candidate 1,3-PDO producers using the vitamin B12-dependent pathway

Oxidative branch: Following the function-based search strategy outlined above and summarised in figure 4.1, a proximity search with the two DAK configurations in parallel resulted in a 55 % reduction of the search space. The reduced search space consisted of 37,791 genomes with in total 834 distinct species in total from 158 genera. The most abundantly present species were *Streptococcus pneumoniae*, *Listeria monocytogenes* and *Klebsiella pneumoniae*, each with more than a thousand strains. The reduced search space was used as input for a proximity search of the reductive branch.

Reductive branch: Identification of B12-dependent strains according to the criteria used for known strain validation, resulted in 187 species being found, including the two *Trichococcus* species (4,142 strains). Known producers' strains of *C. pasteurianum*, *L. brevis*, *L. panis* and *L. reuterii* were identified in the screening. Strains of *L. monocytogenes*

(1,777) and *K. pneumoniae* (1,605), both pathogenic species were overrepresented.

At species level, domain persistency of the B12-dependent-1,3-PDO trait showed a high degree of similarity with the training set (Table 4.1). The most frequently observed additional domains showing a local persistency of 0,80 and above are PF03928 (0,97), PF01923 (0,90), PF00936 (0,87), PF00171 (0,84), PF03319 (0,82) and PF06130 (0,80) (Figure 4.3). PF03928, a haem degrading domain often flanked by PF03319 and PF06130 (in the order of PF03319-PF03928-PF06130), can be of importance for aldehyde reductase which can have different cofactor specificities including haem¹⁸⁶. PF01923 represents an enzyme catalysing the conversion of cobalamin (vitamin B12) into one of its coenzyme forms, adenosylcobalamin (coenzyme B12, AdoCbl), which is vital for the functioning of B12-dependent glycerol dehydratase¹⁸⁷. The compartmentalisation domain, PF00936, could be used to encapsulate enzymatic steps important for 1,3-PDO production¹⁸⁸. PF00171 has been shown to enhance 1,3-PDO production¹⁸⁹. The ethanolamine utilisation protein EutN/carboxysome (PF03319) is involved in the cobalamin-dependent degradation of ethanolamine. PF06130 a signature for phosphate propanoyltransferase is involved in phosphorylation of 3-hydroxypropionyl-CoA to 3-hydroxypropionyl phosphate¹⁹⁰.

Experimental validation of candidate strains

Phenotype selection: Currently, the best known 1,3-PDO natural producers are either pathogenic or opportunistic pathogenic microorganisms such as *K. pneumonia* and *C. butyricum*¹⁵², a trait that should be avoided when possible. Secondly, as the candidate species are dependent on vitamin B12 to produce 1,3-PDO it could be beneficial to additionally select for vitamin B12 synthesis. In order to obtain the biosafety level of each candidate species another BacDive database was queried¹⁷⁹. Forty-seven candidate species were classified at biosafety level 2 and higher and 48 species were non-pathogenic (biosafety level 1) while 92 species remained unclassified. A function-based screening was developed for the distribution of the trait for

vitamin B12 synthesis among the candidate species. The metabolic pathway of this trait is composed of seven enzymes, and 86 species were found to contain all domain signatures required for vitamin B12 synthesis. Finally, the ability to synthesize vitamin B12 synthesis in combination with biosafety level 1 resulted in nine different species being identified.

Experimental validation: From the nine-different species, *Clostridium magnum* was selected for experimental validation as a non-pathogenic strain with the capability of B12 synthesis. *Trichococcus* strain ES5 and *T. pasteurii* were used as positive control in the trials. As a negative control, *A. wieringae* and the closely related *A. woodii* were used. The putative 1,3-PDO operon of *A. wieringae* lacks the key domain PF04055 and therefore the operon lacks the aldehyde reductase function. *A. woodii* lacks the complete operon (Figure 4.4, panel B)

All five strains were grown on glycerol. *C. magnum*, *T. pasteurii* and *Trichococcus* strain ES5 produced 1,3-PDO as the main product with acetate as a by-product. *A. wieringae* showed significantly slower growth. In comparison to *C. magnum* and the two *Trichococcus* species, the production of 1,3-PDO was significantly reduced and the main product was acetate (Figure 4.4, panel C). In our hands, *A. woodii* was not able to grow on glycerol.

DISCUSSION

Bioprospecting entails the systematic search for economically valuable genetic and biochemical resources from nature with a view to creating biotechnological applications¹⁹¹. On the other hand genome prospecting, the *in silico* mining of sequenced genomes and metagenomes for new biotechnologically relevant proteins and enzymes, is a relatively new field. In genome prospecting two approaches can be applied, a “top-down” approach that begins by searching for a function and is followed by identification of the corresponding gene(s), and a “bottom-up” approach that starts with a gene of interest to find a corresponding function in genome(s) of interest.

For bottom-up approaches many sequence similarity tools exist, such

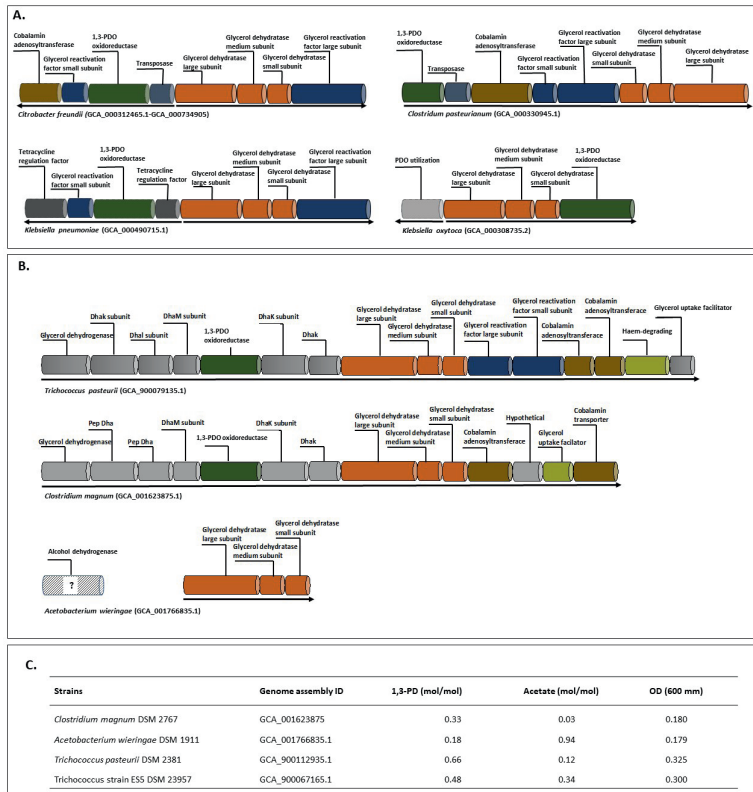


Figure 4.4 | Representation of operonic structures for the metabolic trait of 1,3-propanediol production. A. Genetic architecture of the 1,3-PDO operon of previously molecularly characterized species. B. Genetic architecture of the 1,3-PDO operon of selected species. C. 1,3-PDO and acetate yields from glycerol fermentations of selected strains. The coloured blocks represent genes. Arrows indicate direction of transcription. Species names are indicated with genome sequences in brackets. Note that *A. wieringae* operonic structure does not include a 1,3-PDO dehydrogenase.

as BLAST¹⁶. Bottom-up studies usually start with a selection of (meta) genome sequences followed by sequence similarity-based clustering and selection of candidate sequences and reannotation of interesting candidates thereby avoiding ambiguity related problems in current functional annotations. For bacterial species, however, a priori gene fusion-fission events can be expected. This hampers sequence similarity-based detection¹⁴⁹ and clustering of multi-domain proteins

encoding genes, while the same domains may also be present in multiple proteins meaning that first-matches genomic sequences in a sequence similarity search may not encode the desired function.

When searching for polygenic traits, these problems are aggravated and a function-based approach searching for key functions may be more effective, especially when there is insufficient understanding of the genetic architecture of trait in terms of the minimal number of genes required for the trait and function of the domains encoded by the different genes.

Protein domains have been shown to provide an accurate representation of the functional capabilities of a protein¹⁵¹. To overcome ambiguity related problems in functional annotations SAPP identifies and annotates protein function based on domain architecture. Hidden Markov Models (HMM) profiles favour in their scoring functionally important sites in proteins. Therefore, they are robust in annotating proteins domain over large phylogenetic distances.

KDD is a multi-step process involving data preparation and transformation, pattern searching, evaluation and iteration after modification. By using protein domain architectures as proxy for protein functions a high level of standardisation is obtained. As protein domain architectures can be directly transformed into highly interoperable strings of Pfam domain identifiers “top-down” functional screenings can be carried out efficiently.

Applying the KDD approach to Linked Data allows for the validation of initial results in multiple ways and iteration after modification. By using molecular knowledge obtained from molecular characterisations of the 1,3-PDO operon of four species and validating and iterating the search pattern on a training data set of genome sequences of twelve known 1,3-PDO producers, a large collection of 84,300 publicly available bacterial could be efficiently mined in a top-down approach yielding 178 new candidate species, and the successful experimental verification of three of these species was achieved.

Driven by a strong need to be able to integrate and analyse biodata across databases, there has been a considerable increase in the adoption of Semantic Web technologies in the life-sciences¹⁹². However, a

SPARQL endpoint for phenotypic data is only available via WikiData¹⁹¹. Other resources for phenotypic data such as BacDive, here used to obtain biosafety levels of candidate species, do provide API's to mine their data, but currently cannot be directly queried using SPARQL. With the growing importance of Semantic Web technologies for the life sciences interoperability levels on all aspect will increase, enhancing mining possibilities, and aiding the discovery of new traits at an unprecedented pace.

CONCLUSION

Through the transformation of genomic data into a FAIR linked-data format, iterative function-based approaches can be developed to mine the large genome repositories. By presenting functional annotation as unambiguous protein domain architectures an elevated level of interoperability is obtained allowing for the development of efficient function-based top-down searches not limited to supervised trait identification.

ACKNOWLEDGEMENTS

This research was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science. The work conducted by the U.S. Department of Energy Joint Genome Institute (DOE-JGI), a DOE Office of Science User Facility, was supported by the Office of Science of the DOE under Contract No. DE-AC02-05CH11231.

IBISBA 1.0 - Project ID: 730976 - Funded under: H2020-EU.1.4.1.2. - Integrating and opening existing national and regional research infrastructures of European interest. This work was carried out on the Dutch national e-infrastructure with the support of the SURF foundation.

CHAPTER 5

PHYSIOLOGICAL AND PROTEOMICS ANALYSIS OF 1,3-PROPANEDIOL METABOLISM IN *TRICHOCOCCUS* *PASTEURII* AND *CLOSTRIDIUM* *BUTYRICUM*

Nikolaos Strepis^{1,2}, Lala Akhundova¹, Arianna Palma¹, Tom van de Weijer¹, Tom Schonewille², Sjeff Boeren³, Peter J Schaap², Alfons JM Stams^{1,4}, Diana Z Sousa^{1*}

1. Laboratory of Microbiology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

2. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

3. Laboratory of Biochemistry, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands

4. Centre of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

ABSTRACT

Trichococcus pasteurii is a 1,3-propanediol (1,3-PDO) producer with an operon of 17 genes related to its production, including the vitamin B12-dependent glycerol dehydratase gene. In contrast, another 1,3-PDO producer, *C. butyricum* VPI 3266 does not contain a related operon, but just vitamin B12-independent glycerol dehydratase, glycerol dehydratase reactivase and 1,3-PDO dehydrogenase genes. In this study, we performed physiological trials and proteomics to assess the 1,3-PDO metabolism in the two species. In media with 5 g/L glycerol and 1 g/L yeast extract, *T. pasteurii* had a 1,3-PDO production yield of 0.70 ($\text{g}_{1,3\text{-PDO}}/\text{g}_{\text{consumed glycerol}}$), which was higher than *C. butyricum*'s yield (0.47, $\text{g}_{1,3\text{-PDO}}/\text{g}_{\text{consumed glycerol}}$). All the enzymes encoded in the *T. pasteurii* operon were detected in the proteome. Comparably, in the *C. butyricum* proteome, the B12-independent glycerol dehydratase, 1,3-PDO dehydrogenase and glycerol dehydratase reactivase were detected, supporting the different route of glycerol fermentation. Additional studies of *T. pasteurii*, included proteomics at 45 g/L glycerol and fermentation of glycerol in continuous bioreactors. The production of 1,3-PDO was stable in the continuous fermenters. However, at high glycerol concentration, an inhibition of growth was observed in batch and bioreactors fermentations, which was caused by the formation of 3-hydroxypropionaldehyde. Proteomics indicated that glycerol dehydratase was significantly more abundant than 1,3-PDO dehydrogenase at 45 g/L glycerol. This study verifies the presence of enzymes encoded from the large operon regarding 1,3-PDO production in *T. pasteurii* and reports the metabolic response of the species in low and high glycerol concentration. The metabolic knowledge of *T. pasteurii* promotes its potential as a natural producer of 1,3-PDO.

INTRODUCTION

The worldwide production of biodiesel is estimated to reach 30×10^6 T in 2021¹⁹³. Glycerol is the main by-product of the process (10 % w/w)⁵¹. A possible use of glycerol is its conversion to 1,3-propanediol (1,3-PDO)⁵². As a versatile building molecule, 1,3-PDO can be used for the production of pharmaceuticals, cosmetics, foods and biodegradable polyesters such as polytrimethylene terephthalate⁵². The chemical catalysis of glycerol to 1,3-PDO requires high temperatures, the usage of expensive catalysts and results in the formation of toxic by-products¹¹⁰. An alternative and environmentally friendly option would be a microbial conversion of glycerol to 1,3-PDO.

Production yields of the natural microbial producers ranges from 0,50 to 0,70 ($\text{g}_{1,3\text{-PDO}}/\text{g}_{\text{consumed glycerol}}$) (Table S5.1). The main producers are *Clostridium pasteurianum*, *Clostridium butyricum*, *Klebsiella pneumoniae* and *Citrobacter freundii*. The microbial fermentation of glycerol is based on reductive and oxidative reactions. 1,3-PDO is produced by the reductive branch, where glycerol is dehydrated to 3-hydroxypropionaldehyde (3-HPA) by a glycerol dehydratase. Then, 3-HPA is reduced to 1,3-PDO by a NADH-dependent 1,3-propanediol dehydrogenase¹⁹⁴. Most of the described microbial 1,3-PDO producers described have a glycerol dehydratase that requires vitamin B12 (cobalamin) as a co-factor¹⁸⁷. A vitamin B12-independent variant also exists and was reported for two strains of *C. butyricum* (VPI 3266 and VPI 1718)^{185,195}. Recently, *Trichococcus pasteurii* and *Trichococcus* strain ES5 were observed to produce 1,3-PDO^{48,108}. These two bacteria have a conserved operon of 17 genes dedicated to 1,3-PDO production, including a B12-dependent glycerol dehydratase (Chapter 3).

Here, we conducted physiological and proteomics studies with *T. pasteurii* and *C. butyricum* to gain insights into the metabolic differences between the B12-dependent and B12-independent routes for the production of 1,3-PDO. To further evaluate the potential of *T. pasteurii* continuous bioreactors were used for 1,3-PDO production.

MATERIAL AND METHODS

Microorganisms and growth medium

Trichococcus pasteurii (DSM 18806^T) and *Clostridium butyricum* VPI 3266 (DSM 10702^T) were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). Growth experiments for *T. pasteurii* were performed with an anaerobic basal medium prepared as previously described⁸⁰. Cultivations were conducted on serum bottles of 120 mL contained 50 mL of the medium and were sealed with rubber stoppers and aluminum caps. The headspace of the bottles was flushed with N₂/CO₂ (80/20 (v/v)) to a final pressure of 1.5 atm. Yeast extract, glycerol and vitamins were added to the medium (Figure S5.1). For the physiological trials, all growth and soluble metabolites measurements were taken at 0 and 24 hours for *T. pasteurii* and at 0 and 44 hours for *C. butyricum*.

Analytical methods

Growth was monitored by measuring the optical density (OD 600 nm) with a spectrometer (Hitachi U-1500; Labstuff, Capelle aan den IJssel, the Netherlands). Cultures of 250 ml were centrifuged for 5 min at 7,000 x g at 4 °C and after supernatant was removed, 2.5 mL of Milli-Q water (Millipore Corporation, Burlington, Ma) was added. The re-suspended pellet was added to the pre-weighed aluminum dish. Dishes were placed in an oven (105 °C) for 24 h. After that, the dishes were cooled down and weighed again. Dry biomass was measured for cultures with different OD_{600nm} resulting in the following correlation: $OD_{600nm} = 2.3711 \cdot X_{biomass} + 2E-16$, where $X_{biomass}$ is dry biomass in g/L.

Organic soluble compounds, such as glycerol, 1,3-propanediol, lactate, acetate and formate were quantified by high performance liquid chromatography (HPLC) in SpectraSystem HPLC system using Agilent Metacarb 67H column (300x6.5 mm) (ThermoFisher Scientific, Waltham, MA), which had sulfuric acid (5 mM) as a mobile phase at a flow rate of 0.8 mL min⁻¹ and the temperature was adjusted to 70 °C (for a clear peak separation of lactate). Lactate was quantified by us-

ing an ultra violet (UV) detector, while acetate, formate, 1,3-PDO and butyrate were quantified using an infrared detector (IR). The yield of 1,3-PDO was calculated from triplicate (independent) incubations and expressed as $g_{1,3-PDO}/g_{consumed\ glycerol}$.

The compound 3-HPA quantification was based on the colorimetric method accordingly to Cohen and Altshuler¹⁹⁶ and microtiter plates were used. The absorbance of each well was measured at 605 nm with a microplate reader Synerg Mx Microplate (Biotek, Winooski, WI). Samples used for 3-HPA derived from cultures that were grown without resazurin which interferes with the measurement of optical density. These analyses were conducted in samples from duplicates (independent) incubations.

Proteomics sample preparation

Proteomics conditions were defined as 1 g/L yeast extract and 5 g/L glycerol for *T. pasteurii* and for *C. butyricum*. Additionally, proteomics samples were prepared for *T. pasteurii* on 1 g/L and 45 g/L glycerol. Three replicates of 250 mL culture were conducted for each pre-defined condition. The cells from each individual culture were harvested by centrifugation (15,700 x g for 10 min), when an average OD of 0.75 was achieved (Figure S5.2). The cell pellet was transferred to a STD-lysis buffer (100mM Tris/HCl pH 7.6+4 % SDS, vol/vol+0.1 M dithiothreitol). Lysis of the cells was conducted with a French press cell Thermo IEC (ThermoFisher Scientific, Waltham MA) with four runs at 1,900 psi. To collect proteins, samples were centrifuged. For every sample, 50 ug of protein was loaded in SDS-PAGE on a 10-well SDS-PAGE 10% (wt/vol) Bis-Tris Gel (Mini Protean System, Bio-Rad, San Diego, CA). The voltage was maintained at a constant voltage of 120V for 55 min using Tris-SDS as a running buffer. Staining of the gels was performed with a Colloidal blue staining kit (Life Technologies, Carlsbad, CA) and they were treated for reduction and alkylation using 10mM dithiothreitol and 15mM iodoacetamide in 50mM ammonium bicarbonate. All samples lanes on the gel were sliced in 4 four pieces and then cut again to small slices of approximately 1-2 mm. Peptide digestion was performed with

50 μL of sequencing grade trypsin (5 ng/ μL in 50 mM ammonium bicarbonate) and at room temperature, incubated overnight while shaking. The digested peptides were desalted and were LC-MS/MS ready. The nano LC-MS/MS is using a Proxeon Easy nanoLC and an LTQ-Orbitrap XL instrument (Thermo Fisher Scientific, Waltham, MA) as previously described⁴¹.

Proteomics analysis

Spectra from the quantified peptide were analysed with the software Maxquant¹⁹⁷ using the parameter of "Specific Trypsin/P" Digestion mode with maximum two missed cleavages. For the Andromeda peptide search engine, the default parameters were maintained. The genomes as well the annotation of *T. pasteurii* (GCA_900079135.1) and *C. butyricum* (GCA_000409755.1) are available in the public genome repository ENA (EMBL, Cambridgeshire, UK). As a protein database for *T. pasteurii* and *C. butyricum* was used annotated proteins from UniProt database were used. The proteomic analysis was conducted with the MaxQuant pipeline and the software Perseus was used for statistical analysis¹⁹⁷. To calculate the protein fold change from the two *T. pasteurii* conditions the division of Label Free Quantification (LFQ) values in Maxquant was used. For an estimation of proteins abundance in the samples of *C. butyricum* and *T. pasteurii* were used the values of Intensity Based Absolute Quantification (IBAQ) in Maxquant¹⁹⁷. The FDR for matching peptides with proteins were set at 0,01. Protein fold change was described as significant when the p-value was lower than 0,05. Further filtering criteria were implemented such as every protein must have minimum of two peptides, of which at least one should be unique and at least one should be unmodified. Functional analysis was based on protein functional domains based on Pfam families¹⁹ and InterProScan⁹⁰. The transmembrane regions were predicted based on TMHMM¹¹⁹ and SignalP¹⁹⁸. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to study the identified proteins in metabolic pathways.

Batch and continuous bioreactors

Batch fermentations were performed under anaerobic conditions in Sartorius Biostat® Q Plus bioreactors (Sartorius, Göttingen, Germany) of 1 L total volume and 500 mL working volume. Conditions were set and controlled at 30 °C, 200 rpm, pH (7.0) and an N₂ flow rate of 0.1 vvm. NaOH of 5 mM was used to control the pH. The media in the bioreactors was identical as in the bottle batch cultivations supplemented with 1 g/L yeast extract. Cysteine of 0.05 g/L was added to reduce the medium. Glycerol was used as the sole substrate in concentrations of 1.8, 3.7 and 20 g/L. A grown culture of *T. pasteurii* was used to inoculate the bioreactors (1 %, (v/v)). Continuous fermentations were carried out in triplicates using the same conditions as in the batch fermentation. The continuous feed had a dilution rate of 0.06 h⁻¹. The growth medium for continuous fermentation was prepared in 20 L bottles with 1 g/L yeast extract. The feed was supplemented with increasing concentrations of glycerol: 1.8 g/L (days 0-71), 3.7 g/L (days 72-91) and 20 g/L (days 92-105).

RESULTS

Effect of yeast extract concentration

Previously, *K. pneumonia*, *Ct. freundii*, *C. pasteurianum* and *Pantoea agglomerans* were reported to achieve 1,3-PDO yields ranging from 0.47-0.75 (g_{1,3-PDO}/g_{consumed glycerol}) with 1 to 5 g/L yeast extract in the medium (Table S5.1). The effect of yeast extract on glycerol fermentation in *T. pasteurii* was studied in a similar range. With 1 g/L yeast extract, *T. pasteurii* reached an OD_{600nm} of 1.00 after 24 h of incubation and glycerol was completely consumed. The OD_{600nm} of cultures grown without yeast extract or with concentrations lower than 1 g/L was significantly lower and glycerol was not completely consumed (Table 5.1). Growth on the yeast extract alone (5 g/L) was not observed. Therefore, 1 g/L yeast extract was considered as the optimal yeast extract concentration for achieving a high OD_{600nm}, and was chosen for the glycerol concentration trials. The effect of yeast extract on the growth of *C. butyricum* was previously studied in a range of 1, 2 and 5 g/L yeast extract (Table S5.1).

Effect of glycerol concentration

To identify the optimal glycerol concentration for the highest 1,3-PDO yield, a range of 0 to 45 g/L was tested for *T. pasteurii* and *C. butyricum*. The yeast extract was 1 g/L and the comparisons were conducted with an incubation of 24 hours. Complete conversion of glycerol was observed at 5 g/L and the maximum OD_{600nm} value was achieved (Table 5.1). At that concentration, we observed the highest production yield of 1,3-PDO, 0.70 (g_{1,3-PDO}/g_{consumed glycerol}). Higher glycerol concentrations led to a lower 1,3-PDO production, the substrate was not completely consumed, and low OD_{600nm} values were observed.

Besides 1,3-PDO in all the fermentation trials, other measured soluble compounds were acetate, lactate and formate. In most glycerol cultivations, the acetate/lactate ratio was observed to be stable at 2.40 (mM/mM). The highest ratio was 3.76 (mM/mM) on 10 g/L glycerol concentration. For the fermentations on high glycerol concentrations a low carbon recovery was calculated (less than 71 molC recovered in the products) which would indicate the formation of another major product. Additionally, these cultures were described as having an orange/yellow colour. Subsequently, this product was identified as 3-hydroxypropanaldehyde (3-HPA) with concentrations ranging from 16 to 36 mM depending on the initial glycerol concentration.

In comparison, *C. butyricum* cultures with 5 and 10 g/L glycerol shown a complete substrate consumption (Table 5.1). Specifically, at 5 g/L glycerol, 1,3-PDO production was the maximum with a yield of 0.47 (g_{1,3-PDO}/g_{consumed glycerol}). In higher glycerol concentrations than 10 g/L, the substrate consumption and the 1,3-PDO production was reduced. In addition to 1,3-PDO production, *C. butyricum* produced acetate, butyrate, formate and lactate. Lactate was only detected at glycerol concentrations higher than 10 g/L. Butyrate was produced in all glycerol concentrations with the highest concentration observed at 30 g/L glycerol (27 mM). The acetate/ butyrate ratio during *C. butyricum* fermentations ranged from 0.2 to 1.8 (mM/mM), increasing in higher

Table 5.1 | Cultivation trials for defining the most suitable variables for *T. pasteurii* and *C. butyricum*. Samples for *T. pasteurii* were collected at 24 h and for *C. butyricum* at 44 h. *T. pasteurii* was tested in a range from 0 to 5 g/L yeast extract and in a range of 0 to 45 g/L glycerol in batch cultures. *C. butyricum* was tested in a range of glycerol concentration from 0 to 45 g/L in batch cultures. N.D.: Not Detected

Yeast extract(g/L)	Initial substrate (mM)	Consumed substrate (mM)	1,3-PDO (mM)	Acetate (mM)	Formate (mM)	Lactate (mM)	3-HPA (mM)	Butyrate (mM)	Yield 1,3-PDO production (mol/mol)	OD (600 nm)	Carbon recovery (%)
<i>T. pasteuri</i>											
Yeast extract trial											
0	5.0 (54.8±0.3)	48.1±2.0	19.1±2.9	9.0±0.9	N.D.	1.8±0.2	N.D.	N.D.	0.39±0.07	0.3±0.0	90.1
0.1	5.0 (53.8±0.9)	53.8±0.9	29.1±5.4	11.2±2.2	4.4±0.9	0.8±0.3	N.D.	N.D.	0.54±0.10	0.6±0.0	101.5
1	5.0 (54.7±3.5)	54.7±3.5	38.3±1.9	13.2±0.0	5.4±0.3	N.D.	N.D.	N.D.	0.70±0.05	1.0±0.0	113.3
2	5.0 (55.6±4.7)	55.6±4.7	31.5±1.1	11.0±0.2	5.9±0.1	0.4±0.0	N.D.	N.D.	0.56±0.06	1.0±0.0	99.7
3	5.0 (52.6±2.2)	52.6±2.2	31.9±3.3	11.1±1.1	5.6±0.3	0.6±0.0	N.D.	N.D.	0.60±0.06	1.0±0.0	110.9
4	5.0 (51.0±3.0)	51.0±3.0	29.8±3.1	9.8±0.9	4.1±0.1	0.7±0.0	N.D.	N.D.	0.58±0.07	1.1±0.0	113.4
5	5.0 (57.7±4.0)	57.7±4.0	34.1±0.3	11.4±0.5	5.4±0.3	0.8±0.0	N.D.	N.D.	0.59±0.05	1.0±0.0	112.9
Glycerol trial											
1	0.0 (0.0±0.0)	0.0±0.0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	0.0±0.0	0
1	5.0 (54.7±3.5)	54.7±3.5	38.3±1.9	13.2±0.0	5.4±0.3	N.D.	N.D.	N.D.	0.7±0.05	1.0±0.0	113.3
1	10.0 (110.9±2.4)	90.0±2.9	49.3±0.7	18.8±0.3	N.D.	5.0±0.0	16.6±0.7	N.D.	0.54±0.03	0.8±0.1	112.6
1	20.0 (218.0±6.6)	91.1±8.2	43.2±1.0	20.2±0.0	N.D.	7.6±0.1	28.0±2.4	N.D.	0.47±0.08	0.7±0.0	120.9
1	30.0 (323.9±6.8)	102.1±7.2	43.7±0.1	19.1±1.2	N.D.	8.1±0.3	30.2±0.8	N.D.	0.42±0.07	0.7±0.0	109.5
1	40.0 (425.3±10.0)	114.2±14.0	42.6±0.4	19.8±0.3	N.D.	8.2±0.3	30.7±2.8	N.D.	0.37±0.13	0.7±0.0	98.3
1	45.0 (487.2±11.1)	118.2±14.2	43.4±1.6	20.1±0.9	N.D.	8.3±0.7	33.6±3.6	N.D.	0.36±0.13	0.6±0.0	98.3
<i>C. butyricum</i>											
1	0.0 (0.0±0.0)	0.0±0.0	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	0.0±0.0	0
1	5.0 (54.6±1.4)	53.8±1.4	25.8±1.0	2.2±0.0	21.5±0.2	N.D.	N.D.	10.6±0.4	0.47±0.02	1.0±0.0	97.4
1	10.0 (107.0±6.6)	74.9±13.3	33.2±0.8	3.2±0.3	32.8±3.04	4.9±0.2	N.D.	12.2±3.2	0.44±0.15	1.3±0.0	96.2
1	20.0 (222.4±6.7)	82.6±9.9	23.7±1.9	4.1±3.1	33.0±15.4	5.4±0.2	N.D.	7.6±5.3	0.28±0.11	0.9±0.0	68.5
1	30.0 (319.1±7.9)	101.0±10.2	17.7±5.8	7.3±5.8	13.3±5.6	4.7±0.1	N.D.	27.6±4.0	0.17±0.17	0.8±0.0	70.9
1	40.0 (425.3±6.7)	108.9±11.4	29.6±8.4	5.9±5.1	42.7±10.3	4.7±0.2	N.D.	4.4±3.0	0.27±0.18	0.8±0.0	56.3
1	45.0 (845.5±8.6)	96.9±18.8	37.7±0.6	12.5±0.5	N.D.	4.6±0.1	N.D.	6.7±0.5	0.38±0.19	0.8±0.0	64.6

glycerol concentrations. In *C. butyricum* cultures, 3-HPA production was not observed.

Bioreactors

In all bioreactors, 1,3-PDO was the main product (Figure 5.1). Although, the OD_{600nm} increased with higher substrate concentration, the 1,3-PDO production yield was highest at 1,8 g/L glycerol with a value of 0.60 ($g_{1,3-PDO}/g_{consumed\ glycerol}$). At 3,7 g/L and 23 g/L glycerol, the 1,3-PDO production yield was 0.32 and 0.20 ($g_{1,3-PDO}/g_{consumed\ glycerol}$), respectively. At 1.8 g/L glycerol, 1,3-PDO production stabilized after five days of fermentation, and for the higher glycerol concentrations after three days. Bioreactors operating in glycerol concentrations of 1,8 and 3,7 g/L showed a complete substrate depletion. In the glycerol concentration of 20 g/L, remaining glycerol was quantified at 1,6 g/L (Figure 5.1). Acetate and formate production was not observed at all. The average acetate/formate ratio in the concentrations was 1.07 (mM/mM), 0.61 (mM/mM) and 0.09 (mM/mM), for the 1.8 g/L, 3.7 g/L and 20 g/L glycerol concentrations respectively. At 20 g/L lactate was additionally observed with an average acetate/ lactate ratio of 0.14 (mM/mM).

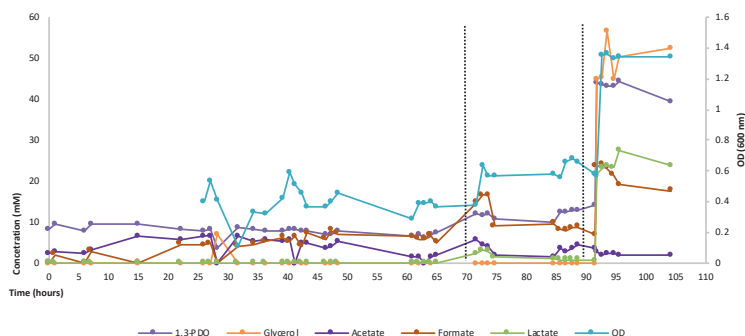


Figure 5.1 | Continuous fermentation of glycerol by *T. pasteurii*. The different glycerol concentrations that were tested (1.8, 3.7 and 20 g/L) are separated with dash bars. Although, the experiment was performed in triplicates, here is presented the bioreactor A2, as the best performing fermenter. Dash lines describe the feeding injections of 3,7 and 20 g/L glycerol.

Proteomics comparison of *T. pasteurii* and *C. butyricum*

For proteomics analyses the optimal substrate concentration of 5 g/L glycerol and 1 g/L yeast extract was used. As previously reported (Chapter 3), the genome of *T. pasteurii* codes for an operon with 17 genes that was correlated with a conversion of glycerol to 1,3-PDO. All the proteins encoded from the operon were in the proteome, except the hypothetical proteins (Tpas_2912) and the glycerol uptake facilitator (Tpas_2927). The most abundant proteins in the operon were glycerol kinase (Tpas_2914), glycerol dehydrogenase (Tpas_2913) and 1,3-PDO dehydrogenase (Tpas_2917). In *T. pasteurii*, the most abundant enzymes were pyruvate kinase (Tpas_354), phosphate acetyltransferase (Tpas_3046) and formate kinase (Tpas_273).

Acetate kinase (WP_027634810.1), dihydroxyacetone kinase subunit DhaK (WP_035764349.1), pyruvate kinase (WP_002582956.1) and glycerol dehydrogenase (WP_027634810.1) were the proteins with the highest abundance in the proteome of *C. butyricum*. Lactate dehydrogenase was not detected in the *C. butyricum* proteome, although the encoding gene is present in the genome. Additionally, the complete set of enzymes involved in the formation of butyrate, was present in the proteome of *C. butyricum* (Table 5.2).

In the reductive pathway, 1,3-PDO dehydrogenase (WP_002581611.1), B12-independent glycerol dehydratase (WP_035765399.1), glycerol dehydrogenase and glycerol dehydratase reactivase (radical SAM domain) (WP_003431290.1) were detected with a high fold change in the proteome of *C. butyricum* (Table 5.2).

Proteomics comparison of 5 g/L and 45 g/L in *T. pasteurii*

Quantitative proteomics for *T. pasteurii* in the conditions of 5 g/L and 45 g/L glycerol indicated different fold change of proteins for glycerol metabolism (Table 5.3). The main enzymes involved in acetate formation were present in the proteome of the two conditions. Abundant in 5 and 45 g/L glycerol were lactate dehydrogenase and pyruvate formate lyase.

B12-dependent glycerol dehydratase activator (Tpas_2921), acetal-

T. pasteurii			C. butyricum		
Description	Log(IBAQ)	Locus tag	Description	Log(IBAQ)	Locus tag
Glycerol metabolism			Glycerol metabolism		
Glycerol kinase	N.D.	Tpas_2911	B12-independent glycerol dehydratase	7.93	WP_035765399.1
Dihydroxyacetone activator	N.D.	Tpas_2912	Glycerol dehydratase reactivase	6.16	WP_003431290.1
Glycerol dehydrogenase	8.98	Tpas_2913	1,3-PDO dehydrogenase	7.79	WP_002581611.1
Dihydroxyacetone kinase	8.35	Tpas_2914	Pyruvate kinase	5.63	WP_002582956.1
Dihydroxyacetone kinase subunit L	8.08	Tpas_2915	Pyruvate kinase	8.20	WP_002582496.1
Phosphotransferase system mannose	N.D.	Tpas_2916	Acetaldehyde dehydrogenase	6.23	WP_003407685.1
1,3-propanediol dehydrogenase	8.73	Tpas_2917	Acetate kinase	8.30	WP_027634810.1
Dihydroxyacetone kinase	8.26	Tpas_2918	Glycerol dehydrogenase	8.30	WP_027634810.1
Dihydroxyacetone kinase subunit L	8.57	Tpas_2919	Phosphate acetyltransferase	5.90	WP_002579602.1
Glycerol dehydratase large subunit	8.23	Tpas_2920	Dihydroxyacetone kinase subunit	8.64	WP_035764349.1
Glycerol dehydratase medium subunit	8.04	Tpas_2921	Dihydroxyacetone kinase subunit	7.79	WP_003413552.1
Glycerol dehydratase small subunit	6.78	Tpas_2922	Dihydrolipoamide dehydrogenase	N.D.	WP_043853069.1
Glycerol dehydratase reactivase alpha subunit	7.72	Tpas_2923	Pyruvate-flavodoxin oxidoreductase	N.D.	WP_027637172.1
Glycerol dehydratase reactivase beta subunit	6.12	Tpas_2924	Glycerol kinase	N.D.	WP_002580667.1
			Butyrate metabolism		
Cobalamin adenosyltransferase	6.66	Tpas_2925	Phosphate butyryltransferase	N.D.	WP_003410761.1
Heme-degrading protein	7.54	Tpas_2926	3-Hydroxybutyryl-CoA dehydrogenase	8.62	WP_002582766.1
Glycerol uptake facilitator	N.D.	Tpas_2927	Crotonase	7.92	WP_002582762.1
Hypothetical protein	N.D.	Tpas_2928	Acetyl-CoA C-acetyltransferase	8.71	WP_002582294.1
Pyruvate kinase	8.42	Tpas_354	Butyrate kinase	7.67	WP_002582660.1
Pyruvate dehydrogenase complex	8.14	Tpas_496	Phosphate butyryltransferase	N.D.	WP_003410761.1
Acetaldehyde dehydrogenase	7.14	Tpas_414	3-Hydroxybutyryl-CoA dehydrogenase	8.62	WP_002582766.1
Lactate dehydrogenase	N.D.	Tpas_1037			
Phosphate acetyltransferase	8.09	Tpas_3046			
Acetate kinase	7.95	Tpas_587			
Formate kinase	8.14	Tpas_273			

Table 5.2 | Protein IBAQ abundances for enzymes involved in glycerol metabolism and the operons correlated to 1,3-PDO for *T. pasteurii* and *C. butyricum*. The proteins in the *T. pasteurii* proteome had log(IBAQ) from 3.96-9.62 and in *C. butyricum* from 3.03-8.71. N.D.: Not Detected

dehyde dehydrogenase and glycosyl transferase group 1 (Tpas_2246) had a significant high fold change in comparison with 45 g/L with 5 g/L glycerol. All subunits of B12-dependent glycerol dehydratase had a high fold change in 45 g/L with the small subunit (Tpas_2922) observed as the most abundant (Table 5.3). The 1,3-PDO dehydrogenase was abundant in both concentrations (Table 5.3). In 45 g/L glycerol, the hypothetical protein (Tpas_2926) was highly abundant and it was identified as a heme-degrading protein (Figure 5.3). Other proteins with high fold and significant change in 45 g/L glycerol were six-hair-pin glycosidase (Tpas_2475), transcription regulator Laci (Tpas_2472), glucose/ribitol dehydrogenase (Tpas_789) and the iron dependent repressor diotheria toxin (Tpas_413) (Table 5.3). At 5 g/L glycerol, enzymes involved in the oxidation and reduction branches were present in a lower abundance than at 45 g/L glycerol (Table 5.3). Proteins with high abundance at 5 g/L were transcription LysR (Tpas_2602), LuxR regulatory protein (Tpas_2189) and membrane protein insertase YidC (Tpas_1747).

DISCUSSION

For the biotechnological production of 1,3-PDO, it is essential to define the optimal growth conditions and metabolic variables. We investigated the effect of the glycerol concentration and yeast extract concentration for *T. pasteurii* and compared it with *C. butyricum*. On a yeast extract of 1 g/L and glycerol concentration of 5 g/L, the 1,3-PDO production yield of *T. pasteurii* is comparable to the *K. pneumoniae* strain Xi-Li, which had a yield of 0,75 ($\text{g}_{1,3\text{-PDO}}/\text{g}_{\text{consumed glycerol}}$) (Table S5.1). Regarding *C. butyricum*, the highest 1,3-PDO production yield in this study was lower than previously reported, which ranged from 0.50 to 0.65 ($\text{g}_{1,3\text{-PDO}}/\text{g}_{\text{consumed glycerol}}$).

Proteomic analysis revealed that the enzymes of the respective gly-

High abundant proteins					
Description	LFQ (Fold Change)	Locus tag	Description	LFQ (Fold Change)	Locus tag
Glycosyl transferases group 1	-2.60	Tpas_2451	Degv	2.34	Tpas_399
Glycerol dehydratase small subunit	-2.57	Tpas_2922	Sodium dicarboxylate symporter	2.30	Tpas_2516
Aldolase reductase	-2.50	Tpas_1637	Saccharopine dehydrogenase	2.13	Tpas_1242
Six-hairpin glycosidase	-2.47	Tpas_2475	LuxR bacterial regulatory protein	1.90	Tpas_2189
ABC transporter permease protein domain	-2.42	Tpas_2171	ABC transporter	1.43	Tpas_501
tRNA guanine-N(7)-methyltransferase	-2.41	Tpas_428	Transcription LysR	1.24	Tpas_2602
ABC transporter	-2.39	Tpas_592	ATPase AAA-type	0.82	Tpas_624
ABC transporter permease protein	-2.37	Tpas_678	Peptide chain release factor 1	0.79	Tpas_2687
Aminotransferases class IV signature	-2.31	Tpas_1231	UvrABC system protein B	0.74	Tpas_950
ABC transporter permease protein	-2.16	Tpas_1868	Exodeoxyribonuclease 7 large subunit	0.66	Tpas_219
Aminotransferases class IV signature	-2.14	Tpas_1846	Cysteine-tRNA ligase	0.64	Tpas_583
Heme-degrading protein	-2.13	Tpas_2926	Coenzyme A biosynthesis bifunctional	0.63	Tpas_186
Pyruvate formate-lyase-activating enzyme	-1.99	Tpas_272	Threonylcarbamoyltransferase	0.60	Tpas_2660
Histidine biosynthesis bifunctional protein	-1.98	Tpas_3030	Cell shape-determining protein MreC	0.60	Tpas_597
Iron dependent repressor diptheria toxin	-1.92	Tpas_413	Transcriptional regulatory protein C terminal	0.60	Tpas_344
Xanthine phosphoribosyltransferase	-1.90	Tpas_1343	Phosphotransferase system EIIb component	0.59	Tpas_3038
Hypothetical protein	-1.85	Tpas_704	Ribosomal protein S18	0.59	Tpas_1737
ABC transporter	-1.84	Tpas_1112	Membrane protein insertase YidC	0.59	Tpas_1747
Glycerol dehydratase medium subunit	-1.82	Tpas_2921	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	0.56	Tpas_1441
3-isopropylmalate dehydrogenase	-1.80	Tpas_1239	ATP phosphoribosyltransferase	0.55	Tpas_988
Transcription regulator LaCl	-1.80	Tpas_2472	Fructose-1,6-bisphosphatase class 3	0.53	Tpas_463
Glucose/ribitol dehydrogenase	-1.63	Tpas_789	TGS protein	0.50	Tpas_2316
NUDIX hydrolase	-1.42	Tpas_617			
50S Ribosomal protein L19	-0.90	Tpas_1059			
Aldolase-type Tim barrel	-0.57	Tpas_78			
Acetaldehyde dehydrogenase	-0.56	Tpas_414			
Peptidase M42	-0.56	Tpas_2359			

Table 5.3 | Proteins LFQ fold changes in the conditions of 45 g/L and 5 g/L in *T. pasteurii*. All proteins listed in the high abundant section of the table have p-value lower than 0.05 and fold change higher than 0.50. Values with - have higher LFQ fold change in 45 g/L glycerol. N.D.: Not Detected

Glycerol metabolism in <i>T. pasteurii</i>			
Description	LFQ (Fold Change)	Locus tag	p-value
Glycerol kinase	0.37	Tpas_2911	0.11
Dihydroxyacetone activator	N.D.	Tpas_2912	N.D.
Glycerol dehydrogenase	-0.25	Tpas_2913	0.09
Dihydroxyacetone kinase	-0.09	Tpas_2914	0.15
Dihydroxyacetone kinase subunit L	0.07	Tpas_2915	0.43
Phosphotransferase system mannose	N.D.	Tpas_2916	N.D.
1,3-propanediol dehydrogenase	-0.03	Tpas_2917	0.78
Dihydroxyacetone kinase	0.06	Tpas_2918	0.52
Dihydroxyacetone kinase subunit L	-0.19	Tpas_2919	0.27
Glycerol dehydratase large subunit	-0.19	Tpas_2920	0.10
Glycerol dehydratase medium subunit	-0.13	Tpas_2921	0.53
Glycerol dehydratase small subunit	-2.57	Tpas_2922	0.00
Glycerol dehydratase reactivase alpha subunit	-0.14	Tpas_2923	0.27
Glycerol dehydratase reactivase beta subunit	-1.80	Tpas_2924	0.01
Cobalamin adenosyltransferase	-0.42	Tpas_2925	0.50
Heme-degrading protein	-2.13	Tpas_2926	0.12
Glycerol uptake facilitator	N.D.	Tpas_2927	N.D.
Hypothetical protein	N.D.	Tpas_2927	N.D.
Pyruvate kinase	-0.19	Tpas_354	0.03
Pyruvate ferredoxin	-0.26	Tpas_169	0.08
Dihydrolipoic transacetylase	0.08	Tpas_496	0.37
Acetaldehyde dehydrogenase	-0.55	Tpas_414	0.01
Lactate dehydrogenase	-0.04	Tpas_1037	0.86
Phosphate acetyltransferase	-0.15	Tpas_3046	0.10
Acetate kinase	-0.21	Tpas_587	0.19
Pyruvate formate-lyase	-0.15	Tpas_273	0.17
Pyruvate formate-lyase	-0.06	Tpas_2137	0.56

erol fermentation pathways were present in the proteome of *T. pasteurii* and *C. butyricum*. When compared at 5 g/L glycerol, *T. pasteurii* produces lactate and lactate dehydrogenase was present in the proteome, while *C. butyricum* produces butyrate and enzymes of the butyrate pathway were detected in the proteome. The main difference was in the reductive pathway. In *T. pasteurii*, enzymes from the large operon were abundant in the proteome, while in *C. butyricum* only three enzymes were detected. The proteome abundance of the previously described operon of *T. pasteurii* indicates its impact in 1,3-PDO production. In contrast, only the B12-independent glycerol dehydratase, the glycerol dehydratase activator (radical SAM) and 1,3-PDO dehydrogenase were abundant in *C. butyricum*.

In high glycerol concentrations, 1,3-PDO production in *T. pasteurii* was significantly lower. In the continuous fermenters, the production yield was also low at higher concentrations although, stable 1,3-PDO production was observed. The low production yield of metabolites at high glycerol concentrations in batch and fermenters cultures is most likely due to 3-HPA accumulation, which is toxic and causes cell death²⁹. The 3-HPA production was reported for *Lactobacillus reuterii* and other *Lactobacillus* bacteria¹⁹⁹⁻²⁰², but also for the species: *C. butyricum* VPI 3266²⁰³, *K. pneumoniae* ATCC 25955²⁰⁴, *Ct. freundii* ATCC 8090²⁰⁴, *Enterobacter agglomerans* CNCM1210²⁰⁴.

Considering the detection of 3-HPA in high concentrations, the high fold change of B12-dependent glycerol dehydratase could be correlated with the accumulation of 3-HPA. The fact that 1,3-PDO dehydrogenase is not further induced in 45 g/L indicates a reduced metabolic capability for the biotransformation of 3-HPA to 1,3-PDO. Enzymes responsible for the required co-factor regulation such as cobalamin-adenosyl transferase and heme-degrading protein had high fold change in 45 g/L. Glycerol dehydrogenase had a high fold change in 45 g/L and it is responsible for the first step of the oxidative pathway, which generates NADH. 1,3-PDO dehydrogenase abundance remained stable and it was possibly influenced by the toxicity of 3-HPA. While 3-HPA formation is hampering the conversion of glycerol to 1,3-PDO, its production may be biotechnologically applied too. Currently, 3-HPA is being used as a food preservative¹⁹⁹. Alternatively, it can be

converted to the valuable compound of 3-hydroxypropionic acid²⁰⁵.

Generally, in *T. pasteurii*, proteins had a higher fold change at 45 g/L glycerol compared to 5 g/L glycerol. Interestingly at 45 g/L glycerol, the glucose/ribitol dehydrogenase was abundant, which has a similar function to glucose dehydrogenase²⁰⁶. We observed a six-hairpin glycosidase to have one of the highest fold changes 45 g/L. This is a glycosyltransferase, which is involved in the biosynthesis of oligosaccharides and polysaccharides contributing to the bacterial biomass²⁰⁷. At 45 g/L glycerol, the iron dependent repressor dietheria toxin had a high fold change and is correlated to the iron regulation²⁰⁸. At 5 g/L glycerol transcription LysR, which was previously connected with glycerol metabolism²⁶, was abundant. Additionally, LuxR type regulatory protein²⁰⁷ was abundant. This protein activates the Lux operon. Finally, at 5 g/L glycerol, we observed the abundance of the membrane protein insertase YidC²⁰⁹, which integrates with other proteins in the membrane.

In conclusion, this proteomic and physiological study identified that *T. pasteurii* can produce 1,3-PDO at high yields. At a high glycerol concentration, 3-HPA is produced, which may be linked with the difference in proteomics abundance of B12-dependent glycerol dehydratase and 1,3-PDO dehydrogenase. In *T. pasteurii*, all the genes in the operon were present in the proteome, which signifies the importance of the accessory proteins. This study provides insights into the metabolism of 1,3-PDO in *T. pasteurii* and introduces the species as a possible efficient producer for biotechnology.

ACKNOWLEDGEMENTS

This research was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science.

SUPPLEMENTARY DATA

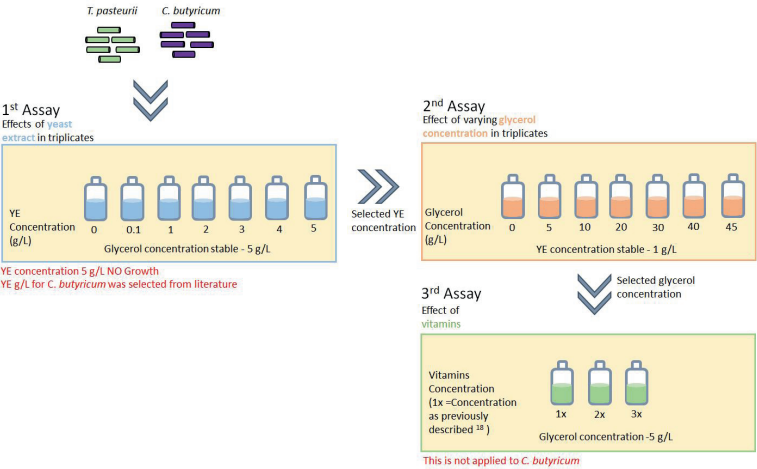


Figure S5.1 | Experimental set-up and identification of variables analysed in this study. For every phase the most optimal concentration of the variable is selected. YE: Yeast Extract

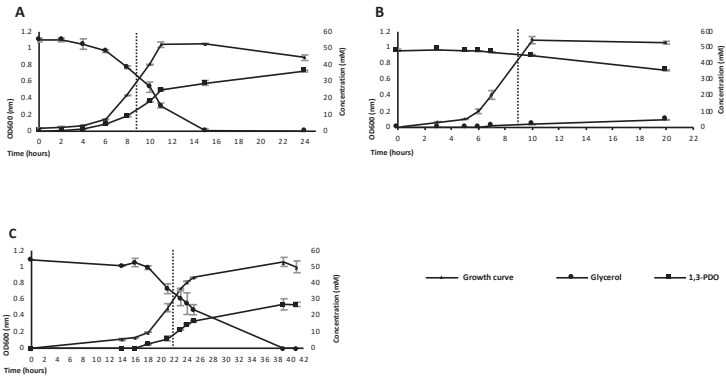


Figure S5.2 | Protein extraction time points for the condition of *T. pasteurii* 5 g/L (A) and 45 g/L (B) and *C. butyricum* 5 g/L (C). Bars indicate time point for samples extraction.

Table S5.1 | Microbial production of 1,3-PDO in previous studies indicating the concentrations of yeast extract and glycerol used. -: No Data

Biosafety Level	Fermentation Type	Yeast Extract (g/L)	Glycerol Concentration (g/L)	1,3-PDO Concentration (g/L)	Yield 1,3-PDO/ Glycerol (mol/mol)	Reference
<i>Klebsiella pneumoniae</i>						
DSM 2026						
2	Shake flasks	1.5	20	9	0.62	Mu et al., 2006162
2	Shake flasks	1.5	20	8	0.54	Mu et al., 2006162
2	Shake flasks	1.5	20	7	0.47	Mu et al., 2006162
2	Fed-batch bioreactor	1.5	40	61	0.49	Mu et al., 2006162
2	Fed-batch bioreactor	1.5	40	51	0.46	Mu et al., 2006162
2	Fed-batch bioreactor	1.5	40	53	0.47	Mu et al., 2006162
Xi-Li						
-	Fed-batch bioreactor	-	20	38	0.70	Zhang et al., 200756
-	Batch bioreactor	-	20	12	0.75	Zhang et al., 200756
ATCC 25955						
2	Batch culture	2	20	-	0.65	Barbirato et al., 199857
2	Batch culture	2	70	-	0.55	Barbirato et al., 199857
M5a1						
2	Fed batch bioreactor	1.5	15-40	51	0.52	Cheng et al., 2004111
2	Fed batch bioreactor	1.5	15-40	58	0.53	Cheng et al., 2004111
<i>Citrobacter freundii</i>						
ATCC 8090						
2	Batch culture	2	20	-	0.65	Barbirato et al., 199857
2	Batch culture	2	70	-	0.54	Barbirato et al., 199857
FMCC-B294 (VK-19)						
-	Batch bioreactor	2.5	20	11	0.65	Metsoviti et al., 2013210
-	Batch bioreactor	2.5	20	10	0.64	Metsoviti et al., 2013210
-	Batch bioreactor	2.5	20	10	0.64	Metsoviti et al., 2013210
-	Batch bioreactor	1	20	10	0.64	Metsoviti et al., 2013210

<i>Clostridium pasteurianum</i>						
ATCC 6013						
1	Batch cultures	1	5	-	0.50	Taconi et al., 200955
1	Batch cultures	1	10	-	0.33	Taconi et al., 200955
1	Fed-batch bioreactor	1	29	-	0.26	Biebl et al., 2001211
<i>Clostridium butyricum</i>						
CNCM 1211						
-	Batch culture	2	20	-	0.62	Barbirato et al., 199857
-	Batch culture	2	70	-	0.64	Barbirato et al., 199857
VPI 3266						
2	Batch culture	-	39	-	0.58	Gonzalez-Pajuelo et al., 2004160
2	Batch culture	-	44	-	0.51	Gonzalez-Pajuelo et al., 2004160
2	Batch culture	-	33	-	0.56	Gonzalez-Pajuelo et al., 2004160
2	Batch culture	-	65	35	0.65	Saint-Amans et al., 1994212
2	Fed-batch culture	-	25	65	0.68	Saint-Amans et al., 1994212
2	Continuous bioreactor	-	58	29	0.62	Gonzalez-Pajuelo et al., 2004160
2	Continuous bioreactor	-	60	30	0.60	Gonzalez-Pajuelo et al., 2004160
2	Continuous bioreactor	-	30	15	0.62	Gonzalez-Pajuelo et al., 2004160
VPI 1718						
-	Batch bioreactor	1	20	11	0.68	Chatzifragkou et al., 2011110
-	Batch bioreactor	1	20	11	0.68	Chatzifragkou et al., 2011110
F2b						
-	Batch bioreactor	1	90	47	0.63	Papanikolaou et al., 2004213
-	Continuous bioreactor	1	30	16	0.66	Papanikolaou et al., 2004213
-	Continuous bioreactor	1	60	37	0.64	Papanikolaou et al., 2004213

CHAPTER 6

PROTEOMICS OF PSYCHROTOLERANT *TRICHOCOCCUS PATAGONIENSIS* AND CHARACTERISATION OF THE PRODUCED EPS AS INULIN

Nikolaos Strepis^{1,2}, Laura de Nies¹, Yuemei Lin³, Sjef Boeren⁴, Steven Aalvink¹, Peter J Schaap², Michel Konnen⁵, Jaap Sinninghe Damsté⁶, Alfons JM Stams^{1,6}, Diana Z Sousa¹

1. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

2. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands

3. Faculty of Applied Sciences, TU Delft, Delft, The Netherlands

4. Laboratory of Biochemistry, Wageningen University & Research, Wageningen, The Netherlands

5. NWO-NIOZ Royal Netherlands Institute for Sea Research, 't Horntje, The Netherlands

6. Centre of Biological Engineering, University of Minho, Braga, Portugal

ABSTRACT

Trichococcus patagoniensis is an anaerobic psychrotolerant bacterium. The ability to grow at sub-zero temperatures is due to cold adaptation mechanisms and the secretion of extracellular polymeric substances (EPS). To gain further insight into cold adaptation, we performed physiological trials and proteomics on *T. patagoniensis* grown at 30 °C and 0 °C. Proteomics showed that at 0 °C, the cell metabolism aims for survival by improving the translation mechanisms with the generation of multiple DEAD-box helicases and ribosomal proteins. At 30 °C, metabolism shifts towards proliferation with enzymes for DNA repair and synthesis dominating the proteome. The produced EPS was identified as inulin, a prebiotic. Subsequently, the pathway involved in inulin production by *T. patagoniensis* was characterised. Inulin is possibly linked to the protein EpsA, which belongs to the EPS assembly proteins epsABCDE. Inulin is produced by a limited number of species and currently *T. patagoniensis* is the only known psychrotolerant species able to produce inulin, which highlights its biotechnological potential.

INTRODUCTION

Cold environments (below 5 °C) are widespread on Earth²¹⁵ and host well-adapted microbial communities which are able to survive and thrive at low temperatures²¹⁶. Psychrophilic microbes show optimal growth at temperatures lower than 15 °C. Psychrotolerant microorganisms can survive at low temperatures but have optimal growth at mesophilic temperatures^{58,215}. So far, most psychrophilic/psychrotolerant bacteria belong to the genera of *Clostridium*, *Cryobacterium*, *Psychrobacter*, *Psychromonas*, and *Bacillus*²¹⁷.

In cold environments microorganisms encounter stress conditions such as freezing of the cell wall, loss of membrane fluidity, reduction of metabolic activity and lack of nutrients⁵⁸. Consequently, microbes have evolved adaptable structural and metabolic mechanisms to cope with such harsh conditions. Cold shock proteins belong in the adaptation arsenal of microbial systems and are often conserved among psychrophilic and psychrotolerant species²¹⁸. Specific metabolites such as extracellular polymeric substances (EPS) have antifreeze properties that prevent water crystallisation around the bacterial cell⁶⁶. Species such as *Winogradskyella* sp., *Pseudoalteromonas* sp. MER144 and *Colwellia psychrerythraea* 34H can produce EPS as cryoprotectant with a varying monomer composition (mannose, glucose, galactose, arabinose and glucuronic acid)^{65,219,220}.

The psychrotolerant *Trichococcus patagoniensis* was previously reported to produce EPS that covers the cells, when grown at a low temperature⁴⁹. Although *T. patagoniensis* has an optimal growth at 30 °C, it can still grow at -5 °C. This capability of *T. patagoniensis* to grow in sub-zero temperatures may be reflected in the proteome, membrane lipid composition and EPS production. This study aimed to gain insights into the cold adaptation mechanisms of *T. patagoniensis*. We analysed the proteome and membrane lipids of cells grown at 0 and 30 °C and the EPS that is formed.

MATERIALS AND METHODS

Media preparation for bacterial growth

Trichococcus patagoniensis DSM 18806^T was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). Growth experiments were set up in serum bottles containing a mineral medium, which was prepared as previously described⁸⁰. Bottles were sealed with rubber stoppers and aluminium caps seals and the headspace was exchanged by N₂/CO₂ (80/20 (v/v) to a final pressure of 1.5 atm. The inoculation medium was supplemented with 1 g/L of yeast extract and 20 mM of glucose (added from sterile solutions). Growth experiments were carried out at 30 °C and 0 °C; incubations at 30 °C were performed in a lab incubator (Hettich Benelux, Geldermalsen, the Netherlands) and incubations at 0 °C were performed in a water/ethanol (75/25 %) Huber CC 410 bath (Huber; Offenburg, Germany). Cells grown at both temperatures were collected for proteomics, exopolymer and membrane lipids analysis.

Analytical measurements

Growth was monitored by measuring the optical density (OD) of the cultures at 600 nm wavelength with a spectrometer (Hitachi U-1500; Labstuff, Capelle aan den IJssel, the Netherlands). Soluble compounds were quantified by high performance liquid chromatography in Thermo Electron HPLC system using Agilent Metacarb 67H column (300x6.5 mm) (ThermoFisherScientific, Waltham, MA,) and a refractive index detector, which had as a mobile phase sulfuric acid (5 mM) at a flow rate of 0.8 ml min⁻¹ and a temperature of 45 °C.

For scanning electron microscopy analysis, cells were fixed in 2 % glutaraldehyde in a 0.1 M phosphate buffer (pH 7.4) for 2 h at room temperature, and post-fixed with 2 % osmium tetroxide for 30–60 min. Later, cells were dehydrated in a graded alcohol series (50, 70, 96 and 100 %), treated with hexamethyldisilazane, and mounted into aluminium stubs and coated with platinum. Finally, samples were implemented on a FEI Quanta 250 FEG-scanning electron microscope.

Biomass extraction for EPS and lipid analysis

The experiment was performed in duplicate of 125 ml cultures grown at 30 °C and 0 °C and in a cold shock condition from 30 °C to 0 °C. For the cold shock, cultures were first grown at 30 °C and after 14 hours of incubation the bottles were transferred to the 0 °C incubator, where incubation was continued for 168 hours. After centrifugation, the cell biomass was washed in phosphate buffered saline (PBS) and then cells were freeze dried overnight with an Eppendorf concentrator plus (Eppendorf, Hamburg, Germany).

EPS identification

Lyophilised cells (0.1 g) were extracted in 50 ml demineralised water at 70 °C for 30 min. After centrifugation, the supernatant was lyophilised. The Fourier transform infrared (FTIR) spectrum of the lyophilised extracellular polymeric substances (EPS) and inulin standard (Inulin from Chicory, Sigma-Aldrich, St. Louis, MO) were recorded on a FTIR spectrometer (Perkin Elmer, Waltham, MA) equipped with a DTGS Mid-infrared detector and a Golden Gate single reflection diamond attenuated total reflectance (ATR) cell. Spectra were recorded from 600 cm^{-1} to 4,000 cm^{-1} at room temperature. The absorption of the samples and background were measured as 64 scans each. Spectral analyses were performed with PeakFit® (Systat Software Inc., San Jose, CA), to calculate 1,050/1,540 cm^{-1} peak area ratio.

Lipids identification

Lyophilised cells (50 μg) were hydrolysed with 1.5 N HCl in methanol by refluxing for 3 hours. The hydrolysate was adjusted to pH 5 with 2 N KOH/MeOH (1:1, v/v) and, after the addition of water to a final 1:1 ratio of H_2O /MeOH, extracted three times with dichloromethane (DCM). The DCM fractions were collected and dried over sodium sulfate. The obtained extract was methylated with diazomethane and separated over an activated Al_2O_3 column into an apolar and a polar fraction using DCM and DCM/MeOH (1:1, v/v) respectively, as eluent. The apolar fraction (containing the fatty acid methyl esters – FAMES) was analysed

by gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS). GC was performed using an Agilent Technologies 7890B instrument (Agilent Technologies, Santa Clara, CA) equipped with an on-column injector and a flame ionisation detector. A fused silica capillary column (25 m x 0.32 mm) coated with CP Sil-5 CB (d_f 0.12 μ m) was used with helium as carrier gas. The samples were injected at 70 °C and the oven temperature was programmed to 130 °C at 20 °C/min and then at 4 °C/min to 320 °C, at which it was held for 10 min. GC-MS was performed on an Agilent Technologies 7890B gas chromatograph interfaced with Agilent Technologies MSD 5975 mass spectrometer operated at 70 eV with a mass range of m/z 50-600. The gas chromatograph was equipped with a fused silica capillary column as described for GC. The carrier gas was helium. The same temperature program as for GC was used. Double bond positions of unsaturated FAMES were determined on the basis of the mass spectra of their dimethyl disulfide derivatives as previously described²²¹.

Intact polar lipids were extracted from the lyophilised cells using a modified Bligh and Dyer technique as previously described²²². An aliquot of the obtained extract was dissolved in hexane/2-propanol/water (72:27:1), filtered through a 0.45 μ m regenerated cellulose filter, and analysed by HPLC/ESI-MS. Intact polar lipids (IPLs; i.e. glycerol ester membrane lipids with attached polar head groups) were analysed according to a published method with some modifications²²³. An Agilent 1200 series LC was used equipped with thermostatted auto-injector and column oven, coupled to a Thermo LTQ XL linear ion trap with Ion Max source with electrospray ionisation (ESI) probe (Thermo Scientific, Waltham, MA) using conditions previously described²²⁴. The lipid extract was analysed by positive ion scanning (m/z 400-2000), followed by data dependent MS/MS, where the four most abundant masses in the mass spectrum were fragmented successively. Each MS/MS was followed by data dependent MS/MS, where the base peak of the MS/MS spectrum was fragmented under identical fragmentation conditions.

Proteomics sample preparation and analysis

For the two conditions, four replicates were prepared of 250 mL cell suspensions. As the defined growth point was considered the glucose consumption of about 17.5 mM., cells were harvested by centrifugation at 14,700 xg for 20 min. The centrifugation pellet was collected and resuspended in a SDS-lysis buffer (100 mM Tris/HCl pH 7.6+4 % SDS, vol/vol+0.1 M dithiothreitol). The cells were lysed using French pressure cell (Thermo IEC; ThermoFisherScientific, MA) with four runs per each sample under 1,900 psi pressure. A centrifugation step was conducted to collect the pellet proteins.

The proteome quantification of *T. patagoniensis* cells that had been grown in two temperature conditions were performed using nanoLC-MS/MS. For every sample, 50 ug of protein was loaded in SDS-PAGE on a 10-well SDS-PAGE 10% (wt/vol) Bis-Tris Gel (Mini Protean System, Bio-Rad, San Diego, CA), for 55 min at a constant voltage of 120V using Tris-SDS as a running buffer. Staining of the gels was performed with Colloidal Blue Staining Kit (Life Technologies, Carlsbad, CA) and treated for reduction and alkylation using 10 mM dithiothreitol and 15 mM iodoacetamide in 50 mM ammonium bicarbonate. The sample lanes on the gel were sliced into 4 four pieces and then cut again into small slices of approximately 1-2 mm. Peptide digestion was activated with 50 uL of sequencing grade trypsin (5 ng/μL in 50 mM ammonium bicarbonate) and incubated at room temperature overnight while shaking. After enzymatic digestion, for the extraction of the peptides the samples were acidified to a Ph of 2-4 by adding 10 % TFA. After this, the peptides were extracted using the uColumn clean-up procedure with the C18 column. The digested peptides were desalted and were LC-MS/MS ready. The nanoLC-MS/MS used a Proxeon Easy nanoLC and an LTQ-Orbitrap XL instrument (Thermo Fisher Scientific, Naarden, the Netherlands) as previously reported⁴¹.

The quantified peptide spectra were analysed with Maxquant¹⁹⁷ using the "Specific Trypsin/P" Digestion mode with a maximum two missed cleavages and regarding the Andromeda search engine the default parameters were used. The protein database was based on the translated coding sequences of *T. patagoniensis* genes predicted with by Prodigal 2.6.2011⁸⁹. Proteomic analysis was conducted following the

MaxQuant pipeline and the statistical analysis was done by Perseus¹⁹⁷. The FDR for matching peptides with proteins were defined at 0,01 and p-value at 0,05. Additional filtering steps were the criteria that every protein must have a minimum of two peptides, of which at least one should be unique and at least one should be unmodified.

Genomic analysis

The genome as well the annotation of *T. patagoniensis* is available in the public genome repository ENA (EMBL, Germany)¹⁰⁸. Functional analysis was conducted using protein functional domains based on Pfam families¹⁹ and InterProScan⁹⁰. Prediction of transmembrane regions was performed with TMHMM¹¹⁹ and SignalIP¹⁹⁸. The Kyoto Encyclopedia of Genes and Genomes (KEGG)²²⁵ was used for studying the identifying proteins in metabolic pathways and statistical graphs were generated in R¹²⁰.

RESULTS

Growth results

At the optimal temperature of 30 °C, *T. patagoniensis* reached a stationary phase within 24 hours (Figure S6.2). At 0 °C, however, growth was only observed after 400 hours and a stationary phase was achieved at 750 hours. Adapted cells in 0 °C had a shorter lag phase of 336 hours.

Identification of EPS

The spectral region between 1,200 and 900 cm⁻¹ is dominated by a sequence of intensive peaks due to functional groups of oligo- and polysaccharides in FTIR spectra (Figure 6.1). The typical peaks of inulin (fructose polymer, C₆H₁₀O₅)_n are a broad intensive peak with a maximum at around 1,050 cm⁻¹ and two shoulders at 980 and 1,130 cm⁻¹. Apparently, all the extracted EPS had the same typical peaks as the in-

ulin standard²²⁶, which implies that the microorganism produces inulin polymer. In addition to the typical inulin peaks, there are amide I and II peaks at $1,640\text{ cm}^{-1}$ and $1,540\text{ cm}^{-1}$, respectively, in the spectra of the EPS. Those two amide peaks indicate that there are polypeptides/proteins in the EPS complexed with inulin-like polymers.

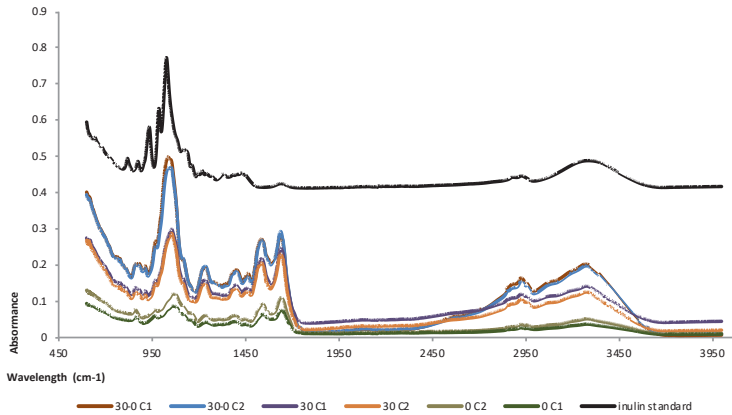


Figure 6.1 | Infrared spectrum for the temperature conditions of $30\text{ }^{\circ}\text{C}$, $0\text{ }^{\circ}\text{C}$ and the cold shock from $30\text{ }^{\circ}\text{C}$ to $0\text{ }^{\circ}\text{C}$. The peaks in specific wavelength indicate that the polysaccharide is inulin ($\text{C}_6\text{H}_{10}\text{O}_5$). The absolute peak area of FTIR peak is referring to the relative value. The blue line is the inulin standard and the abundance were increase by 0.4 for clearer indication of the peak's match with the type of the polysaccharide.

The signal intensity of the inulin EPS spectrum at $0\text{ }^{\circ}\text{C}$ is different from that at $30\text{ }^{\circ}\text{C}$ (Figure 6.1). To compare the relative amount of inulin polymers produced at these temperatures, $1,050/1,540\text{ cm}^{-1}$ peak area ratio was obtained. The ratios are 3.6 at $0\text{ }^{\circ}\text{C}$, 3.1 at $30\text{ }^{\circ}\text{C}$ and 4.6 when the temperature changed from $30\text{ }^{\circ}\text{C}$ to $0\text{ }^{\circ}\text{C}$.

Morphologically, inulin-EPS surrounded the cell and was shared among dividing cells (Figure 6.2). Inulin-EPS was observed to extend out of the cell around the membrane. Characteristic is the inulin-EPS "belt" that is created around the cells. This formation of inulin-EPS was observed in both conditions.

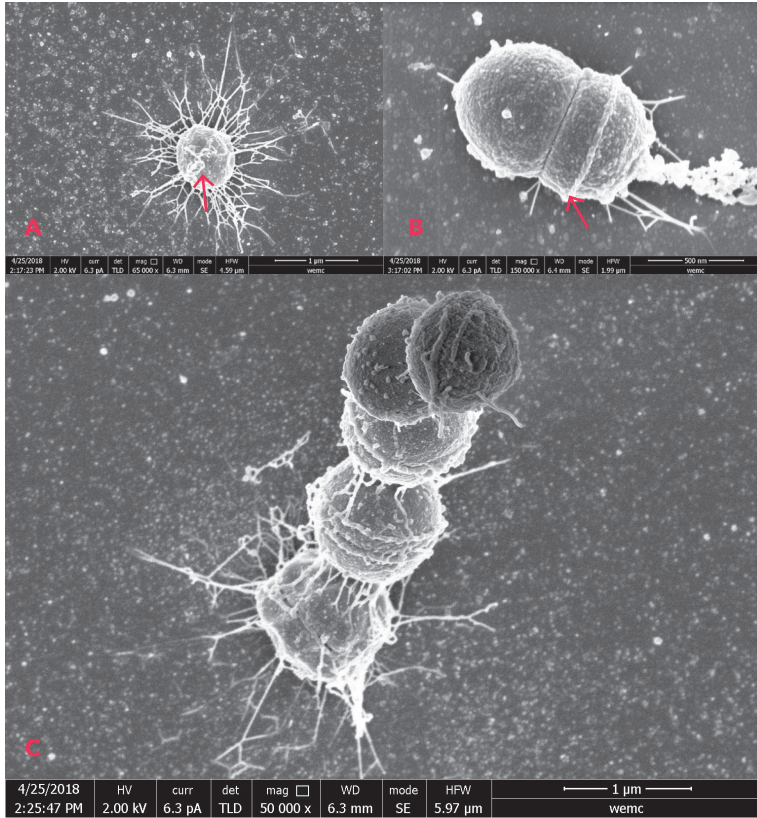


Figure 6.2 | Inulin-EPS production and morphology of cells as observed from scanning electron microscopy pictures for *T. patagoniensis* at 30 °C (A, C) and at 0 °C (B). Red arrow indicates the formation of inulin-EPS around the cell.

Genomic analysis for EPS synthesis and cold adaptation mechanisms

In silico analysis of *T. patagoniensis* revealed the existence of all genes involved in the Wzy-dependent pathway, involved in EPS formation (Figure 6.3). The Wzy-dependent pathway uses glucose as input for producing repeat-units that compose EPS²²⁷. Multiple homologues of beta galactosidases exist in the genome for converting lactose to glucose. All genes of the enzymatic cascade are present for converting glucose to fructose-biphosphate and later to five different repeat units (Figure 6.3).

The Wzy-dependent pathway is complemented with genes for assembling EPS. In the genome of *T. patagoniensis*, *epsABCDE*, as well as two homologs for *wzx* flippase and *wzy* polymerase, were identified (Figure 6.3). Furthermore, we identified multiple genes encoding membrane-bound EPS assembly genes. In *T. patagoniensis*, the *epsC* encoded enzyme was predicted to have two transmembrane regions and a cytoplasmic C-terminal domain required for kinase activation which matches with the previously studied protein structure²²⁷. A similar protein structure was also conserved in the *epsD* gene product. The enzymes *wzx* flippase and *wzy* polymerase are large proteins that are known to be firmly bound to the membrane, and in *T. patagoniensis* were predicted *in silico* to have 14 transmembrane regions.

Regarding cold tolerance mechanisms, five homologs of cold-shock protein A (CspA) were predicted in *T. patagoniensis* (Table S6.2). One of the CspA genes (Tpas_1901) was identified in an operon (Tpat_1889-1909) containing genes encoding for tRNA synthetases, DNA-repair protein, rod shape determining proteins and ribosomal proteins (Table S6.2). In addition, two GyrA genes were predicted in the genome, which can be activated by CspA enzymes. Finally, transcription termination factor Rho (Tpat_2557) was detected in the genome, which is a large cold-shock protein.

Proteomics analysis

In the proteome, all enzymes participating in the Wzy-dependent pathway were present in both conditions. Proteins with a high fold change at 0 °C were Pgi, Fbp and ManB, which are involved in Fru-6-P production (Figure 6.3). Glk and LacZ, which convert lactose to Glc-6-P, had significantly higher fold change at 30 °C. Regarding the fold change of EpsABCDE unit proteins, EpsA had higher fold change at 0 °C, while EpsC and *epsD* did not. Transmembrane proteins NagE, Wzx flippase, Wzy polymerase, EpsB and EpsE could not be identified in the soluble protein fraction.

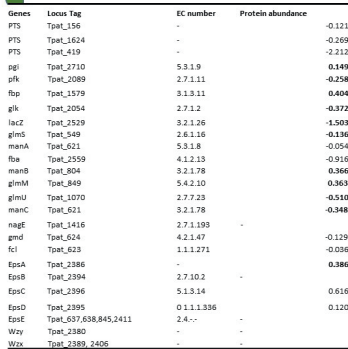


Figure 6.3 | EPS production from the Wzy-dependent pathway in *T. patagoniensis*. The pathway utilizes glucose or lactose and with a cascade of enzymes generates the repeat units UDP-GlcNac, UDP-ManNac, UDP-GalNac, GDP-Man and GDP-Fuc. Proteomics indicated that all enzymes are active in the two conditions of 0 °C and 30 °C. There are individual genes that have significantly high fold change in specific conditions (Bold colour). The EPS genes are located in the transmembrane, the protein unit of EpsABCDE, Wzx flipase and Wzy polymerase. Due to the constant interaction of Eps genes with the EPS molecule, it is expected that EPS is attached with EPS genes.

The CspA-like cold-shock proteins were not detected in the proteome. However, the two GyrA enzymes had a high fold change at 0 °C. Transcription termination factor Rho has a high fold change at 0 °C as well. A thiol reductase thioredoxin gene is neighbour to the transcription factor Rho and the encoded protein was identified only at 0 °C. At 0 °C, proteins related to transcription and translation mechanisms had high fold change. In addition, DEAD-box proteins had a high fold change, DEAD-box helicase rhIE, cshB, yxiN, fml, cshA. Fifty-nine ribosomal proteins had high fold change at 0 °C, with the highest fold change observed in the ribonuclease HIII, ribosomal protein S14, ribosome-binding factor, ribosomal protein L21 and small ribosomal biogenesis GTPase (Figure 6.4).

At 30 °C, two heat shock proteins had a high fold change; heat shock protein (Tpat_1085) and heat shock protein 70 family (Tpat_2139). Furthermore, there are fewer ribosomal proteins with a high fold change at 30 °C (39 ribosomal proteins). We detected 27 DNA targeting proteins at 30 °C versus 18 at 0 °C (which had a low fold change) (Table S6.1). A high fold change was observed in DNA ligase, HNH nuclease, four ribose-phosphate diphosphokinase proteins, SAICAR synthetase and the 10 kDA chaperon.

Membrane lipid composition

The composition of intact polar lipids in the membrane cell wall at the temperatures 0 °C and 30 °C were phosphatidylaminopentananetetrol (PG), phosphatidylglycerol, dihexose, cardiolipin and hexose. There were minor differences between the two conditions (Table S6.2). In detail, at 0 °C dihexose is more abundant in the membrane compared to 30 °C, where PG was higher. There are no significant differences in the poly unsaturated/saturated ratio of lipids between the 0 °C and 30 °C condition (Figure S6.2).

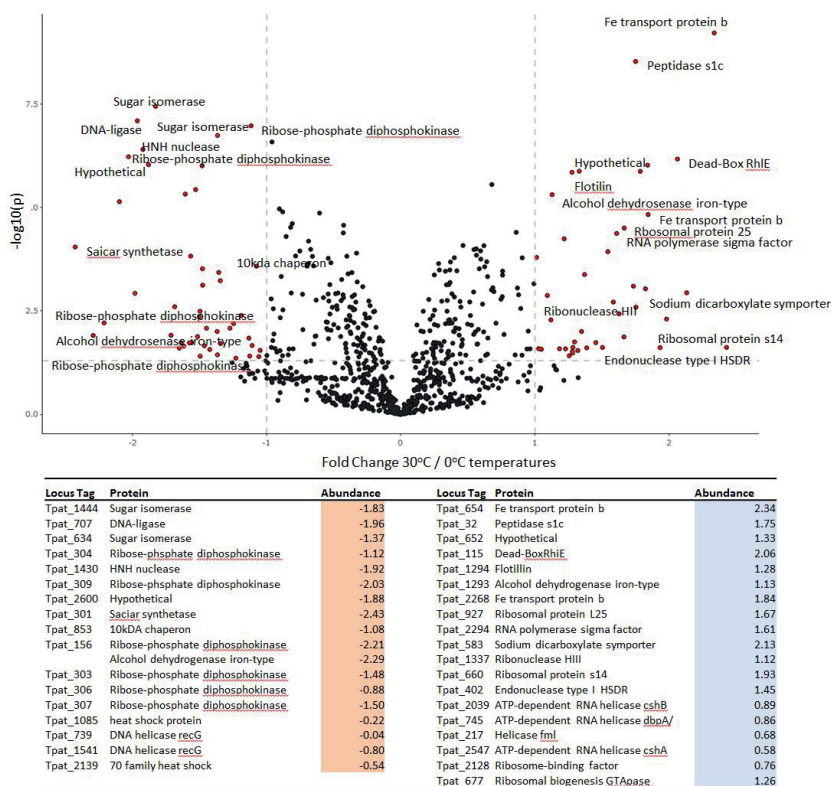


Figure 6.4 | Volcano plot showing highly abundant (fold change >1) and significantly different (p-value < 0.05) proteins from *T. patagoniensis* grown at 0 °C and 30 °C. The positive fold change is for proteins abundant in 0 °C and the negative change is for proteins abundant at 30 °C. In 0 °C the most abundant proteins are related to transcription and translation and at 30 °C the most significant abundant proteins are related in DNA synthesis and polymerization. The proteins without of label were either sugar related or ABC transporters.

DISCUSSION

The EPS excreted by *T. patagoniensis* is characterized as inulin. Inulin is water-soluble, thermally stable and inhibits water crystallization at low temperatures²²⁸. Inulin production in bacteria is very uncommon. Species producing inulin or inulin oligosaccharides are *Lactobacillus gasseri* DSM 20640²²⁹, *Leuconostoc citreum*²³⁰, *Bacillus* sp. 217C-11²³¹

and *Lactobacillus johnsonii* NCC533²³² and two more recombinant species who were engineered to produce inulin; *Lactobacillus reuteri* strain 121²³³ and *Lactobacillus gasseri* DSM 20243²²⁹. In psychrophilic/psychrotolerant bacteria, inulin as EPS had not been observed before. FTIR and SEM imaging revealed that inulin was present in 0 °C and 30 °C conditions. Increase in inulin from the high to low temperature signifies inulin production as a stress response by *T. patagoniensis*. Inulin-EPS is observed to be attached around the cells, which may provide additional protection such as resistance to mechanical stresses and defences to viral attacks²³⁴. The industrial production of inulin is usually based on plants such as chicory, dahlia and Jerusalem artichoke^{229,235}. Inulin has significant biotechnological value, especially due its cryoprotectant properties. In addition, inulin is applied during the freeze drying process for probiotics production²²⁸, and is a prebiotic as it supports a healthy gut microbiota²³⁶. Inulin production from *T. patagoniensis* could be of value, because it is a natural producer with a robust phenotype.

T. patagoniensis possibly uses the Wzy-dependent pathway for producing inulin, which is supported by genomics and proteomics analysis. Wzy-dependent pathway proteins were detected in the proteome during both temperature conditions. LAB species use Wzy-dependent pathway for producing homopolymeric polysaccharides of glucans and fructans²²⁷. The biosynthesis of polysaccharides via the Wzy-dependent pathway is a complex process that includes the synthesis of activated sugar precursors, and their assembly in the cell membrane²²⁷ (Figure 6.2). In *T. patagoniensis*, inulin is produced linked with a peptide, which most certainly is the protein EpsA, as it is the last protein in EPS assembly and in expansion of the EPS chain²²⁷. The EpsA protein was significantly more abundant at 0 °C, which may be associated with the observed increase in EPS. The gene *epsA* encodes a LytR-CpsA-Psr family protein, which participates in the assembly of EPS in the membrane²³⁷. Other enzymes may be involved in the assembly of inulin, like EpsE, which is a glucosyltransferase (GT) involved in creating Und-P-P link in repeat units in the EPS chain. *T. patagoniensis* EpsE is predicted to be a transmembrane protein with one transmembrane region, which confirms that this is a GT dedicated to EPS synthesis²³⁷. Following the same perspective, EpsC and EpsD together

form an active bacterial tyrosin (BY)-kinase that is responsible for chain length control of the EPS polymer by acting as a scaffold to keep together the assembly machinery from the other EPS genes. In synergy with the other EPS genes, Wzx flippases are polysaccharides that flip EPS over the cell wall. Wzy polymerase generate glycosidic bonds for adding single repeat units²³⁷. We believe that EpsACBDE unit is responsible for the chain control, assembly, excretion and management of inulin in *T. patagoniensis*.

At the optimal temperature of 30 °C, *T. patagoniensis* growth was completed within 24 hours while at 0 °C, growth was severely reduced. There were no significant differences in the composition of the lipid fraction of the cell wall when exposed to 0 °C, 30 °C or a cold shock (30 to 0 °C). However, there was a clear shift in the soluble proteome from translation at 0 °C to DNA synthesis and repairing at 30 °C. Cold shock proteins were not detected in the proteome, although other proteins with similar function were present. The only cold shock protein with high fold change in 0 °C was transcription termination factor Rho, which terminates transcription by suppressing antisense strand²³⁸. The neighbour gene thioredoxin protects factor Rho from bicyclomycin²⁷. GyrA and GyrB contribute to the adaptation of the cell at low temperatures²²⁷. DEAD-box helicases had a high fold change in the proteome at 0 °C (Figure 6.3), as they are influenced by environmental stress. Previously, it was observed that the deletion of any of the three DEAD-box-helicases encoding genes (*cshA*, *cshB* and *yfmL*) resulted in a cold-sensitive phenotype²³⁹. The gene *yxiN* contributes to the assembly of the ribosomal large subunit by unwinding secondary RNA structures¹⁹ and the genes *cshA* and *cshB* influence cold adaptation of bacteria by initiating transcription at low temperature²³⁹. RhlE helicase had a high fold change at 0 °C and is known to regulate the accumulation of immature ribosomal RNA ribosome precursors by ribosome maturation²⁴⁰. Enzymes involved in the translation mechanism had a high fold change at 0 °C and can affect the formation of 16S and 23S ribosomal units. In addition, high fold change had ribonuclease HIII had a high fold change, which cleaves RNA or DNA-RNA molecules. Mutation of ribonuclease HIII has shown a temperature sensitive phenotype²⁴¹. Ribosomal binding factor had high fold changes and it can promote tolerance in low temperatures by maturing the 16S rRNA

molecule²⁴². Small ribosomal biogenesis GTPase operates synergistically with the binding factor for maturation of ribosomal units²⁴³.

At 30 °C the proteome landscape changes as the proteins that are correlated with replication and DNA regulatory effects had a high fold change. The DEAD-box DNA helicases had a high fold change and are correlated with the correct unfolding of the DNA molecule, in contrast with DEAD-box helicases focused on RNA at 0 °C. RecG had a high fold change and is involved in unfolding branched duplex DNA²⁴⁴. DNA ligases had a high fold change, especially 10 kDA chaperon, which is a stress related chaperon involved in temperature changes²⁴⁵. In this perspective, SAICAR synthetase and phosphate diphosphokinas are involved in the production of purines. Finally, heat shock proteins are chaperones that unfold protein-protein interactions and are expressed during heat stress conditions²⁴⁶. The heat shock 70 family protein is a highly conserved protein among all prokaryotes as a chaperon, which contributes to correct assembly of multi-proteins blocks²⁴⁷.

At low temperature, growth was hampered, and the efficiency of transcription and translation was highly decreased. The low metabolic activity at low temperature requires adjustments in the transcription/translation mechanism for achieving survival. We observed high abundance of enzymes involved in the proper functioning and structuring of RNA and ribosomal units at 0 °C. Inulin production may provide another layer of cryoprotection for the cells. In contrast, the fast growth at 30 °C requires proper DNA proliferation, which is secured by chaperons and nucleotide synthesis enzymes.

ACKNOWLEDGEMENTS

This research was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science.

SUPPLEMENTARY DATA

Table S6.1 | Proteins involved in transcription and translation for the temperatures of 0 °C and 30 °C. With bolt are the fold changes that were significant. N.D.: Non-detected

Protein	Locus Tag	Log(LFQ) Fold Change (0 °C/30 °C)		Locus Tag	Log(LFQ) Fold Change (0°C/30°C)
Cold-sock protein A	Tpat_494	N.D.	Nuclease	Tpat_1430	-1.92
Cold-sock protein A	Tpat_1801	N.D.	Hosphoribosylformylglycinamide synthase subunit PurQ	Tpat_303	-1.47
Cold-sock protein A	Tpat_1802	N.D.	Mitochondrial biogenesis protein aim24	Tpat_50	-1.46
Cold-sock protein A	Tpat_1901	N.D.	mta/sah nucleosidase	Tpat_1917	-1.44
Cold-sock protein A	Tpat_1923	N.D.	Phosphoribosylformylglycinamide synthase subunit PurL	Tpat_304	-1.11
Transcription termination factor Rho	Tpat_2557	0.32	10 kDa chaperonin	Tpat_853	-1.07
Ribosomal protein S6	Tpat_283	0.72	Bifunctional purine biosynthesis protein PurH	Tpat_308	-0.90
GyrB	Tpat_2318	0.24	Phosphoribosylformylglycinamide cyclo-ligase	Tpat_306	-0.87
GyrA	Tpat_284	0.36	Exonuclease VII large subunit	Tpat_519	-0.81
GyrB	Tpat_285	0.34	Homocysteine s-methyltransferase	Tpat_757	-0.80
3-phosphoshikimate-carboxyvinyltrans- ferase (EPSp)	Tpat_2558	0.54	Probable DNA-directed RNA polymerase subunit delta	Tpat_2561	-0.76
Thiol reductase thioredoxin	Tpat_2556	1.76	Mur ligase central	Tpat_2548	-0.73
Acetate kinase	Tpat_1889	-0.27	UDP-N-acetylmuramyl-tripeptide synthetase	Tpat_839	-0.49
S-adenosyl-l-methionine-dependent methyltransferase	Tpat_1890	-0.16	Chaperonins clpa/b signature 1	Tpat_1961	-0.46
Ribosomal protein l27	Tpat_1891	0.94	UDP-N-acetylmuramate--L-alanine ligase	Tpat_2001	-0.41
Hypothetical protein	Tpat_1892	N.D.	Chaperone protein DnaJ	Tpat_2140	-0.38
Ribosomal protein L21	Tpat_1893	-0.05	60 kDa chaperonin	Tpat_852	-0.22
ABC transporter	Tpat_1894	-0.18	Chaperonins clpa/b signature 1	Tpat_1359	-0.23
Amino acid ABC transporter permease	Tpat_1895	N.D.	DNA topoisomerase type IA	Tpat_1263	-0.18
Hypothetical protein	Tpat_1896	-1.25	DNA polymerase III subunit beta	Tpat_288	-0.10
Hypothetical protein	Tpat_1897	-1.13	Mur ligase central	Tpat_1934	-0.03
Cell shape-determining protein mred	Tpat_1898	N.D.	Elongation factor P	Tpat_523	-0.03
Rod shape-determining protein mreC	Tpat_1899	-0.25	Cell division protein FtsZ	Tpat_1930	-0.02
Cell shape determining protein mreB/mrl	Tpat_1900	0.39			
RadC	Tpat_1902	N.D.			
Hypothetical protein	Tpat_1903	N.D.			
Mur ligase -central	Tpat_1904	N.D.			
Aminoacyl-transfer RNA synthetases class-I	Tpat_1905	-0.19			
Alkyl hydroperoxide reductase subunit C	Tpat_1906	N.D.			
Thiamine biosynthesis protein	Tpat_1907	-0.25			
Aminotransferases class-v pyridoxal-phosphate	Tpat_1908	N.D.			
Septation ring formation regulator EzrA	Tpat_1909	0.21			

Table S6.2 | Intact polar lipids for the two conditions of 0 °C and 30 °C and the cold shock from 30 °C to 0 °C. There are differences in the intact polar lipid composition regarding dihexose and cardiolipin.

Temperature (°C)	Hexose (%)	APT (%)	PG (%)	Dihexose (%)	Cardiolipin (%)
0	2.5 ± 1.4	0.0 ± 0.0	23.0 ± 14.0	35.7 ± 5.7	41.5 ± 8.4
30	1.4 ± 0.2	0.0 ± 0.0	47.8 ± 2.8	18.8 ± 1.1	33.4 ± 1.8
30-0	3.5 ± 0.1	10.2 ± 3.5	45.9 ± 1.1	26.4 ± 0.4	17.5 ± 2.5

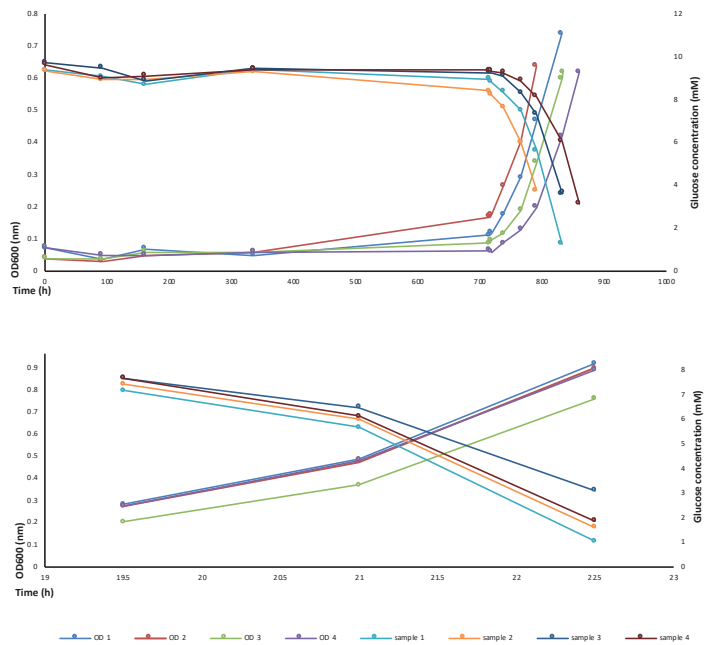


Figure S6.1 | The time points where the proteomics samples were selected for cell harvest. The glucose consumption of the range of 3.5-3.7 mM was considered as the extraction point. On the left the 0 °C growth and on the right the 30 °C growth.

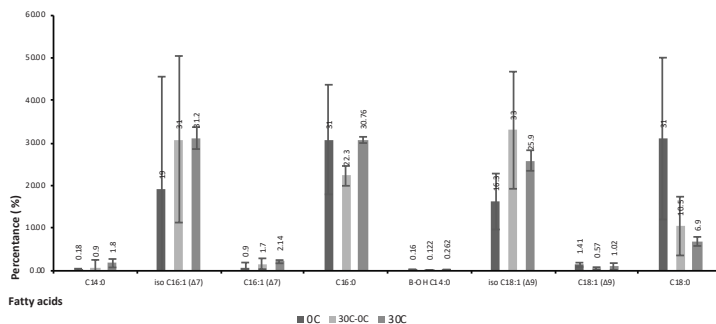


Figure S6.2 | Graph representation of the lipid composition in the two temperatures of 0 °C and 30 °C and cold shock 30 to 0 °C. Although, there is a significant deviation through the replicates, we observe an exchange of unsaturated fatty acids at 30 °C and saturated in 0 °C.

CHAPTER 7

***TRICHOCOCCUS SHCHERBAKOVIAE* SP. NOV. ISOLATED FROM A LABORATORY-SCALE ANAEROBIC EXPANDED GRANULAR SLUDGE BED BIOREACTOR OPERATED AT LOW TEMPERATURE**

Sofiya N Parshina^{1,2*}, Nikolaos Strepis^{2,3}, Steven Aalvink², Alla N Nozhevnikova¹, Alfons JM Stams², Diana Z Sousa²

1. Laboratory of Winogradsky Institute of Microbiology, Federal Research Centre of Biotechnology of Russian Academy of Sciences, Moscow, Russia

2. Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

3. Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands

ABSTRACT

A new species of the *Trichococcus* genus, strain Art1^T, was isolated from a psychrotolerant syntrophic propionate-oxidizing consortium, obtained before from a low-temperature EGSB fed with a mixture of VFA's (acetate, propionate and butyrate). 16S rRNA gene sequence of strain Art1^T was highly similar to those of other *Trichococcus* species (> 99 %), but digital DNA-DNA hybridization (dDDH) values were lower than 70 % indicating that strain Art1^T is a novel species of the genus *Trichococcus*. Cells of strain Art1^T are immotile cocci with a diameter of 0.5-2.0 µm and were observed singularly, in pairs, short chains and irregular conglomerates. Cells of Art1^T stained Gram-positive and produced extracellular polymeric substances (EPS). Growth was optimal at pH 6-7.5 and cells could grow in a temperature range of -2 to 30 °C (optimum 25-30 °C). Strain Art1^T can degrade several carbohydrates, and the main products from glucose fermentation are lactate, acetate, formate, and ethanol. The genomic DNA G+C content of strain Art1^T is 46.7 %. The major components of the cellular fatty acids are C_{16:1ω9c}, C_{16:0} and C_{18:1ω9c}. Based on genomic and physiological characteristics of strain Art1^T, a new species of the *Trichococcus* genus, *Trichococcus shcherbakoviae* is proposed. The type strain of *Trichococcus shcherbakoviae* is Art1^T (=DSM 107162^T, VKM B-3260^T).

The genus *Trichococcus* was first described in 1984 by Scheff *et al.*²⁴⁸, with the isolation of the type strain *Trichococcus flocculiformis* from activated sludge. Since then, other seven *Trichococcus* species have been characterised from different ecosystems: *T. collinsii*, *T. pasteurii*, *T. palustris*, *T. patagoniensis*, *T. ilyis*, *T. paludicola*, *T. alkaliphilus*^{49,67,68,108}. These species show variations in their phenotypic and physiological properties, but all have very high similar sequences of their 16S rRNA gene (99.7-99 %) ^{68,108}. To correctly assign new species within this genus DNA-DNA hybridization, either experimental or *in silico*, is required⁷². Currently, all the isolated *Trichococcus* species have their genome sequenced, which facilitates *in silico* approaches. Tools like the Genome-to-Genome Distance Calculator (GGDC)¹⁷ and the method of Average Nucleotide Identity (ANI)¹² provide a fast and reliable basis for the delineation of new species. Cut-off points for species delinea-

tion have been set at 70 % for dDDH and 95 % for ANI²⁴⁹.

The genus *Trichococcus* belongs to the *Carnobacteriaceae* family within the phylum *Firmicutes*. Members are often detected in anaerobic digester sludges²⁵⁰⁻²⁵⁵, where they are thought to be involved in the conversion of carbohydrates to lactate, acetate, formate and other acids or alcohols. The high metabolic versatility of *Trichococcus*, for example their tolerance to oxygen and metals and capacity to adapt to temperature variations and to high salinity environments, may give members of this genus an advantage in these ecosystems.

In this study, we describe strain Art1^T, a psychrotolerant fermentative bacterium isolated from a low-temperature anaerobic expanded granular sludge bed bioreactor (EGSB) seeded with mesophilic granular sludge and fed with a mixture of acetate, propionate and butyrate. The bioreactor was operating at temperatures of 3-8 °C for period for 1.5 years²⁵⁶. After series of dilutions on propionate in serum bottles a psychrotolerant syntrophic propionate-oxidizing consortium was obtained²⁵⁷. Strain Art1^T was isolated from this consortium after series of dilutions on glucose at 20 °C. Products from glucose fermentation were lactate, formate, acetate and ethanol.

The culture medium used for the isolation of the bacterium and growth tests was a bicarbonate-buffered medium⁸⁰ supplemented with salts, vitamins and 0.1 g l⁻¹ yeast extract and 20 mM glucose. Cultivations were performed in serum bottles of 120 mL with 50 mL of medium or in Hungate tubes with 5 mL of medium. The headspace of all vials was flushed and pressurized with N₂/CO₂ (80/20, v/v) to a final pressure of 1.5 bar. Substrates were added to the medium from sterile stock solutions. Growth was measured by optical density (OD_{600nm}) with a DiluPhotometer OD₆₀₀ Spectrophotometer (Westburg, Leusden, the Netherlands). Soluble metabolites (sugars, volatile fatty acids and alcohols) were measured with a Thermo Electron HPLC system equipped with an Agilent Metacarb 67H column (300 mm x 6.5 mm) and refractive index and UV detectors (Thermo, Waltham, MA). The mobile phase used was sulfuric acid (0.01 N) at a flow rate of 0.8 mL min⁻¹. The column temperature was set at 45 °C. Possible gaseous products (H₂, CO₂) were monitored with a Shimadzu GC-2014 gas chromatograph equipped with a Molsieve 13X column (2 m x 3 mm)

and a thermal conductivity detector (TCD) (Shimadzu, Kyoto, Japan). Argon was used as carrier gas at a flow rate of 50 mL min⁻¹, and temperatures in the injector, column and detector were 80, 100 and 130 °C respectively.

For amplification of the 16S rRNA gene, DNA was isolated by a modified alkaline extraction procedure²⁵⁸ and further on purified using the Wizard® Kit (Promega, Washington, DC). For amplification of the nearly full-sized 16S rRNA gene, universal bacterial primers 27f and 1492r were used. PCR was conducted in 50 µL reactions with the following reagents: 16 PCR buffer [17 µL (NH₄)₂SO₄, 67 mM tris-HCl, pH 8.8, 2 mM MgCl₂]; 12.5 nmol of each dNTP, 50 ng of DNA template; 5 pmol of corresponding primers and 3 units of DNA polymerase BioTaq (Dialat, Moscow, Russia). The temperature-time profile for the amplification was: initial denaturation at 94 °C for 9 min followed by 30 cycles consisting of denaturation at 94 °C for 1 min, primer annealing at 55 °C for 1 min and elongation at 72 °C for 2 min and a final elongation at 72 °C for 7 min. A Tetrad2 thermocycler (Bio-Rad, Moscow, Russia) was used. PCR products were analyzed by electrophoresis in 2 % agarose at 6 V/cm². Gels were photographed using the BioDocII video documentation system (Biometra, Göttingen, Germany). The amplification products were sequenced on an ABI PRISM 3730 automatic sequencer (Applied Biosystems, Foster City, CA) using the Big Dye Terminator v.3.1 kit (Applied Biosystems) according to the manufacturer's instructions. The sequence chromatograms were edited using the Chromas software, version 1.45 (<http://www.technelysium.com.au/chromas.html>).

16S rRNA gene sequences of strain Art1^T and all the currently isolated *Trichococcus* spp. were added to the all-species living tree project⁹⁹. Phylogenetic tree was reconstructed using ARB v6.0⁹⁷, applying the neighbour-joining method (Jukes-Cantor correction) (Figure 7.1a). 16S rRNA gene identity between strain Art1^T and other *Trichococcus* species was higher than 99%. To further clarify the phylogenetic assignment of strain Art1^T its genome was sequenced. Genomic DNA of strain Art1^T was extracted using MasterPure Gram positive DNA purification Kit (Epicenter, Madison, WI) according to manufacturers' instructions. The genome of strain Art1^T was sequenced at NovoGene

(Beijing, China) using an Illumina Hiseq X Ten platform (Illumina Inc., San Diego, CA) and assembled as previously described¹⁰⁸.

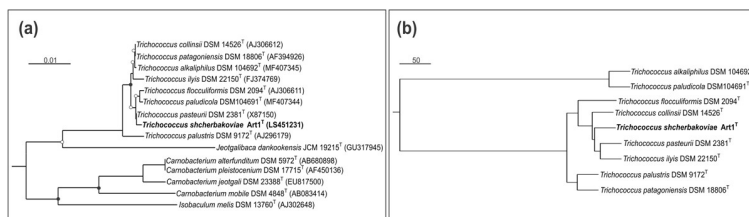


Figure 7.1 | **(a)** 16S rRNA gene-based phylogenetic tree showing the position of strain Art1^T relative to other neighbouring isolated strains. Tree (bootstrapping value of 1000 trees) was reconstructed using ARB software package and applying the neighbour-joining method with Jukes-Cantor correction. The significance of each branch is indicated at the nodes with filled-circles (values above 70 %) and open-circles (values above 50 %). Bar, 1 % sequence divergence. **(b)** Functional clustering of *Trichococcus* species based on matrix of presence/absence of protein domains. The tree was generated using a hierarchical clustering. Bar, 50 domains difference between species.

Strain Art1^T genomic sequence was assembled in 55 scaffolds with a total size of 3.1 Mb and a G+C content of 46.7 %. For the overall genome related index (OGRI)²⁵⁹, the values of average nucleotide identity (ANI) and digital DDH (dDDH) were calculated for members of the genus *Trichococcus*. The dDDH values between strain Art1^T and other *Trichococcus* species were well below the cut off value for species delineation (Table 7.1), indicating that strain Art1^T is a novel species of the genus *Trichococcus*. Strain Art1^T was most similar with *T. alkalicophilus* (dDDH 47.2%; ANI 91.5 %) and *T. pasteurii* (dDDH 46.7 %; ANI 91.7 %). Genomic annotation was conducted with functional protein domains using the platform SAPP¹¹⁷. For improving phylogenomics, a functional clustering of all *Trichococcus* species based on matrix of presence /absence protein domains was generated in R¹²⁰ using hierarchical clustering method (Figure 7.1b).

Table 7.1 | dDDH (and confidence intervals) and ANI values between strain Art1^T and other type strains the *Trichococcus* genus. Thresholds for the definition of new species are 70 % and 94 % for dDDH and ANI, respectively²⁴⁹.

Strain Art1 ^T in comparison with:	dDDH (%)	ANI (%)
<i>Trichococcus collinsii</i> DSM 14526T	43.0 ± 5.0	90.6
<i>Trichococcus pasteurii</i> DSM 2381T	46.7 ± 5.2	91.7
<i>Trichococcus patagoniensis</i> DSM 18806T	42.6 ± 5.0	89.8
<i>Trichococcus ilyis</i> DSM 22150T	31.1 ± 4.9	85.8
<i>Trichococcus flocculiformis</i> DSM 23957T	42.5 ± 5.0	90.2
<i>Trichococcus flocculiformis</i> DSM 2094T	41.6 ± 5.0	90.3
<i>Trichococcus palustris</i> DSM 9172T	23.5 ± 4.8	77.3
<i>Trichococcus paludicola</i> DSM 104691T	44.1 ± 5.1	90.9
<i>Trichococcus alkaliphilus</i> DSM 104692T	47.2 ± 5.2	91.5

Cell morphology of strain Art1^T (glucose-grown cultures) was observed using a phase-contrast microscope Axiomager DI (CarlZeiss, Germany) in the Core Facility "UNIQEM" Collection. Cells of strain Art1^T are immotile cocci with a diameter of 0.5-2.0 µm and were observed singularly, in pairs, short chains and irregular conglomerates (Figure 7.2a). Gram staining was done according to standard procedures²⁶⁰. Cells of Art1^T stained Gram-positive. For scanning electron microscopy analysis (SEM), the culture was adhered to poly L-lysine 12 mm coated coverslips (Corning, BioCoat, Horsham, PA) and incubated for 1 h at room temperature. The cells were then fixed in 2.5 % glutaraldehyde in 0.1 M phosphate buffer (pH 7.4) for 1 h at room temperature, rinsed 3 times with 0.1 M phosphate buffer (pH 7.4) and post-fixed with 1 % osmium tetroxide for 60 min. Hereafter cells were dehydrated in a graded alcohol series (10, 30, 50, 70, 80, 96 and 100 %), dried to critical point in 100 % ethanol with CO₂ in a Leica EM CPD300 system (Leica Microsystems, Wetzlar, Germany) and mounted into aluminium stubs and coated with tungsten. Cells were subsequently studied with a FEI Magellan 400 FE scanning electron microscope (FEI Company, Hillsboro, OR). SEM observations revealed that cells of strain Art1^T were enclosed within a thick extracellular matrix (Figure 7.2b). Fimbria-like structures could also be observed. The analysis of extracellular polymeric substances (EPS) according to the method described by Dubois *et al.*²⁶¹ confirmed their presence. The presence

of EPS structures was observed previously in *T. patagoniensis*⁴⁹. EPS can function as a protection layer from mechanical and temperature stress and possibly lead to biofilm creation.

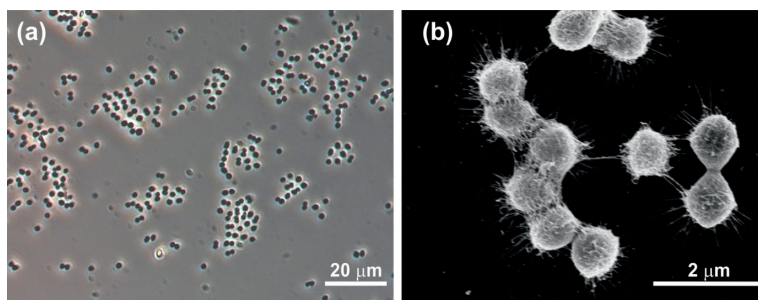


Figure 7.2 | (a) Phase contrast micrograph and (b) SEM images of strain Art1^T (glucose-grown cultures).

The fifth dilution of the morphologically monotypic culture was chosen for growth of colonies in roll-tubes with a 2 % (w/v) agar medium. After incubation for 5 days at 25 °C, colonies of strain Art1^T were white, circular, with a denser centre and had a diameter of 0.5-3.0 mm. The margins were smooth and thinner than the central part of colony. The consistency of the colonies was mucoid and slimy.

Fatty acids analysis of strain Art1^T was performed in DSMZ (Braunschweig, Germany). The analysis was done with the system of Sherlock MIS (MIDI Inc, Newark, NJ). Cells for lipids analysis were cultivated in pre-reduced tryptic soy broth (BD, Franklin Lakes, NJ) as previously used for other *Trichococcus* strains⁶⁷. The major fatty acids found in *Trichococcus* Art1^T and of other *Trichococcus* species are shown in the Table 7.2. Major fatty acid methyl esters (FAMES) in strain Art1^T were C_{14:0} (16.13 %), C_{16:0} (19.68 %) and C_{18:1 ω⁹C} (19.65 %).

Strain Art1^T is aerotolerant. The oxidase and catalase tests were negative. Optimal conditions for growth were tested with glucose as substrate. Growth was observed between -2 °C and 37 °C with an optimum at 25 - 30 °C (tested range: -4 - 45 °C). The optimal pH for growth was 6.0 - 7.5 with growth occurring between 5.0 and 9.5 (tested range: 5.0-9.5). Strain Art1^T could grow at NaCl concentrations from 0 to 5.5 g/L (optimum 0.0-2.5) (tested range: 0.0-8.0).

Table 7.2 | Cellular fatty acids profile of strain Art1^T in comparison with other members of the genus *Trichococcus*. Highlighted in bold are three most predominant fatty acids. Strains: 1, Art1^T (data from this study); 2, *T. paludicola* DSM 104691^T⁶⁸; 3, *T. alkaliphilus* 104692^T⁶⁸; 4, *T. ilyis* DSM 22150^{T108}; 5, *T. patagoniensis* DSM 18806^{T49}; 6, *T. pasteurii* DSM 2381^{T67}; 7, *T. collinsii* DSM 14526^{T67}; 8, *T. palustris* DSM 9172^{T67}; 9, *T. flocculiformis* DSM 2094^{T248}.

Fatty acids	1	2	3	4	5	6	7	8	9
C _{12:0}	1.1			0.7	0.7	0.0	12.0	0.0	1.0
C _{14:0}	9.0			16.1	11.2	14.0	57.0	21.0	28.0
C _{15:0}				0.6	0.0	0.0	0.0	0.0	0.0
C _{16:1ω9c}	38.6			35.5	0.0	46.0	18.0	20.0	0.0
C _{16:0}	10.9	25.4	31.4	19.7	16.5	15.0	14.0	15.0	16.0
C _{12:0}	1.1			0.7	0.7	0.0	12.0	0.0	1.0
C _{14:0}	9.0			16.1	11.2	14.0	57.0	21.0	28.0
C _{15:0}				0.6	0.0	0.0	0.0	0.0	0.0
C _{16:1ω9c}	38.6			35.5	0.0	46.0	18.0	20.0	0.0
C _{16:0}	10.9	25.4	31.4	19.7	16.5	15.0	14.0	15.0	16.0
C _{18:1ω9c}	29.5			19.7	21.8	18.0	0.0	22.0	6.0
C _{18:1ω7c}	1.5			0.6	0.0	0.0	0.0	0.0	0.0
C _{18:0}	2.2	14.5	31.4	3.0	3.3	2.0	0.0	4.0	2.0
C _{20:1ω9c}	1.3			0.8	0.0	0.0	0.0	0.0	0.0
C _{20:1ω7c}	1.0			0.8	0.0	0.0	0.0	0.0	0.0

Trichococcus strain Art1^T utilized (at a concentration 20 mM, unless stated otherwise) a broad range of carbon sources for growth, namely: glucose, D-arabinose, D-galactose, D-fructose, D-lactose, maltose, D-mannitol, sucrose, raffinose, trehalose, D-ribose, cellobiose, D-mannose, pyruvate, D-gluconate. Weak growth was observed with D-xylose and D-sorbitol. No growth was observed with formate, D-galacturonate, D-gluconate, L-sorbose, D-tagatose, xylose, starch, glycerol, ethanol, methanol, inositol, dulcitol and D-sorbitol. The strain does not grow autotrophically on H₂/CO₂ (80/20 (v/v), 1.5 atm). Physiological characteristics are summarized in Table 7.3.

Table 7.3 Morphological and physiological characteristics of described species of the genus *Trichococcus*. Strains: 1, Art1^T (data from this study); 2, *T. paludicola* DSM 104691^{T68}; 3, *T. alkaliphilus* 104692^{T68}; 4, *T. ilyis* DSM 22150^{T108}; 5, *T. patagoniensis* DSM 18806^{T49}; 6, *T. pasteurii* DSM 2381^{T67}; 7, *T. collinsii* DSM 14526^{T67}; 8, *T. palustris* DSM 9172^{T67}; 9, *T. flocculiformis* DSM 2094^{T248}. +, Positive; - Negative; ND, Not determined.

Source	1	2	3	4	5	6	7	8	9
	Low temperature anaerobic bioreactor	Soil of the Zoige Wetland	Soil of the Zoige Wetland	Sulfate-reducing bioreactor	Guano penguin	Septic pit	Hydrocarbon spill site	Swamp	Activated Sludge
Gram stain	+	+	+	+	Variable	+	+	+	+
Cell length (µm)	0.5-2.0	0.8-2.0	0.8-2.0	0.63-1.4	1.3-2.0	1.0-1.5	1.0-2.5	1.0-2.5	1.0-2.5
Cell shape	Cocci	Cocci	Cocci	Cocci	Cocci	Short chains of cocci	Rods	Cocci	Long chains of cocci
pH range (optimum)	5.0-9.5 (6.0-7.5)	7.0-10.5	7.0-10.5	6.0-9.6 (7.8)	6.0-10 (8.5)	5.5-9.0	6.0-9.0 (7.5)	6.2-8.4 (7.5)	5.8-9.0
Temperature range (optimum)	-2-37 (20-30)	4-37 (25)	4-37 (25)	4-40 (30)	-5-35 (28-30)	0-42 25-30	7-36 (23-30)	0-33	4-40 (25-30)
DNA G+C content									
In vitro	ND	46.0	46.7	ND	45.8	45.0	47.0	48.0	48.0
In silico	46.7	ND	ND	47.9	46.9	45.7	44.3	45.9	44.3
Major cellular fatty acids (%)	38.57 10.94 29.54	20.0 15.0 22.0	0.0 16.5 21.8	46.0 15.0 18.0	35.5 19.7 19.7 ^T	25.4	31.4 11.4	18.0 14.0 0.0	0.0 16.0 6.0
Glucose	+	+	+	+	+	+	+	+	+
D-Arabinose	+	ND	ND	+	+	+	ND	ND	+
D-Galactose	+	+	+	+	ND	+	+	ND	+
D-Fructose	+	+	+	+	+	+	ND	ND	+
D-lactose	+	+	+	+	-	+	-	-	+
Maltose	+	+	+	+	+	+	-	-	+
D-Mannitol	+	+	+	+	+	+	-	-	-
Sucrose	+	+	+	+	+	+	ND	ND	+
Raffinose	+	+	+	+	+	-	+	-	-
Trehalose	+	+	+	ND	-	+	+	-	+
D-Ribose	+	ND	ND	ND	-	+	+	-	-
Cellobiose	+	+	+	+	+	+	+	+	+
D-Mannose	+	+	+	+	+	+	+	+	+
Pyruvate	+	ND	ND	ND	+	+	+	+	ND
D-Gluconate	+	+	+	+	ND	+	-	ND	ND
D-Xylose	-	ND	ND	+	ND	+	ND	ND	+
Ethanol	-	ND	ND	-	-	ND	ND	ND	ND
D-sorbitol	-	+	+	-	+	+	+	+	-
Glycerol	-	-	-	-	-	+	-	-	-

Description of *Trichococcus shcherbakoviae* sp. nov.

T. shcherbakoviae (shcherbakov'i.ae. N.L. gen. n. *shcherbakoviae* of Scherbakova, named after Dr. Viktoriya A. Shcherbakova, a Russian microbiologist, for her big contribution to investigation of low-temperature environments and psychrotolerant microorganisms).

Cells are non-motile Gram-stain-positive cocci with a diameter of 0.5–2.0 µm. Cells grow single, in pairs, short chains or irregular aggregates. Cells grow in a temperature range of -2 °C to 37 °C and the optimum is 20–30 °C. The ranges pH for growth are from 5.0 to 9.5 and the optimal growth is achieved at pH 6.0–7.5. The range of NaCl concentrations that allow growth is 0–5.5 g L⁻¹ (optimum 0–2.5 g L⁻¹). Strain Art1^T can utilize a broad range of carbon sources for growth, such as pyruvate, malate, arabinose, citrate, fructose, galactose, D-glucuronate, glucose, mannitol, mannose, rhamnose, cellobiose, lactose, maltose, sucrose and raffinose. No growth was observed with formate, D-galacturonate, D-gluconate, L-sorbose, D-tagatose, xylose, starch, glycerol, ethanol, methanol, inositol, dulcitol and D-sorbitol. Strain Art1^T did not grow autotrophically on H₂/CO₂. Major fatty acid methyl esters (FAMES) in strain Art1^T were C_{14:0}, C_{16:0} and C_{18:1 ω9c}.

The type strain of *Trichococcus shcherbakoviae* is Art1^T (=DSM 107162^T = VKM B-3260^T), isolated from psychrophilic methanogenic community that degraded a mixture of acetate, propionate and butyrate. The genomic G+C content of the type strain is 46.7 mol%.

ACKNOWLEDGEMENTS

This research was supported by Governmental Funding from the Russian Federation Federal Agency of Scientific Organizations for fundamental studies (N° 0104-2018-0037) in Research Center of Biotechnology, Russian Academy of Science under the topics "Collection of unique and extremophile microorganisms for biotechnology purposes" (N° ISGZ 0104-2016-001). The work performed by SNP at Wageningen University & Research was funded by a visit-

ing scientist grant attributed by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) (grant number 040.11.645). NK, AJMS and DZS were financed by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science.

CHAPTER 8

| 159 |

GENERAL DISCUSSION AND FUTURE RECOMMENDATIONS

DISCUSSION

Approximately 149,000 genomes are currently present in the NCBI database and new genomes are constantly becoming available²⁶². There is a vast amount of knowledge stored in these genomic repositories. Such knowledge can contribute to the genome based identification of microbial processes. Specific phenotypes of microbes can be used for biotechnological prospecting especially those that are related to the production of valuable compounds. The use of omics studies (genomics and proteomics) and physiological trials can characterize the metabolic potential of each microbe. This is the strategy that was applied in this thesis for linking genotypes to interesting phenotypes (Figure 8.1). The main focus of this approach rested on *Trichococcus* species which have distinct phenotypes with a set of unique genes. Members of the genus *Trichococcus* can produce high concentrations of 1,3-propanediol (1,3-PDO) and exopolysaccharides. Furthermore, the knowledge obtained from the genotypes related to 1,3-PDO production in *Trichococcus* strain ES5 and *T. pasteurii* was used for investigating all deposited bacterial sequenced genomes.

Distinguishing diverse microbial phenotypes

Although, *Trichococcus* species were isolated from diverse environments⁵⁰, they all displayed a number of shared characteristics. They utilize a broad spectrum of carbohydrates, have high-salinity resistance and survive at low temperatures. In this thesis, besides the shared physiological properties, metabolic diversity was addressed for the genus and this was mainly illustrated by the production of 1,3-PDO and inulin in specific members. Additionally, the search for interesting phenotypes, led to the description of two novel *Trichococcus* species. Though the description of *Trichococcus ilyis* and *Trichococcus shcherbakoviae* the genus was expanded to a total of 9 validated species and one subspecies (*Trichococcus* strain ES5).

The description of novel species for the genus *Trichococcus* may be more challenging than in other genera. The 16S rRNA gene of the *Trichococcus* species is highly conserved. Therefore, the 16S rRNA gene sequence similarity threshold (98,7 %²⁶³) is not valid for delin-

eating bacterial species in the *Trichococcus* genus. One example is *T. patagoniensis* and *T. collinsii*, which have a 100 % similar 16S rRNA gene sequence and that creates complications in the phylogenomic clustering. The *Trichococcus* species is not the only genus with conserved 16S rRNA genes; other examples are *Edwardsiella*, *Clostridium* and *Mycobacterium*^{10,263,264}. In general, 57 % of bacterial species have a high 16S rRNA gene similarity (above 98.7 %)²⁶³. The 16S rRNA gene similarity influences the novel species description, because the precise assignment of bacterial lineage may be questionable, and therefore multiple approaches are based on omics studies²⁶⁵ for species description. In this thesis, classification of new *Trichococcus* species (*T. ilyis* and *T. shcherbakoviae*) was conducted by complementing 16S rRNA gene analysis with a genome-based method, digital DNA-DNA hybridization (dDDH). This approach was also applied for the assignment of *T. paludicola* and *T. alkaliphilus*⁶⁸. I can conclude that the use of genomics information (such as dDDH) can contribute to the taxonomical clustering of novel species in the genus of *Trichococcus* and in other genera as an efficient and accurate approach.

Genomic guidance for phenotypic description

Genomics on *Trichococcus* species indicated clear variations among the species in nucleotide base composition and gene content (protein domains). Many members of the genus *Trichococcus* have genotypes linked with high salinity robustness, low temperature survival and specific substrates utilization (Chapter 3). Several of these phenotypes were not previously reported for individual species, such as production of 1,3-PDO and the degradation of specific substrates (arabinan and alginate). In *T. pasteurii*, a complete operon of 17 genes was related to 1,3-PDO production and in vitro *T. pasteurii* could produce 1,3-PDO from glycerol. This operon, which was completely conserved in genes and architecture, was identified in the genome of *Trichococcus* strain ES5. Arabinan is usually found in plant cell walls and its degradation is difficult. In vitro, *T. patagoniensis* was confirmed as a bacterial species that can degrade arabinan. Similarly, *T. collinsii* was verified to degrade alginate, a polysaccharide that influences EPS formation¹³¹. In conclusion, genotypic information lead to valid predicted phe-

notypes. I would like to underline that in vitro analysis is needed to confirm genomics predictions. Frequently, phenotypes with genomic evidence are not in vitro expressed. This was the case for substrates as tagatose, starch and L-sorbose, where the genome indicated their utilization by *T. ilyis*, but in vitro bacterial growth was not achieved with these compounds.

From the same perspective, all *Trichococcus* species possessed a high number of cold shock domains (CSD), genes related with a psychrotolerance phenotype²⁶⁶. Surprisingly, only four species could grow at 0 °C. In addition, to understand CSD in *Trichococcus* species, we included 20 lactic acid bacteria (LAB), belonging to the genera of *Carnobacterium* and *Aerococcus*. The species of these genera that had multiple CSDs were also detected with numerous genes related to high salinity tolerance, which resembled those in *Trichococcus* species. Furthermore, CDS were in particular present in *Trichococcus*, *Carnobacterium* and *Aerococcus* species isolated from ecosystems with a low temperature. Further studies should be conducted to gain insights into the co-existence of CSD related to high salinity tolerance genes in these psychrophilic/psychrotolerant species.

Similar studies based phenotypic analysis on the genotypic and phylogenetic information. Comparative genomics was implemented in 75 *Clostridium difficile* strains using DNA microarrays for modeling physiology and phylogeny²⁶⁷. The study was successful and multiple *C. difficile* strains were observed to have genotypes with determinate characteristics such as motility, antibiotic resistance and toxicity, properties that were validated in vitro. This methodology differs from the one in the present thesis in the part that not only a genotype-phenotypic link was established, but additionally the architecture of genes in the conserved operons of the trait was defined. In another study, *Deinococcus ficus*, an ionizing radiation resistant bacterium was discriminated from other members in the genus based on genomic information²⁶⁸. A core/pan genome analysis on the genus of *Deinococcus* was conducted to identify unique genes in *D. ficus*. We followed a similar approach for *T. ilyis*, with the exception that we based our analysis on a core/pan protein domains analysis. Therefore, a phylogenomic separation based on functional domain clustering of the species was

conducted. Finally, in a study like *Trichococcus* genotypic analysis (Chapter 3), genomics was conducted in 101 strains of *L. rhamnosus* species, where carbohydrate substrate utilization and other phenotypes were predicted and eventually tested in vitro²⁶⁹. A remarkable finding is the discovery of the EPS assembly genes in *L. rhamnosus* strains and the verification of EPS production in vitro similar to our study of *T. patagoniensis*. The genomic approach used in this study resembled our methods, however, it was based on orthogroups with the use of eggNOG (v4.5)²⁷⁰. That is a powerful tool with well annotated functional categories. However, it still is a homologues method and the database is outdated (last update in 2016). Considering these studies, I would like to advocate the importance of protein domains for an accurate analysis due to their constant precise annotations and their link to the protein function.

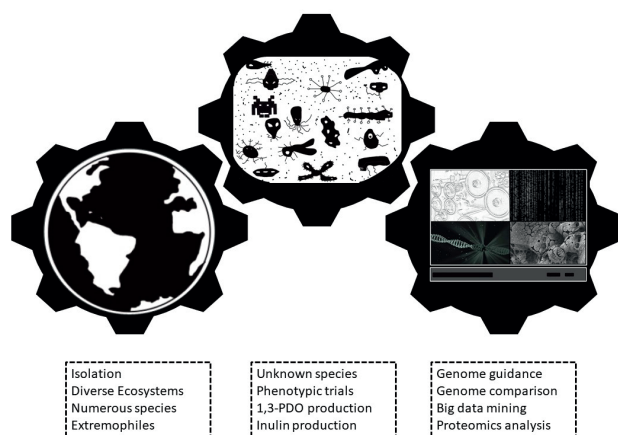


Figure 8.1 | Representation of the three major topics that were the “gears” of my thesis. The isolation of novel microorganisms that fuelled the exploration of unknown species by phenotypic trials. The application of omics studies can be used for novel, but also described species and knowledge on the physiology of individual phenotypes can be obtained. All gears are connected and move in a synchronized fashion. Omics studies can guide the species description for specific metabolic traits and the in vitro confirmation provides further metabolic targets for omics analysis. This three gears approach, I would like to recommend further studies on the physiological characterization of novel microorganisms.

Mining big data

In this thesis, we performed a mining of the complete public repository of genomes. An important aspect when considering the mining of available genomes is the quality and frequency of updates in gene annotation. Microbial genomes stored in public databases originate from multiple annotation platforms (orthology based, SEED database and proteins domains)^{39,271,272}, thus the gene description may differ. Specifically, methods based on orthology are prone to errors due to propagating mis-annotations. Consistency and accuracy in the description of genomic sequences is a necessity, and for this purpose, we choose functional analysis for all microbial genomes. The functional analysis is based on predicted protein domains, conserved structure protein units that are annotated by X-ray crystallography²⁷³ and Hidden Markov Models. Functional analysis contributes to precise genotypic mapping, avoiding annotation errors. In this thesis, we annotated all genomic sequences with SAPP¹¹⁷, which is based on protein domains and generates a semantic database. Furthermore, this new database of all available microbial genome, is a powerful source of genomic information that can be utilized for multiple research questions as long as it undergoes constant re-annotations in the future. Recently, the idea of a consistent and accurate database was adopted by NCBI RefSeq project, re-annotating 95,000 prokaryotic genomes²⁷⁴. A comprehensive reannotation will always report unbiased and error-free genomic evidences for phenotypes. In this thesis, we recorded cases in which the homological search categorized glycosyltransferases and cold adaptation proteins as hypothetical, whereas the protein domains accurately classified their function (Chapter 3 and 6).

We realized that using all available genomic sequences with an accurate annotation will expand knowledge of the 1,3-PDO production trait and possibly the discovery of novel producers. The vast amount of genomic knowledge and metabolic potential contained in all those genomes are usually not utilized. Often in research, genomes are analyzed for specific genomic targets and not for all the available genomic knowledge. Although not all metabolic traits are explored, genomic information is there and available for mining. In Chapter 4, we focused on the metabolic trait of 1,3-PDO production by searching in all available bacterial genomes for genomic syntenies like the one

in *T. pasteurii*. Based on this holistic analysis, we showed that several other bacteria are potential 1,3-PDO producers. Regarding vitamin B12-dependent glycerol dehydratase species, we identified 187 potential candidates for 1,3-PDO production and several were in vitro validated (Figure 8.2). A major contribution of this study the metabolic landscape of 1,3-PDO production was the identification of species with a vitamin B12-independent glycerol dehydratase: *Clostridium diolis*, *Clostridium beijerinckii*, *Clostridium butyricum* E4 and *Raoultella planticola*. So far, the only B12-independent glycerol dehydratase species was *C. butyricum* (VPI 3266 and VPI 1718)^{203,213}. Many studies have described extensively the properties of this species and the B12-independent glycerol dehydratase gene was targeted for bioengineering and the transformation of microbes into 1,3-PDO producers^{52,275,276}. We managed to provide four potential additional strains with B12-independent glycerol dehydratase. Furthermore, *C. diolis* and *C. beijerinckii* are genetically accessible and thus engineering can be applied on their genomes^{159,277}. That more B12-independent species exist in nature is an important outcome.

We defined the operon architectures that is in a functional metabolic trait and identified potential new natural 1,3-PDO producers. The screening of the data with a single 1,3-PDO dehydrogenase domain revealed the architecture of the B12-dependent-1,3-PDO operon. In this genomic synteny nine conserved domains were identified (Chapter 4). In the case of the B12-independent-1,3-PDO operon six conserved domains were identified. Furthermore, the results were validated and evaluated by operons from known 1,3-PDO producers. In this study, the mining of the data is completely unbiased which removes the need for a priori knowledge (weight factors) for the search. Previously, a screening based on weighted ranks was conducted for free fatty acids producers in all the available species of *Cyanobacteria* (120 genomes)²⁷⁸. A ranking algorithm was designed using physiological criteria and the weight skew for each gene in the trait was generated. Although such studies produce significant results, they are based on background knowledge and sometimes on hypothesis, which may be misleading. In this thesis, we implemented an approach that focuses on design, build, learn and test, which provides a constant flow of information that improves the results.

An example of applying knowledge obtained through big data mining is the proper assignment of the 1,3-PDO dehydrogenase domain. Previously, the domain PF00465 was related to 1,3-propanediol dehydrogenase due to the enzyme being an iron-containing alcohol dehydrogenase²⁷¹. However, in this analysis, the 1,3-PDO dehydrogenase gene in the operons of identified natural producers had a domain with a different e-value from the general PF00465 domain. Therefore, we designed a new HMM by identifying the conserved amino acids in the 1,3-PDO dehydrogenase gene of the producers. The new 1,3-PDO dehydrogenase domain was used to screen the data for 1,3-PDO producers and was validated by assigning the same species as previously. Therefore, this novel 1,3-PDO dehydrogenase domain is optimal for screening 1,3-PDO producers. Future studies can implement this domain in big data and even metagenomics samples. This approach is not only limited for 1,3-PDO dehydrogenase. The mining of big data can generate functional domains for any enzyme, which can be applied to tracing the physiology in multiple genomes or even metagenomics sample.

The phenotypic characteristics of the species were linked with the results of the big data mining. The possible 1,3-PDO producers were matched with *in silico* stored physiological data for growth temperature, pathogenicity and vitamin B12 synthesis. The search for accurate digitally stored phenotypic records was difficult. We used BacDive¹⁷⁹ from DSMZ which is a reliable source of physiological properties, but with a limited number of categorized species. Certainly, the digital recording of physiological metadata is a laborious challenge, but without it the linking of genotypes to phenotypes is significantly hampered. Therefore, there is a need for an *in silico* database of validated well-recorded physiological properties.

The discovery of novel genes related to 1,3-PDO production offers possibilities for using these genes for genomic modification to improve further the production in other species. Multiple genes discovered in Chapter 4 are associated with B12-dependent and B12-independent glycerol dehydratase routes. So far, the B12-independent glycerol dehydratase from *C. butyricum* was used for genetic engineering of *E. coli*⁵². In other genomic engineering studies,

genes of B12-dependent glycerol dehydratase and 1,3-PDO dehydrogenase from *K. pneumoniae* were inserted in *E. coli*²⁷⁹. In this study, the identified accessory genes related with 1,3-PDO production can be obtained from genome accessible producers and applied for bio-engineering. Additionally, we observed that multiple accessory genes in the same operon were related to a higher production of 1,3-PDO. *C. magnum* contains a large operon for the 1,3-PDO production trait and it was able to produce high concentrations of this compound. *C. magnum*'s production yield of 1,3-PDO was 0.56 (mol_{1,3-PDO produced} / mol_{glycerol consumed}) which was comparable to *T. pasteurii* (0.60-0.70). Therefore, the genes of such large operons should be further studied and possibly be considered for genome insertion in other species.

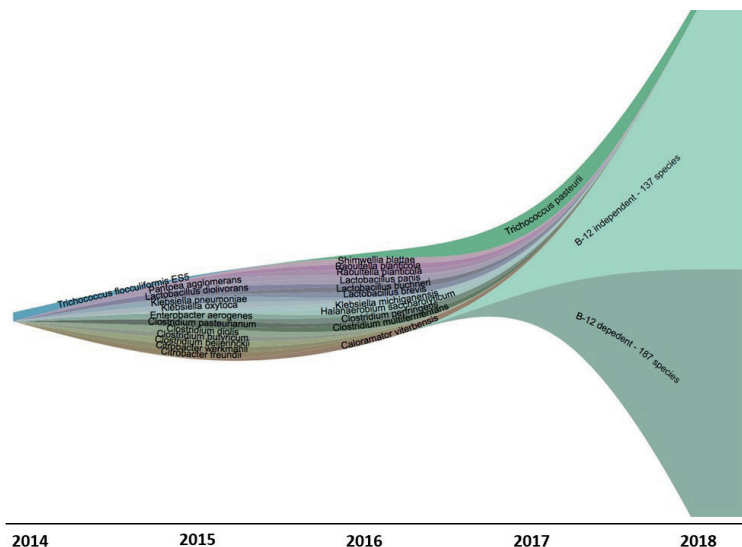


Figure 8.2 | Visualization of known species with the trait of 1,3-PDO production over the four years of this PhD. The different colored lines introduce each species and they are extended through the graph as every bacterium's metabolic knowledge was a corner stone in our study. In 2014, we investigated *Trichococcus floucoliformis* strain ES5, a 1,3-PDO producer. From 2015 and onwards, we started to investigate more strains with the information from the metabolic trait of ES5. In 2017, we had our breakthrough with *Trichococcus* species where we identified 17 genes in an operon structure in *T. pasteurii* (Chapter 3). Finally, in 2018, we studied all the bacterial genomes for this metabolic trait. We managed to identify all 1,3-PDO natural producers (for which genome is available) and added a high number of genes associated with the metabolic trait.

Improving production of metabolites

Addressing the 1,3-PDO production with big data mining, interesting knowledge of the trait was obtained. Therefore, we considered conducting additional omics and physiological studies on individual species to further profile their metabolic trait. Previous studies focused on different fermentation approaches for improving the production of 1,3-PDO in bacteria^{56,57,111,212}. In general, the production of 1,3-PDO in microbes serves as an electron sink in the oxidation of glycerol to acetate and reduces three NAD⁺ to three NADH, which can be utilized in the reductive pathway of 1,3-PDO production. A theoretical equation for fermentation is:



The maximum theoretical yield of produced 1,3-PDO is 75 %. In Chapter 5, we used a combination of proteomic and physiological trials to investigate the oxidative and reductive metabolism of glycerol fermentation in *T. pasteurii* and compare the production yield with the theoretical one. The physiological trials provided information about the influence of variables such as yeast extract, glycerol concentration and vitamins on 1,3-PDO production. At minimal concentrations of glycerol and yeast extract (5 g/L and 1 g/L respectively) and vitamin B12 supply, the maximum yield of 1,3-PDO production was observed in *T. pasteurii* and it was 70 %, which is a bit lower than the theoretically calculated yield. Interestingly, with high concentrations of glycerol, an accumulation of 3 hydroxypropanaldehyde (3-HPA) was observed. This was previously also reported for other bacteria and especially in LAB species²⁰². 3-HPA is a toxic compound that causes cell death, and thus does not favor 1,3-PDO production. Studying the kinetics and expression regulation of glycerol dehydratase and 1,3-PDO dehydrogenase²⁸⁰, 3-HPA may be regulated resulting in more 1,3-PDO produced. Additionally, 3-HPA can be converted to 3-hydroxypropionic acid (3-HP) by aldehyde dehydrogenase²⁸¹. Biotechnological production of 3-HP is of interest because of its commercial value²⁸¹. The pathway of 3-HPA to 3-HP is lacking in *T. pasteurii*, thus 3-HP is not produced. However, a study conducted genome editing for including 3-HP pathway in *K. pneumoniae* and led to the production of 3-HP and 1,3-PDO with a cumulative yield of 81 %²⁸².

The methodology used in this thesis for prospecting by 1,3-PDO producers can be extrapolated to gain insights into other bioconversions. Additionally, pathways to produce valuable compounds may even be detected in metagenomics samples of large microbial communities. As new metabolites are constantly being identified in microbial species, genome mining methods are implemented in several genera. The application of omics studies and physiological trials on *Actinobacteria* species exploited 48,780 bacterial genomes for antitumor drugs. Forty-nine potential producers of leinamycin-type compounds were identified²⁸³. From the perspective of this research and our study, I conclude that physiological trials and the application of omics studies expand the possibilities of bioprospecting microbial processes.

Understanding cold adaptation mechanisms

Due to extreme environmental conditions, psychrophilic and psychrotolerant microorganisms, need to adapt and obtain a versatile physiology²⁸⁴. *T. patagoniensis* was previously isolated from Patagonia as a microbe adapted to anoxic frozen conditions. In this thesis, omics studies were applied to investigate the psychrotolerance phenotype in *T. patagoniensis* (Chapter 6). Previously, EPS formation by *T. patagoniensis* was observed to cover the cells creating a cryoprotectant layer⁴⁹. In this study, we identified that the EPS formed by *T. patagoniensis* is inulin. *T. patagoniensis* is the first psychrotolerant bacterium able to produce inulin. The inulin synthesis was correlated before with the Wzy-dependent pathway and the EPS assembly genes²²⁷. In this thesis, we were able to verify the existence of the pathway in *T. patagoniensis* and its contribution to inulin production by omics studies. The discovery of inulin in a psychrotolerant species signifies the importance of studying extremophiles and their peculiar metabolism at low temperatures.

Inulin is a highly valuable compound due to its description as prebiotic and its medical usage^{236,285,286}. Recently, inulin was found to improve endothelial dysfunction, which can lead to the prevention of metabolic disorder-related cardiovascular diseases²⁸⁷. Industrial production of inulin is usually based on plants such as chicory, dahlia and Jerusalem artichoke^{229,235}. Previously, the production of inulin by bacteria was observed in *Lactobacillus* species: *Lactobacillus reuteri* strain 121, *L.*

johnsonii, *L. gasseri* strains 20077, 20243 and 20604, and other bacteria such as *Leuconostoc citreum* CW28²²⁹. Vegetal inulin is the choice for commercial production and it has a molecular weight (MW) of thousands. In contrast, the MW of microbial inulin is significant higher, which increases its compound value. *L. citreum* CW28 produced an inulin with a MW of up to 1,600,000 Da^{288,289}. Bacterial inulin production has not been fully studied, but the high MW and the fast growth of bacteria may be advantageous for the industrial production of inulin. For *T. patagoniensis* samples, inulin was also not precisely quantified, but it was approximately 10 % of the dry biomass. The robustness of *T. patagoniensis* and the possible high production of EPS can be added values for the bio-refineries of inulin.

In my opinion, microbial life in cryosphere have evolved robust adaptation mechanisms which are reflected in their metabolism. Obviously, the ability to grow at low temperatures will be reflected in the proteome of the psychrophilic/psychrotolerant species. Proteins are dynamic structural forms that react with their environment²⁹⁰. The flexible structure of proteins has to confront the freezing of the water to maintain flexibility⁵⁸. A preliminary analysis showed that the ratio of proline/arginine was high in the proteome of *T. patagoniensis*. Other proteome adaptations at low temperatures, such as the high content of hydrophobicity is difficult to assess currently *in silico*. In my opinion, a big mining data approach using all available psychrophilic/psychrotolerant species may lead to the acquisition of significant knowledge regarding their proteome composition and cold adaptation mechanisms.

The next step

The exploitation of metabolic traits and phenotypes of microbes can contribute to biotechnological prospecting. Novel compounds can be identified, and the production will be ascertained. Every environment is a place for discovering microbes, but I will recommend seeking microbes in ecosystems with extreme conditions, where the need to survive demands multiple adaptation mechanisms. The identification of inulin in *Trichococcus* was an interesting outcome and may turn out to be important for biotechnology. Other psychrotolerant species should be investigated for the ability to produce EPS. I ended my

PhD thesis with a chapter describing a novel bacterium (Chapter 7) isolated from a psychrophilic environment. The novel bacterium was named *T. shcherbakoviae* and this was the second *Trichococcus* species observed to produce an EPS substance as observed in SEM. The analysis and description of the EPS would be a future step and maybe *T. shcherbakoviae* will be another inulin producing psychrotolerant species.

In general, genomic information should be implemented in combination with 16S rRNA gene sequence analysis for a precise understanding and description of novel bacteria. Phylogenomics based on complete genome information can infer further phylogenetic interaction of the species. The genome sequencing should be a standardized application regarding novel bacteria. Two major outcomes can improve the current methods and those are an improved delineation of prokaryotes and the linking genotypes to specific phenotypes. The latter one can guide the physiological trials for the species description. Recently, overall genome related index (OGRI) values have been defined for a more standardized procedure in the use of genomic information especially for taxonomical purposes²⁵⁹. The addition of other omics studies such as transcriptomics and metabolomics may provide further information regarding the pathways associated with the species phenotypes. In Chapter 4, we annotated all available genome sequences with SAPP and stored them as a database. This database has significant value as each genome is annotated by the same pipeline and thus the genomic analysis is not affected by annotation mismatches. If the database were to become publicly available, it would contribute to an easier mining of big data for metabolic traits. Additionally, if the database were coupled to a well-recorded metadata from physiological properties, the genotypic and phenotypic information would be easier to link. With simple queries, access to vast information regarding physiology and genomic evidence would be obtained, increasing the knowledge for the metabolic traits of interest.

In conclusion, this thesis explores the power of omics studies and identifies possible ways for the bioprospecting of microbial processes. The methods presented here can be considered as a platform that future studies can use and adjust for an accurate and precise phenotypic and phylogenetic definition of remarkable metabolic traits.

REFERENCES

- 1 Akondi, K. B. & Lakshmi, V. V. Emerging trends in genomic approaches for microbial bioprospecting. *Omics* **17**, 61-70, doi:10.1089/omi.2012.0082 (2013).
- 2 Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA* **113**, 5970-5975, doi:10.1073/pnas.1521291113 (2016).
- 3 Liu, X. *et al.* Bioprospecting microbial natural product libraries from the marine environment for drug discovery. *J Antibiot (Tokyo)* **63**, 415-422, doi:10.1038/ja.2010.56 (2010).
- 4 Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* **7**, 13219, doi:10.1038/ncomms13219 (2016).
- 5 Van Trappen, S., Vandecandelaere, I., Mergaert, J. & Swings, J. *Flavobacterium degerlachei* sp. nov., *Flavobacterium frigoris* sp. nov. and *Flavobacterium micromati* sp. nov., novel psychrophilic bacteria isolated from microbial mats in Antarctic lakes. *Int J Syst Evol Microbiol* **54**, 85-92, doi:10.1099/ijs.0.02857-0 (2004).
- 6 Panosyan, H., Di Donato, P., Poli, A. & Nicolaus, B. Production and characterization of exopolysaccharides by *Geobacillus thermodenitrificans* ArzA-6 and *Geobacillus toebii* ArzA-8 strains isolated from an Armenian geothermal spring. *Extremophil*, doi:10.1007/s00792-018-1032-9 (2018).
- 7 Sanchez-Andrea, I., Stams, A. J., Hedrich, S., Nancucheo, I. & Johnson, D. B. *Desulfosporosinus acididurans* sp. nov.: an acidophilic sulfate-reducing bacterium isolated from acidic sediments. *Extremophil* **19**, 39-47, doi:10.1007/s00792-014-0701-6 (2015).
- 8 Pollock, J. *et al.* Alkaline iron(III) reduction by a novel alkaliphilic, halotolerant, *Bacillus* sp. isolated from salt flat sediments of Soap Lake. *Appl Microbiol Biotechnol* **77**, 927-934, doi:10.1007/s00253-007-1220-5 (2007).
- 9 Yarza, P. *et al.* Uniting the classification of cultured and uncultured

- tured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**, 635-645, doi:10.1038/nrmicro3330 (2014).
- 10 Beye, M., Fahsi, N., Raoult, D. & Fournier, P. E. Careful use of 16S rRNA gene sequence similarity values for the identification of *Mycobacterium* species. *New Microbes New Infect* **22**, 24-29, doi:10.1016/j.nmni.2017.12.009 (2018).
- 11 Meier-Kolthoff, J. P., Goker, M., Sproer, C. & Klenk, H. P. When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol* **195**, 413-418, doi:10.1007/s00203-013-0888-4 (2013).
- 12 Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**, 2567-2572, doi:10.1073/pnas.0409727102 (2005).
- 13 Kim, M., Oh, H. S., Park, S. C. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* **64**, 346-351, doi:10.1099/ijs.0.059774-0 (2014).
- 14 Scortichini, M., Marcelletti, S., Ferrante, P. & Firrao, G. A Genomic redefinition of *Pseudomonas avellanae* species. *PLoS One* **8**, e75794, doi:10.1371/journal.pone.0075794 (2013).
- 15 Zhang, W. et al. Whole-genome sequence comparison as a method for improving bacterial species definition. *J Gen Appl Microbiol* **60**, 75-78 (2014).
- 16 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
- 17 Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform* **14**, 60, doi:10.1186/1471-2105-14-60 (2013).
- 18 Mukherjee, S. et al. 1,003 reference genomes of bacterial

- and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* **35**, 676-683, doi:10.1038/nbt.3886 (2017).
- 19 Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-230, doi:10.1093/nar/gkt1223 (2014).
 - 20 Lee, B. & Lee, D. Protein comparison at the domain architecture level. *BMC Bioinformatics* **10 Suppl 15**, S5, doi:10.1186/1471-2105-10-S15-S5 (2009).
 - 21 Finn, R. D. Bioprospecting - Why is it so unrewarding? *Biodivers Conserv* **12**, 207-216 (2003).
 - 22 Zotchev, S. B., Sekurova, O. N. & Katz, L. Genome-based bioprospecting of microbes for new therapeutics. *Curr Opin Biotechnol* **23**, 941-947, doi:10.1016/j.copbio.2012.04.002 (2012).
 - 23 Elshahed, M. S. Microbiological aspects of biofuel production: Current status and future directions. *J Adv Res* **1**, 103-111 (2010).
 - 24 Koskinen, P. E. P. *et al.* Bioprospecting thermophilic microorganisms from icelandic hot springs for hydrogen and ethanol production. *Energy Fuels* **22**, 134-140 (2008).
 - 25 Verma, D., Singla, A., Lal, B. & Sarma, P. M. Conversion of biomass-generated syngas into next-generation liquid transport fuels through microbial intervention: Potential and current status. *Cur Science* **110**, 329-336 (2016).
 - 26 Guan, N. *et al.* Comparative genomics and transcriptomics analysis-guided metabolic engineering of *Propionibacterium acidipropionici* for improved propionic acid production. *Biotechnol Bioeng* **115**, 483-494, doi:10.1002/bit.26478 (2018).
 - 27 Leao, T. *et al.* Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus Moorea. *Proc Natl Acad Sci USA* **114**, 3198-3203, doi:10.1073/pnas.1618556114 (2017).
 - 28 Vilanova, C. & Porcar, M. Are multi-omics enough? *Nat Microbiol*

- 1, 16101, doi:10.1038/nmicrobiol.2016.101 (2016).
- 29 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**, 1621-1624, doi:10.1038/ismej.2012.8 (2012).
- 30 Rhoads, A. & Au, K. F. PacBio Sequencing and its applications. *Genom Proteom Bioinform* **13**, 278-289, doi:10.1016/j.gpb.2015.08.002 (2015).
- 31 Land, M. L. *et al.* Quality scores for 32,000 genomes. *Stand Genomic Sci* **9**, 20, doi:10.1186/1944-3277-9-20 (2014).
- 32 Land, M. *et al.* Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics* **15**, 141-161, doi:10.1007/s10142-015-0433-4 (2015).
- 33 Bennett, G. M., Abba, S., Kube, M. & Marzachi, C. Complete genome sequences of the obligate symbionts "*Candidatus Sulcia muelleri*" and "*Ca. Nasuia deltocephalinicola*" from the pestiferous leafhopper *Macrostelus quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announc* **4**, doi:10.1128/genomeA.01604-15 (2016).
- 34 Prakash, O., Shouche, Y., Jangid, K. & Kostka, J. E. Microbial cultivation and the role of microbial resource centers in the omics era. *Appl Microbiol Biotechnol* **97**, 51-62, doi:10.1007/s00253-012-4533-y (2013).
- 35 Vester, J. K., Glaring, M. A. & Stougaard, P. Improved cultivation and metagenomics as new tools for bioprospecting in cold environments. *Extremophiles* **19**, 17-29, doi:10.1007/s00792-014-0704-3 (2015).
- 36 Muller, C. A., Obermeier, M. M. & Berg, G. Bioprospecting plant-associated microbiomes. *J Biotechnol* **235**, 171-180, doi:10.1016/j.jbiotec.2016.03.033 (2016).
- 37 Salgado, H., Moreno-Hagelsieb, G., Smith, T. F. & Collado-Vides, J. Operons in *Escherichia coli*: genomic analyses and predictions. *Proc Natl Acad Sci U S A* **97**, 6652-6657, doi:10.1073/pnas.110147297 (2000).

- 38 Gil, A. I. *et al.* Fecal contamination of food, water, hands, and kitchen utensils at the household level in rural areas of Peru. *J Environ Health* **76**, 102-106 (2014).
- 39 Aziz, R. K. *et al.* The RAST Server: rapid annotations using sub-systems technology. *BMC Genomics* **9**, 75, doi:10.1186/1471-2164-9-75 (2008).
- 40 Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1**, 127-136, doi:10.1038/nrmicro751 (2003).
- 41 Karpievitch, Y. V., Polpitiya, A. D., Anderson, G. A., Smith, R. D. & Dabney, A. R. Liquid chromatography mass spectrometry-based proteomics: biological and technological aspects. *Ann Appl Stat* **4**, 1797-1823, doi:10.1214/10-AOAS341 (2010).
- 42 Cho, W. C. Proteomics technologies and challenges. *Genom Proteom Bioinform* **5**, 77-85, doi:10.1016/S1672-0229(07)60018-7 (2007).
- 43 Armengaud, J. Next-generation proteomics faces new challenges in environmental biotechnology. *Curr Opin Biotechnol* **38**, 174-182, doi:10.1016/j.copbio.2016.02.025 (2016).
- 44 D'Lima, N. G. *et al.* Comparative proteomics enables identification of nonannotated cold shock proteins in *E. coli*. *J Proteome Res* **16**, 3722-3731, doi:10.1021/acs.jproteome.7b00419 (2017).
- 45 Sorek, R. & Cossart, P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**, 9-16, doi:10.1038/nrg2695 (2010).
- 46 Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261, doi:10.1038/nbt0303-255 (2003).
- 47 Mashego, M. R. *et al.* Microbial metabolomics: past, present and future methodologies. *Biotechnol Lett* **29**, 1-16, doi:10.1007/s10529-006-9218-0 (2007).
- 48 van Gelder, A. H., Aydin, R., Alves, M. M. & Stams, A. J.

- 1,3-Propanediol production from glycerol by a newly isolated *Trichococcus* strain. *Microb Biotechnol* **5**, 573-578, doi:10.1111/j.1751-7915.2011.00318.x (2012).
- 49 Pikuta, E. V. *et al.* *Trichococcus patagoniensis* sp. nov., a facultative anaerobe that grows at -5 degrees C, isolated from penguin guano in Chilean Patagonia. *Int J Syst Evol Microbiol* **56**, 2055-2062, doi:10.1099/ij.s.0.64225-0 (2006).
 - 50 Pikuta, E. V. & Hoover, R. B. in *Lactic Acid Bacteria: Biodivers Tax* (John Wiley & Sons, Ltd, 2014).
 - 51 da Silva, G. P., Mack, M. & Contiero, J. Glycerol: a promising and abundant carbon source for industrial microbiology. *Biotechnol Adv* **27**, 30-39, doi:10.1016/j.biotechadv.2008.07.006 (2009).
 - 52 Saxena, R. K., Anand, P., Saran, S. & Isar, J. Microbial production of 1,3-propanediol: Recent developments and emerging opportunities. *Biotechnol Adv* **27**, 895-913, doi:10.1016/j.biotechadv.2009.07.003 (2009).
 - 53 Chung, T. C., Axelsson, L., Lindgren, S. E. & Dobrogosz, W. J. In vitro studies on reuterin synthesis by *Lactobacillus reuteri* in vitro studies on reuterin synthesis by lactobacillus. *Microb Ecol Health Dis* **2**, 137-144 (1989).
 - 54 Jiang, W., Wang, S., Wang, Y. & Fang, B. Key enzymes catalyzing glycerol to 1,3-propanediol. *Biotechnol Biofuels* **9**, 57, doi:10.1186/s13068-016-0473-6 (2016).
 - 55 Taconi Katherine, A., Venkataramanan Keerthi, P. & Johnson Duane, T. Growth and solvent production by *Clostridium pasteurianum* ATCC® 6013™ utilizing biodiesel-derived crude glycerol as the sole carbon source. *Environ Prog Sustain Energy* **28**, 100-110, doi:10.1002/ep.10350 (2009).
 - 56 Zhang, G., Ma, B., Xu, X., Li, C. & Wang, L. Fast conversion of glycerol to 1,3-propanediol by a new strain of *Klebsiella pneumoniae*. *Biochem Eng J* **37**, 256-260, doi:10.1016/j.bej.2007.05.003 (2007).
 - 57 Barbirato, F., Himmi, E. H., Conte, T. & Bories, A. 1,3-propanediol

- production by fermentation: An interesting way to valorize glycerin from the ester and ethanol industries. *Indu C Prod* **7**, 281-289, doi:10.1016/S0926-6690(97)00059-9 (1998).
- 58 De Maayer, P., Anderson, D., Cary, C. & Cowan, D. A. Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* **15**, 508-517, doi:10.1002/embr.201338170 (2014).
 - 59 Del Campo, M. et al. Unwinding by local strand separation is critical for the function of DEAD-box proteins as RNA chaperones. *J Mol Biol* **389**, 674-693, doi:10.1016/j.jmb.2009.04.043 (2009).
 - 60 Pyle, A. M. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* **37**, 317-336, doi:10.1146/annurev.biophys.37.032807.125908 (2008).
 - 61 Chintalapati, S., Kiran, M. D. & Shivaji, S. Role of membrane lipid fatty acids in cold adaptation. *Cell Mol Biol (Noisy-le-grand)* **50**, 631-642 (2004).
 - 62 Deming, J. W. Psychrophiles and polar regions. *Curr Opin Microbiol* **5**, 301-309 (2002).
 - 63 Nichols, C. A., Guezennec, J. & Bowman, J. P. Bacterial exopolysaccharides from extreme marine environments with special consideration of the southern ocean, sea ice, and deep-sea hydrothermal vents: a review. *Mar Biotechnol (NY)* **7**, 253-271, doi:10.1007/s10126-004-5118-2 (2005).
 - 64 Cid, F. P. et al. Properties and biotechnological applications of ice-binding proteins in bacteria. *FEMS Microbiol Lett* **363**, doi:10.1093/femsle/fnw099 (2016).
 - 65 Marx, J. G., Carpenter, S. D. & Deming, J. W. Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. *Can J Microbiol* **55**, 63-72, doi:10.1139/W08-130 (2009).
 - 66 Casillo, A. et al. Structure-activity relationship of the exopolysaccharide from a psychrophilic bacterium: A strategy for

cryoprotection. *Carbohydr Polym* **156**, 364-371, doi:10.1016/j.carbpol.2016.09.037 (2017).

- 67 Liu, J. R. et al. Emended description of the genus *Trichococcus*, description of *Trichococcus collinsii* sp. nov., and reclassification of *Lactosphaera pasteurii* as *Trichococcus pasteurii* comb. nov. and of *Ruminococcus palustris* as *Trichococcus palustris* comb. nov. in the low-G+C gram-positive bacteria. *Int J Syst Evol Microbiol* **52**, 1113-1126, doi:10.1099/00207713-52-4-1113 (2002).
- 68 Dai, Y. M. et al. Characterization of *Trichococcus paludicola* sp. nov. and *Trichococcus alkaliphilus* sp. nov., isolated from a high-elevation wetland, by phenotypic and genomic analyses. *Int J Syst Evol Microbiol* **68**, 99-105, doi:10.1099/ijsem.0.002464 (2018).
- 69 American Society for Microbiology, Bergey, D. H. & Breed, R. S. *Bergey's manual of determinative bacteriology*. (Williams & Wilkins Co., 1957).
- 70 Dai, Y. M. et al. Corrigendum: Characterization of *Trichococcus paludicola* sp. nov. and *Trichococcus alkaliphilus* sp. nov., isolated from a high-elevation wetland, by phenotypic and genomic analyses. *Int J Syst Evol Microbiol* **68**, 2115, doi:10.1099/ijsem.0.002802 (2018).
- 71 Stams, A. J., Huisman, J., Garcia Encina, P. A. & Muyzer, G. Citric acid wastewater as electron donor for biological sulfate reduction. *Appl Microbiol Biotechnol* **83**, 957-963, doi:10.1007/s00253-009-1995-7 (2009).
- 72 Eiteman, M. A. & Ramalingam, S. Microbial production of lactic acid. *Biotechnol Lett* **37**, 955-972, doi:10.1007/s10529-015-1769-5 (2015).
- 73 Scheff, G., Salcher, O. & Lingens, F. *Trichococcus flocculiformis* gen. nov. sp. nov. A new gram-positive filamentous bacterium isolated from bulking sludge. *Appl Microbiol Biotechnol* **19**, 114-119 (1984).

- 74 Schink, B. Fermentation of tartrate enantiomers by anaerobic bacteria, and description of two new species of strict anaerobes, *Ruminococcus pasteurii* and *Ilyobacter tartaricus*. *Arch Microbiol* **139**, 409-414 (1984).
- 75 Zhilina, T. N., Kotsyurbenko, O. R., Osipov, G. A., Kostrikina, N. A. & A., Z. G. *Ruminococcus palustris* sp. nov.: a psychoractive anaerobic organism from a swamp. *Microbiol* **64**, 574-579 (1995).
- 76 Liu, Y. *et al.* Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* **5**, 14082, doi:10.1038/srep14082 (2015).
- 77 Schleifer, K. H. Classification of bacteria and archaea: past, present and future. *Syst Appl Microbiol* **32**, 533-542, doi:10.1016/j.syapm.2009.09.002 (2009).
- 78 Klenk, H. P. & Goker, M. En route to a genome-based classification of archaea and bacteria? *Syst Appl Microbiol* **33**, 175-182, doi:10.1016/j.syapm.2010.03.003 (2010).
- 79 Meier-Kolthoff, J. P., Klenk, H. P. & Goker, M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* **64**, 352-356, doi:10.1099/ij.s.0.056994-0 (2014).
- 80 Stams, A. J., Van Dijk, J. B., Dijkema, C. & Plugge, C. M. Growth of syntrophic propionate-oxidizing bacteria with fumarate in the absence of methanogenic bacteria. *Appl Environ Microbiol* **59**, 1114-1119 (1993).
- 81 Kyrpides, N. C. *et al.* Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol* **12**, e1001920, doi:10.1371/journal.pbio.1001920 (2014).
- 82 Boisvert, S., Raymond, F., Godzaridis, E., Laviolette, F. & Corbeil, J. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* **13**, R122, doi:10.1186/gb-2012-13-12-r122 (2012).
- 83 Gao, S., Sung, W. K. & Nagarajan, N. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-

- end sequences. *J Comput Biol* **18**, 1681-1691, doi:10.1089/cmb.2011.0170 (2011).
- 84 Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Res* **9**, 868-877 (1999).
 - 85 Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinform* **30**, 31-37, doi:10.1093/bioinformatics/btt310 (2014).
 - 86 Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* **108**, 1513-1518, doi:10.1073/pnas.1017351108 (2011).
 - 87 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).
 - 88 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinform* **27**, 578-579, doi:10.1093/bioinformatics/btq683 (2011).
 - 89 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).
 - 90 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).
 - 91 Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964 (1997).
 - 92 Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100-3108, doi:10.1093/nar/gkm160 (2007).
 - 93 Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinform* **23**, 1282-1288, doi:10.1093/bioinformatics/btm098 (2007).

- 94 Consortium., U. Activities at the universal protein resource (UniProt). *Nucleic Acids Research* **42**, 191-198 (2014).
- 95 Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genom* **12**, 402, doi:10.1186/1471-2164-12-402 (2011).
- 96 Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-185, doi:10.1093/nar/gkm321 (2007).
- 97 Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363-1371, doi:10.1093/nar/gkh293 (2004).
- 98 Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).
- 99 Yarza, P. et al. The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**, 241-250, doi:10.1016/j.syapm.2008.07.001 (2008).
- 100 Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
- 101 Kuykendall, L. D., Roy, M. A., O'Neill, J. J. & Devine, T. E. Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. *J Sys Evolu Microbiol* **38**, (1988).
- 102 Miller, L. T. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**, 584-586 (1982).
- 103 Petit, E. et al. Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by *Clostridium phytofermentans*. *PLoS One* **8**, e54337, doi:10.1371/journal.

pone.0054337 (2013).

- 104 Mulligan, C., Fischer, M. & Thomas, G. H. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35**, 68-86, doi:10.1111/j.1574-6976.2010.00236.x (2011).
- 105 Okochi, M., Kurimoto, M., Shimizu, K. & Honda, H. Increase of organic solvent tolerance by overexpression of manXYZ in *Escherichia coli*. *Appl Microbiol Biotechnol* **73**, 1394-1399, doi:10.1007/s00253-006-0624-y (2007).
- 106 Wang, W., Sun, J., Hartlep, M., Deckwer, W. D. & Zeng, A. P. Combined use of proteomic analysis and enzyme activity assays for metabolic pathway analysis of glycerol fermentation by *Klebsiella pneumoniae*. *Biotechnol Bioeng* **83**, 525-536, doi:10.1002/bit.10701 (2003).
- 107 Zhang, Y. M. & Rock, C. O. Membrane lipid homeostasis in bacteria. *Nat Rev Microbiol* **6**, 222-233, doi:10.1038/nrmicro1839 (2008).
- 108 Strepis, N. et al. Description of *Trichococcus ilyis* sp. nov. by combined physiological and in silico genome hybridization analyses. *Int J Syst Evol Microbiol* **66**, 3957-3963, doi:10.1099/ijsem.0.001294 (2016).
- 109 Koehorst, J. J. et al. Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. *Sci Rep* **6**, 38699, doi:10.1038/srep38699 (2016).
- 110 Chatzifragkou, A. et al. Production of 1,3-propanediol by *Clostridium butyricum* growing on biodiesel-derived crude glycerol through a non-sterilized fermentation process. *Appl Microbiol Biotechnol* **91**, 101-112, doi:10.1007/s00253-011-3247-x (2011).
- 111 Cheng, K. K., Liu, D. H., Sun, Y. & Liu, W. B. 1,3-Propanediol production by *Klebsiella pneumoniae* under different aeration strategies. *Biotechnol Lett* **26**, 911-915 (2004).
- 112 Hatti-Kaul, R. & Mattiasson, B. Anaerobes in industrial- and environmental biotechnology. *Adv Biochem Eng Biotechnol* **156**,

- 1-33, doi:10.1007/10_2016_10 (2016).
- 113 Maervoet, V. E. *et al.* 1,3-propanediol production with *Citrobacter werkmanii* DSM17579: effect of a dhaD knock-out. *Microb Cell Fact* **13**, 70, doi:10.1186/1475-2859-13-70 (2014).
 - 114 Franzmann, P. D., Hopfl, P., Weiss, N. & Tindall, B. J. Psychrotrophic, lactic acid-producing bacteria from anoxic waters in Ace Lake, Antarctica; *Carnobacterium funditum* sp. nov. and *Carnobacterium alterfunditum* sp. nov. *Arch Microbiol* **156**, 255-262 (1991).
 - 115 Pikuta, E. V. *et al.* *Carnobacterium pleistocenium* sp. nov., a novel psychrotolerant, facultative anaerobe isolated from permafrost of the Fox Tunnel in Alaska. *Int J Syst Evol Microbiol* **55**, 473-478, doi:10.1099/ijs.0.63384-0 (2005).
 - 116 Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-65, doi:10.1093/nar/gkl842 (2007).
 - 117 Koehorst, J. J. *et al.* SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. *Bioinform* **34**, 1401-1403, doi:10.1093/bioinformatics/btx767 (2018).
 - 118 Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755-763 (1998).
 - 119 Sonnhammer, E. L., von Heijne, G. & Krogh, A. A Hidden Markov Model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182 (1998).
 - 120 Team, R. D. C. R: A language and environment for statistical computing. R foundation for statistical computing. (2008).
 - 121 DARwin Software (2006).
 - 122 Meier-Kolthoff, J. P. *et al.* Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy.

- Stand Genomic Sci* **9**, 2, doi:10.1186/1944-3277-9-2 (2014).
- 123 Goker, M., Garcia-Blazquez, G., Voglmayr, H., Telleria, M. T. & Martin, M. P. Molecular taxonomy of phytopathogenic fungi: a case study in *Peronospora*. *PLoS One* **4**, e6319, doi:10.1371/journal.pone.0006319 (2009).
 - 124 Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Highly parallelized inference of large genome-based phylogenies. *Concurr Comput.* **26**, 1715–1729 (2014).
 - 125 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).
 - 126 Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* **32**, 2798–2800, doi:10.1093/molbev/msv150 (2015).
 - 127 Minkin, I., Pham, H., Starostina, E., Vyahhi, N. & Pham, S. C-Sibelia: an easy-to-use and highly accurate tool for bacterial genome comparison. *F1000Res* **2**, 258, doi:10.12688/f1000research.2-258.v1 (2013).
 - 128 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645, doi:10.1101/gr.092759.109 (2009).
 - 129 Liu, J. Z., Xu, W., Chistoserdov, A. & Bajpai, R. K. Glycerol dehydratases: biochemical structures, catalytic mechanisms, and industrial applications in 1,3-propanediol production by naturally occurring and genetically engineered bacterial strains. *Appl Biochem Biotechnol* **179**, 1073–1100, doi:10.1007/s12010-016-2051-6 (2016).
 - 130 Voegelé, R. T., Sweet, G. D. & Boos, W. Glycerol kinase of *Escherichia coli* is activated by interaction with the glycerol facilitator. *J Bacteriol* **175**, 1087–1094 (1993).
 - 131 Orgad, O., Oren, Y., Walker, S. L. & Herzberg, M. The role of alginate in *Pseudomonas aeruginosa* EPS adherence, viscoelastic properties and cell attachment. *Biofoul* **27**, 787–798, doi:10.108

0/08927014.2011.603145 (2011).

- 132 Kavanagh, K. L., Jornvall, H., Persson, B. & Oppermann, U.
Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* **65**, 3895-3906, doi:10.1007/s00018-008-8588-y (2008).
- 133 Liu, Y. *et al.* Hydroxysteroid dehydrogenase family proteins on lipid droplets through bacteria, *C. elegans*, and mammals. *Biochim Biophys Acta Mol Cell Biol Lipids* **1863**, 881-894, doi:10.1016/j.bbalip.2018.04.018 (2018).
- 134 Henrissat, B. & Bairoch, A. new families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **293 (Pt 3)**, 781-788 (1993).
- 135 Inacio, J. M. & de Sa-Nogueira, I. Characterization of abn2 (yxiA), encoding a *Bacillus subtilis* GH43 arabinanase, Abn2, and its role in arabino-polysaccharide degradation. *J Bacteriol* **190**, 4272-4280, doi:10.1128/JB.00162-08 (2008).
- 136 Wefers, D. *et al.* Enzymatic mechanism for arabinan degradation and transport in the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. *Appl Environ Microbiol* **83**, doi:10.1128/AEM.00794-17 (2017).
- 137 Kim, M. S., Roh, S. W., Nam, Y. D., Yoon, J. H. & Bae, J. W. *Carnobacterium jeotgali* sp. nov., isolated from a Korean traditional fermented food. *Int J Syst Evol Microbiol* **59**, 3168-3171, doi:10.1099/ijs.0.010116-0 (2009).
- 138 Joborn, A., Dorsch, M., Olsson, J. C., Westerdahl, A. & Kjelleberg, S. *Carnobacterium inhibens* sp. nov., isolated from the intestine of Atlantic salmon (*Salmo salar*). *Int J Syst Bacteriol* **49 Pt 4**, 1891-1898, doi:10.1099/00207713-49-4-1891 (1999).
- 139 Collins, M. D., Farrow, J. A. E., Phillips, B. A., Feresu, S. & Jones, D. Classification of *Lactobacillus divergens*, *Lactobacillus piscicola*, and some catalase-negative, asporogenous, rod-shaped bac-

- teria from poultry in a new genus, *Carnobacterium*. *I J Sys Evol Microbiol* **37**, 310-316 (1987).
- 140 Boch, J., Kempf, B., Schmid, R. & Bremer, E. Synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*: characterization of the gbsAB genes. *J Bacteriol* **178**, 5121-5129 (1996).
 - 141 Canovas, D. *et al.* Genes for the synthesis of the osmoprotectant glycine betaine from choline in the moderately halophilic bacterium *Halomonas elongata* DSM 3043, USA. *Microbiol* **146** (Pt 2), 455-463, doi:10.1099/00221287-146-2-455 (2000).
 - 142 Kempf, B. & Bremer, E. OpuA, an osmotically regulated binding protein-dependent transport system for the osmoprotectant glycine betaine in *Bacillus subtilis*. *J Biol Chem* **270**, 16701-16713 (1995).
 - 143 Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**, 181-186, doi:10.1038/nrmicro1793 (2008).
 - 144 Wesolowski, J. & Paumet, F. SNARE motif: a common motif used by pathogens to manipulate membrane fusion. *Virul* **1**, 319-324, doi:10.4161/viru.1.4.12195 (2010).
 - 145 Roberts, M. C. Update on acquired tetracycline resistance genes. *FEMS Microbiol Lett* **245**, 195-203, doi:10.1016/j.femsle.2005.02.034 (2005).
 - 146 Corr, S. C. *et al.* Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *Proc Natl Acad Sci USA* **104**, 7617-7621, doi:10.1073/pnas.0700440104 (2007).
 - 147 Palazzotto, E. & Weber, T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Curr Opin Microbiol* **45**, 109-116, doi:10.1016/j.mib.2018.03.004 (2018).
 - 148 Consortium., U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699, doi:10.1093/nar/gky092 (2018).

- 149 Pasek, S., Risler, J. L. & Brezellec, P. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **22**, 1418-1423, doi:10.1093/bioinformatics/btl135 (2006).
- 150 van Dam, J. C. J., Koehorst, J. J., Vik, J. O., Schaap, P. J. & Suarez-Diez, M. Interoperable genome annotation with GBOL, an extendable infrastructure for functional data mining. *bioRxiv* (2017).
- 151 Koehorst, J. J., Saccenti, E., Schaap, P. J., Martins Dos Santos, V. A. P. & Suarez-Diez, M. Protein domain architectures provide a fast, efficient and scalable alternative to sequence-based methods for comparative functional genomics. *F1000Res* **5**, 1987, doi:10.12688/f1000research.9416.3 (2016).
- 152 Gonzalez-Pajuelo, M., Meynial-Salles, I., Mendes, F., Soucaille, P. & Vasconcelos, I. Microbial conversion of glycerol to 1,3-propanediol: physiological comparison of a natural producer, *Clostridium butyricum* VPI 3266, and an engineered strain, *Clostridium acetobutylicum* DG1(pSPD5). *Appl Environ Microbiol* **72**, 96-101, doi:10.1128/AEM.72.1.96-101.2006 (2006).
- 153 Ristoski, P. & Paulheim, H. Semantic Web in data mining and knowledge discovery: A comprehensive survey. *J Web Sem* **36**, 1-22 (2016).
- 154 Sun, J., van den Heuvel, J., Soucaille, P., Qu, Y. & Zeng, A. P. Comparative genomic analysis of dha regulon and related genes for anaerobic glycerol metabolism in bacteria. *Biotechnol Prog* **19**, 263-272, doi:10.1021/bp025739m (2003).
- 155 Yang, G., Tian, J. & Li, J. Fermentation of 1,3-propanediol by a lactate deficient mutant of *Klebsiella oxytoca* under micro-aerobic conditions. *Appl Microbiol Biotechnol* **73**, 1017-1024, doi:10.1007/s00253-006-0563-7 (2007).
- 156 Barbirato, F., Himmi, E. H., Conte, T. & Bories, A. 1,3-propanediol production by fermentation: An interesting way to valorize glycerin from the ester and ethanol industries. *Industrial Crops*

- and Products **7**, 281-289 (1998).
- 157 Taconi, K. A., Venkataramanan, K. P. & Johnson, D. T. Growth and solvent production by *Clostridium pasteurianum* ATCC 6013 TM utilizing biodiesel-derived crude glycerol as the sole carbon source. *Environ Prog Sustain Energy* **28**, 100-110 (2009).
 - 158 Kivisto, A., Santala, V. & Karp, M. 1,3-Propanediol production and tolerance of a halophilic fermentative bacterium, *Halanaerobium saccharolyticum* subsp. *saccharolyticum*. *J Biotechnol* **158**, 242-247, doi:10.1016/j.jbiotec.2011.10.013 (2012).
 - 159 Otte, B., Grunwaldt, E., Mahmoud, O. & Jennewein, S. Genome shuffling in *Clostridium diolis* DSM 15410 for improved 1,3-propanediol production. *Appl Environ Microbiol* **75**, 7610-7616, doi:10.1128/AEM.01774-09 (2009).
 - 160 Gonzalez-Pajuelo, M., Andrade, J. C. & Vasconcelos, I. Production of 1,3-propanediol by *Clostridium butyricum* VPI 3266 using a synthetic medium and raw glycerol. *J Ind Microbiol Biotechnol* **31**, 442-446, doi:10.1007/s10295-004-0168-z (2004).
 - 161 Jarvis, G. N., Moore, E. R. & Thiele, J. H. Formate and ethanol are the major products of glycerol fermentation produced by a *Klebsiella planticola* strain isolated from red deer. *J Appl Microbiol* **83**, 166-174 (1997).
 - 162 Mu, Y., Teng, H., Zhang, D. J., Wang, W. & Xiu, Z. L. Microbial production of 1,3-propanediol by *Klebsiella pneumoniae* using crude glycerol from biodiesel preparations. *Biotechnol Lett* **28**, 1755-1759, doi:10.1007/s10529-006-9154-z (2006).
 - 163 Gungormusler, M., Gonen, C. & Azbar, N. Continuous production of 1,3-propanediol using raw glycerol with immobilized *Clostridium beijerinckii* NRRL B-593 in comparison to suspended culture. *Bioprocess Biosyst Eng* **34**, 727-733, doi:10.1007/s00449-011-0522-2 (2011).
 - 164 Seyfried, M., Lyon, D., Rainey, F. A. & Wiegel, J. *Caloramator vit-erbiensis* sp. nov., a novel thermophilic, glycerol-fermenting bacterium isolated from a hot spring in Italy. *Int J Syst Evol Microbiol*

- 52**, 1177-1184 (2002).
- 165 Wilkens, E., Ringel, A. K., Hortig, D., Willke, T. & Vorlop, K. D. High-level production of 1,3-propanediol from crude glycerol by *Clostridium butyricum* AKR102a. *Appl Microbiol Biotechnol* **93**, 1057-1063, doi:10.1007/s00253-011-3595-6 (2012).
 - 166 Asad ur, R., Wijesekara R.G, S., Nomura, N., Sato, S. & Matsumura, M. Pre-treatment and utilization of raw glycerol from sunflower oil biodiesel for growth and 1,3-propanediol production by *Clostridium butyricum*. *J Chem Technol Biotechnol* **83**, 1072-1080, doi:10.1002/jctb.1917 (2008).
 - 167 Petitdemange, E., Dürr, C., Abbad-Andaloussi, S. & Raval, G. Fermentation of raw glycerol to 1, 3-propanediol by new strains of *Clostridium butyricum*. *J Indu Microbiol Biotechnol* **15**, 498-502 (1995).
 - 168 Papanikolaou, S., Fick, M. & Aggelis, G. The effect of raw glycerol concentration on the production of 1,3-propanediol by *Clostridium butyricum*. *J Chem Technol Biotechnol* **79**, 1189-1196 (2004).
 - 169 Biebl, H., Menzel, K., Zeng, A. P. & Deckwer, W. D. Microbial production of 1,3-propanediol. *Appl Microbiol Biotechnol* **52**, 289-297 (1999).
 - 170 Jun, S. A. et al. Microbial fed-batch production of 1,3-propanediol using raw glycerol with suspended and immobilized *Klebsiella pneumoniae*. *Appl Biochem Biotechnol* **161**, 491-501, doi:10.1007/s12010-009-8839-x (2010).
 - 171 Xu, Y. Z. et al. Metabolism in 1,3-propanediol fed-batch fermentation by a D-lactate deficient mutant of *Klebsiella pneumoniae*. *Biotechnol Bioeng* **104**, 965-972, doi:10.1002/bit.22455 (2009).
 - 172 Zhang, G. L., Ma, B., Xu, X. L., Li, C. & Wang, L. Fast conversion of glycerol to 1,3-propanediol by a new strain of *Klebsiella pneumoniae*. *Biochem Engineer J* **37**, 256-260 (2007).
 - 173 Vivek, N., Pandey, A. & Binod, P. An efficient aqueous two phase systems using dual inorganic electrolytes to separate 1,3-pro-

- panediol from the fermented broth. *Bioresour Technol* **254**, 239-246, doi:10.1016/j.biortech.2018.01.076 (2018).
- 174 Schütz, H. & Radler, F. Anaerobic reduction of glycerol to propanediol-1.3 by *Lactobacillus brevis* and *Lactobacillus buchneri*. *System Appl Microbiol* **5**, 169-178 (1984).
 - 175 Kang, T. S., Korber, D. R. & Tanaka, T. Metabolic engineering of a glycerol-oxidative pathway in *Lactobacillus panis* PM1 for utilization of bioethanol thin stillage: potential to produce platform chemicals from glycerol. *Appl Environ Microbiol* **80**, 7631-7639, doi:10.1128/AEM.01454-14 (2014).
 - 176 Rodriguez, A., Wojtusik, M., Ripoll, V., Santos, V. E. & Garcia-Ochoa, F. 1,3-Propanediol production from glycerol with a novel biocatalyst *Shimwellia blattae* ATCC 33430: Operational conditions and kinetics in batch cultivations. *Bioresour Technol* **200**, 830-837, doi:10.1016/j.biortech.2015.10.061 (2016).
 - 177 Pradima, J. Review on enzymatic synthesis of value added products of glycerol, a by-product derived from biodiesel production. *Resource-Efficient Technol* **3**, 394-405 (2017).
 - 178 Leinonen, R. et al. The european nucleotide archive. *Nucleic Acids Res* **39**, D28-31, doi:10.1093/nar/gkq967 (2011).
 - 179 Sohngen, C. et al. BacDive--The Bacterial Diversity Metadatabase in 2016. *Nucleic Acids Res* **44**, D581-585, doi:10.1093/nar/gkv983 (2016).
 - 180 Schink, B. *Clostridium magnum* sp. nov., a non-autotrophic homoacetogenic bacterium. *Archiv Microbiol* **137**, 250-255 (1984).
 - 181 Balch, W. E., Schoberth, S., Tanner, R. S. & Wolfe, R. S. *Acetobacterium*, a New genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria. *Int J Syst Evol Microbiol* **27**, 355-361 (1977).
 - 182 Dawyndt, P. et al. Mining fatty acid databases for detection of novel compounds in aerobic bacteria. *J Microbiol Methods* **66**, 410-433, doi:10.1016/j.mimet.2006.01.008 (2006).

- 183 Jin, R. Z., Forage, R. G. & Lin, E. C. Glycerol kinase as a substitute for dihydroxyacetone kinase in a mutant of *Klebsiella pneumoniae*. *J Bacteriol* **152**, 1303-1307 (1982).
- 184 Silvester, N. et al. The European Nucleotide Archive in 2017. *Nucleic Acids Res* **46**, D36-D40, doi:10.1093/nar/gkx1125 (2018).
- 185 Raynaud, C., Sarcabal, P., Meynial-Salles, I., Croux, C. & Soucaille, P. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci USA* **100**, 5010-5015, doi:10.1073/pnas.0734105100 (2003).
- 186 Machielsen, R., Uria, A. R., Kengen, S. W. & van der Oost, J. Production and characterization of a thermostable alcohol dehydrogenase that belongs to the aldo-keto reductase superfamily. *Appl Environ Microbiol* **72**, 233-238, doi:10.1128/AEM.72.1.233-238.2006 (2006).
- 187 Knietsch, A., Bowien, S., Whited, G., Gottschalk, G. & Daniel, R. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* **69**, 3048-3060 (2003).
- 188 Zarzycki, J., Erbilgin, O. & Kerfeld, C. A. Bioinformatic characterization of glycol radical enzyme-associated bacterial microcompartments. *Appl Environ Microbiol* **81**, 8315-8329, doi:10.1128/AEM.02587-15 (2015).
- 189 Lee, S. M. et al. Enhancement of 1,3-propanediol production by expression of pyruvate decarboxylase and aldehyde dehydrogenase from *Zymomonas mobilis* in the acetolactate-synthase-deficient mutant of *Klebsiella pneumoniae*. *J Ind Microbiol Biotechnol* **41**, 1259-1266, doi:10.1007/s10295-014-1456-x (2014).
- 190 Matsakas, L., Hrůzová, K., Rova, U. & Christakopoulos, P. Biological production of 3-hydroxypropionic acid: an update on the current status. *Ferment* **4**, 13 (2018).

- 191 Artuso, A. Bioprospecting, benefit sharing, and biotechnological capacity building. *World Develop* **30**, 1355-1368 (2002).
- 192 Cheung, K. H., Prud'hommeaux, E., Wang, Y. & Stephens, S. Semantic web for health Care and life sciences: a review of the state of the art. *Brief Bioinform* **10**, 111-113, doi:10.1093/bib/bbp015 (2009).
- 193 Koutinas, A. A. *et al.* Valorization of industrial waste and by-product streams via fermentation for the production of chemicals and biopolymers. *Chem Soc Rev* **43**, 2587-2627, doi:10.1039/c3cs60293a (2014).
- 194 Johnson, E. A. & Lin, E. C. *Klebsiella pneumoniae* 1,3-propanediol:NAD⁺ oxidoreductase. *J Bacteriol* **169**, 2050-2054 (1987).
- 195 Biebl, H. & Sproer, C. Taxonomy of the glycerol fermenting clostridia and description of *Clostridium diolis* sp. nov. *Syst Appl Microbiol* **25**, 491-497, doi:10.1078/07232020260517616 (2002).
- 196 Cohen, I. & Altshuller, A. A new spectrophotometric method for determination of acrolein in combustion gases and in the atmosphere. *Anal Chem* **33**, 726-733 (1961).
- 197 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372, doi:10.1038/nbt.1511 (2008).
- 198 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**, 785-786, doi:10.1038/nmeth.1701 (2011).
- 199 Sauvageot, N., Gouffi, K., Laplace, J. M. & Auffray, Y. Glycerol metabolism in *Lactobacillus collinoides*: production of 3-hydroxypropionaldehyde, a precursor of acrolein. *Int J Food Microbiol* **55**, 167-170 (2000).
- 200 Pasteris, S. E. & Strasser de Saad, A. M. Sugar-glycerol cofermentations by *Lactobacillus hilgardii* isolated from wine. *J Agric Food*

Chem **57**, 3853-3858, doi:10.1021/jf803781k (2009).

- 201 Martin, R. *et al.* Characterization of a reuterin-producing *Lactobacillus coryniformis* strain isolated from a goat's milk cheese. *Int J Food Microbiol* **104**, 267-277, doi:10.1016/j.ijfood-micro.2005.03.007 (2005).
- 202 Bauer, R., du Toit, M. & Kossmann, J. Influence of environmental parameters on production of the acrolein precursor 3-hydroxypropionaldehyde by *Lactobacillus reuteri* DSMZ 20016 and its accumulation by wine lactobacilli. *Int J Food Microbiol* **137**, 28-31, doi:10.1016/j.ijfoodmicro.2009.10.012 (2010).
- 203 Gonzalez-Pajuelo, M., Andrade, J. C. & Vasconcelos, I. Production of 1,3-propanediol by *Clostridium butyricum* VPI 3266 in continuous cultures with high yield and productivity. *J Ind Microbiol Biotechnol* **32**, 391-396, doi:10.1007/s10295-005-0012-0 (2005).
- 204 Barbirato, F., Grivet, J. P., Soucaille, P. & Bories, A. 3-Hydroxypropionaldehyde, an inhibitory metabolite of glycerol fermentation to 1,3-propanediol by enterobacterial species. *Appl Environ Microbiol* **62**, 1448-1451 (1996).
- 205 Raj, S. M., Rathnasingh, C., Jo, J.-E. & Park, S. Production of 3-hydroxypropionic acid from glycerol by a novel recombinant *Escherichia coli* BL21 strain. *Process Biochem* **43**, 1440-1446, doi:/10.1016/j.procbio.2008.04.027 (2008).
- 206 Jornvall, H., von Bahr-Lindstrom, H., Jany, K. D., Ulmer, W. & Froschle, M. Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases. *FEBS Lett* **165**, 190-196 (1984).
- 207 Campbell, J. A., Davies, G. J., Bulone, V. V. & Henrissat, B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J* **329** (Pt 3), 719 (1998).
- 208 Schmitt, M. P. & Holmes, R. K. Iron-dependent regulation of diphtheria toxin and siderophore expression by the cloned

- Corynebacterium diphtheriae* repressor gene dtxR in *C. diphtheriae* C7 strains. *Infect Immun* **59**, 1899-1904 (1991).
- 209 Scotti, P. A. et al. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *EMBO J* **19**, 542-549, doi:10.1093/emboj/19.4.542 (2000).
- 210 Metsoviti, M. et al. Screening of bacterial strains capable of converting biodiesel-derived raw glycerol into 1,3-propanediol, 2,3-butanediol and ethanol. *Eng Life Sci* **12**, 57-68, doi:10.1002/elsc.201100058 (2011).
- 211 Biebl, H. Fermentation of glycerol by *Clostridium pasteurianum* – batch and continuous culture studies. *J Ind Microbiol Biotechnol* **27**, 18-26, doi:10.1038/sj.jim.7000155 (2001).
- 212 Saint-Amans, S., Perlot, P., Goma, G. & Soucaille, P. High production of 1,3-propanediol from glycerol by *Clostridium butyricum* VPI 3266 in a simply controlled fed-batch system. *Biotechnology Letters* **16**, 831-836, doi:10.1007/BF00133962 (1994).
- 213 Chatzifragkou, A., Dietz, D., Komaitis, M., Zeng, A. P. & Papanikolaou, S. Effect of biodiesel-derived waste glycerol impurities on biomass and 1,3-propanediol production of *Clostridium butyricum* VPI 1718. *Biotechnol Bioeng* **107**, 76-84, doi:10.1002/bit.22767 (2010).
- 214 Barbirato, F. et al. Anaerobic pathways of glycerol dissimilation by *Enterobacter agglomerans* CNCM 1210: limitations and regulations. *Microbiol* **143** (Pt 7), 2423-2432, doi:10.1099/00221287-143-7-2423 (1997).
- 215 Feller, G. Cryosphere and psychrophiles: insights into a cold origin of life? *Life (Basel)* **7**, doi:10.3390/life7020025 (2017).
- 216 Margesin, R. & Miteva, V. Diversity and ecology of psychrophilic microorganisms. *Res Microbiol* **162**, 346-361, doi:10.1016/j.resmic.2010.12.004 (2011).
- 217 Yadav, A. N., Verma, P., Kumar, V., Sachan, S. G. & Saxena, A. K. Extreme cold environments : A suitable niche for selection of novel psychrotrophic microbes for biotechnological applica-

- tions. *Adv Biotechnol and Microbiol* **2** (2017).
- 218 Keto-Timonen, R. *et al.* Cold Shock Proteins: A minireview with special emphasis on Csp-family of *Enteropathogenic Yersinia*. *Front Microbiol* **7**, 1151, doi:10.3389/fmicb.2016.01151 (2016).
- 219 Caruso, C. *et al.* Extracellular polymeric substances with metal adsorption capacity produced by *Pseudoalteromonas* sp. MER144 from Antarctic seawater. *Environ Sci Pollut Res Int* **25**, 4667-4677, doi:10.1007/s11356-017-0851-z (2018).
- 220 Caruso, C. *et al.* Production and biotechnological potential of extracellular polymeric substances from sponge-associated antarctic bacteria. *Appl Environ Microbiol* **84**, doi:10.1128/AEM.01624-17 (2018).
- 221 Nichols, P. D., Guckert, J. B. & White, D. C. Determination of monosaturated fatty acid double-bond position and geometry for microbial monocultures and complex consortia by capillary GC-MS of their dimethyl disulphide adducts. *J Microbiol Methods* **5**, 49-55 (1986).
- 222 Pitcher, A., Hopmans, E. C., Schouten, S. & Sinninghe Damsté, J. S. Separation of core and intact polar archaeal tetraether lipids using silica columns: Insights into living and fossil biomass contributions. *Organic Geochem* **40**, 12-19 (2009).
- 223 Sturt, H. F., Summons, R. E., Smith, K., Elvert, M. & Hinrichs, K. U. Intact polar membrane lipids in prokaryotes and sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry--new biomarkers for biogeochemistry and microbial ecology. *Rapid Commun Mass Spectrom* **18**, 617-628, doi:10.1002/rcm.1378 (2004).
- 224 Damste, J. S. *et al.* 13,16-Dimethyl octacosanedioic acid (iso-diabolic acid), a common membrane-spanning lipid of *Acidobacteria* subdivisions 1 and 3. *Appl Environ Microbiol* **77**, 4147-4154, doi:10.1128/AEM.00466-11 (2011).
- 225 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases

and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).

- 226 Naumann, D. FT-infrared and FT-Raman spectroscopy in biomedical research. *App Spectroscopy Rev* **36**, 239 (2001).
- 227 Zeidan, A. A. et al. Polysaccharide production by lactic acid bacteria: from genes to industrial applications. *FEMS Microbiol Rev* **41**, S168-S200, doi:10.1093/femsre/fux017 (2017).
- 228 Wang, L., Yu, X., Xu, H., Aguilar, Z. P. & Wei, H. Effect of skim milk coated inulin-alginate encapsulation beads on viability and gene expression of *Lactobacillus plantarum* during freeze-drying. *LWT- Food Sci Technol* **68**, 8-13 (2016).
- 229 Anwar, M. A. et al. Inulin and levan synthesis by probiotic *Lactobacillus gasseri* strains: characterization of three novel fructansucrase enzymes and their fructan products. *Microbiol* **156**, 1264-1274, doi:10.1099/mic.0.036616-0 (2010).
- 230 Ortiz-Soto, M. E., Olivares-Illana, V., López-Munguía, A. & Pez-Munguía, A. L. Biochemical properties of inulosucrase from *Leuconostoc citreum* CW28 used for inulin synthesis. *Biocatal Biotransformation* **224**, 275-281 (2004).
- 231 Wada, T., Ohguchi, M. & Iwai, Y. A novel enzyme of *Bacillus* sp. 217C-11 that produces inulin from sucrose. *Biosci Biotechnol Biochem* **67**, 1327-1334, doi:10.1271/bbb.67.1327 (2003).
- 232 Anwar, M. A., Kralj, S., van der Maarel, M. J. & Dijkhuizen, L. The probiotic *Lactobacillus johnsonii* NCC 533 produces high-molecular-mass inulin from sucrose by using an inulosucrase enzyme. *Appl Environ Microbiol* **74**, 3426-3433, doi:10.1128/AEM.00377-08 (2008).
- 233 van Hijum, S. A., van Geel-Schutten, G. H., Rahaoui, H., van der Maarel, M. J. & Dijkhuizen, L. Characterization of a novel fructosyltransferase from *Lactobacillus reuteri* that synthesizes high-molecular-weight inulin and inulin oligosaccharides. *Appl Environ Microbiol* **68**, 4390-4398 (2002).
- 234 Flemming, H. C. & Wingender, J. The biofilm matrix. *Nat Rev*

- Microbiol* **8**, 623-633, doi:10.1038/nrmicro2415 (2010).
- 235 Shoaib, M. *et al.* Inulin: Properties, health benefits and food applications. *Carbohydr Polym* **147**, 444-454, doi:10.1016/j.carbpol.2016.04.020 (2016).
- 236 Kolida, S., Tuohy, K. & Gibson, G. R. Prebiotic effects of inulin and oligofructose. *Br J Nutr* **87**, S193-S197, doi:10.1079/BJN/2002537 (2002).
- 237 Dragos, A. & Kovacs, A. T. The peculiar functions of the bacterial extracellular matrix. *Trends Microbiol* **25**, 257-266, doi:10.1016/j.tim.2016.12.010 (2017).
- 238 Kriner, M. A., Sevostyanova, A. & Groisman, E. A. Learning from the leaders: gene regulation by the transcription termination factor Rho. *Trends Biochem Sci* **41**, 690-699, doi:10.1016/j.tibs.2016.05.012 (2016).
- 239 Lehnik-Habrink, M. *et al.* DEAD-Box RNA helicases in *Bacillus subtilis* have multiple functions and act independently from each other. *J Bacteriol* **195**, 534-544, doi:10.1128/JB.01475-12 (2013).
- 240 Jain, C. The *E. coli* RhIE RNA helicase regulates the function of related RNA helicases during ribosome assembly. *RNA* **14**, 381-389, doi:10.1261/rna.800308 (2008).
- 241 Itaya, M. *et al.* Isolation of RNase H genes that are essential for growth of *Bacillus subtilis* 168. *J Bacteriol* **181**, 2118-2123 (1999).
- 242 Dammel, C. S. & Noller, H. F. Suppression of a cold-sensitive mutation in 16S rRNA by overexpression of a novel ribosome-binding factor, RbfA. *Genes Dev* **9**, 626-637 (1995).
- 243 Guo, Q. *et al.* Structural basis for the function of a small GTPase RsgA on the 30S ribosomal subunit maturation revealed by cryo-electron microscopy. *Proc Natl Acad Sci USA* **108**, 13100-13105, doi:10.1073/pnas.1104645108 (2011).
- 244 Azeroglu, B. *et al.* RecG Directs DNA Synthesis during double-

- strand break repair. *PLoS Genet* **12**, e1005799, doi:10.1371/journal.pgen.1005799 (2016).
- 245 Wilkinson, A., Day, J. & Bowater, R. Bacterial DNA ligases. *Mol Microbiol* **40**, 1241-1248 (2001).
 - 246 De Maio, A. Heat shock proteins: facts, thoughts, and dreams. *Shock* **11**, 1-12 (1999).
 - 247 Karlin, S. & Brocchieri, L. Heat shock protein 70 family: multiple sequence comparisons, function, and evolution. *J Mol Evol* **47**, 565-577 (1998).
 - 248 Scheff, G., Salcher, O. & Lingens, F. *Trichococcus flocculiformis* gen. nov. sp. nov. A new gram-positive filamentous bacterium isolated from bulking sludge. *App Microbiol and Biotechnol* **19**, 114-119 (1984).
 - 249 Goris, J. et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81-91, doi:10.1099/ij.s.0.64483-0 (2007).
 - 250 Baek, G., Kim, J., Cho, K., Bae, H. & Lee, C. The biostimulation of anaerobic digestion with (semi)conductive ferric oxides: their potential for enhanced biomethanation. *App Microbiol and Biotechnol* **99**, 10355-10366, doi:10.1007/s00253-015-6900-y (2015).
 - 251 Bialek, K., Cysneiros, D. & O'Flaherty, V. Hydrolysis, acidification and methanogenesis during low-temperature anaerobic digestion of dilute dairy wastewater in an inverted fluidised bioreactor. *Appl Microbiol Biotechnol* **98**, 8737-8750, doi:10.1007/s00253-014-5864-7 (2014).
 - 252 Keating, C. et al. Biological phosphorus removal during high-rate, low-temperature, anaerobic digestion of wastewater. *Front Microbiol* **7**, 226, doi:10.3389/fmicb.2016.00226 (2016).
 - 253 Lee, C., Kim, J., Shin, S. G. & Hwang, S. Monitoring bacterial and archaeal community shifts in a mesophilic anaerobic batch reactor treating a high-strength organic wastewater. *FEMS Microbiol Ecol* **65**, 544-554, doi:10.1111/j.1574-6941.2008.00530.x

(2008).

- 254 Regueiro, L., Veiga, P., Figueroa, M., Lema, J. M. & Carballa, M. Influence of transitional states on the microbial ecology of anaerobic digesters treating solid wastes. *Appl Microbiol Biotechnol* **98**, 2015-2027, doi:10.1007/s00253-013-5378-8 (2014).
- 255 Zeng, T. et al. Assessment of bacterial community composition of anaerobic granular sludge in response to short-term uranium exposure. *Microb Ecol*, doi:10.1007/s00248-018-1152-x (2018).
- 256 Lettinga, G. et al. High-rate anaerobic treatment of wastewater at low temperatures. *Appl Environ Microbiol* **65**, 1696-1702 (1999).
- 257 Parshina, S. N., Ermakova, A. V., Bomberg, M. & Detkova, E. N. *Methanospirillum stamsii* sp. nov., a psychrotolerant, hydrogenotrophic, methanogenic archaeon isolated from an anaerobic expanded granular sludge bed bioreactor operated at low temperature. *Int J Syst Evol Microbiol* **64**, 180-186, doi:10.1099/ijls.0.056218-0 (2014).
- 258 Birnboim, H. C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-1523 (1979).
- 259 Chun, J. et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* **68**, 461-466, doi:10.1099/ijsem.0.002516 (2018).
- 260 Gerhardt, P. Methods for general and molecular bacteriology. (American Society for Microbiology, 1994).
- 261 Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. t. & Smith, F. Colorimetric method for determination of sugars and related substances. *Anal Chem* **28**, 350-356 (1956).
- 262 Vincent, A. T., Derome, N., Boyle, B., Culley, A. I. & Charette, S. J. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *J Microbiol Methods* **138**, 60-71, doi:10.1016/j.mimet.2016.02.016 (2017).

- 263 Rossi-Tamisier, M., Benamar, S., Raoult, D. & Fournier, P. E. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* **65**, 1929-1934, doi:10.1099/ijss.0.000161 (2015).
- 264 Mignard, S. & Flandrois, J. P. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *J Microbiol Methods* **67**, 574-581, doi:10.1016/j.mimet.2006.05.009 (2006).
- 265 Gutleben, J. *et al.* The multi-omics promise in context: from sequence to microbial isolate. *Crit Rev Microbiol* **44**, 212-229, doi:10.1080/1040841X.2017.1332003 (2018).
- 266 Lee, Y. *et al.* Tyr51: Key determinant of the low thermostability of the *Colwellia psychrerythraea* cold-shock protein. *Biochemistry* **57**, 3625-3640, doi:10.1021/acs.biochem.8b00144 (2018).
- 267 Stabler, R. A. *et al.* Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol* **10**, R102, doi:10.1186/gb-2009-10-9-r102 (2009).
- 268 Matrosova, V. Y. *et al.* High-quality genome sequence of the radioresistant bacterium *Deinococcus ficus* KS 0460. *Stand Genomic Sci* **12**, 46, doi:10.1186/s40793-017-0258-y (2017).
- 269 Petrova, M. I. *et al.* Comparative genomic and phenotypic analysis of the vaginal probiotic *Lactobacillus rhamnosus* GR-1. *Front Microbiol* **9**, 1278, doi:10.3389/fmicb.2018.01278 (2018).
- 270 Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**, D286-293, doi:10.1093/nar/gkv1248 (2016).
- 271 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285, doi:10.1093/nar/gkv1344 (2016).
- 272 Mount, D. W. Using the basic local alignment search tool (BLAST). *CSH Protoc* **2007**, pdb top17, doi:10.1101/pdb.top17

(2007).

- 273 Phillips, D. C. The three-dimensional structure of an enzyme molecule. *Sci Am* **215**, 78-90 (1966).
- 274 Haft, D. H. et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* **46**, D851-D860, doi:10.1093/nar/gkx1068 (2018).
- 275 Gonzalez-Pajuelo, M. et al. Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1,3-propanediol from glycerol. *Metab Eng* **7**, 329-336, doi:10.1016/j.ymben.2005.06.001 (2005).
- 276 Tang, X., Tan, Y., Zhu, H., Zhao, K. & Shen, W. Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of *Escherichia coli*. *Appl Environ Microbiol* **75**, 1628-1634, doi:10.1128/AEM.02376-08 (2009).
- 277 Milne, C. B. et al. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC Syst Biol* **5**, 130, doi:10.1186/1752-0509-5-130 (2011).
- 278 Motwalli, O. et al. In silico screening for candidate chassis strains of free fatty acid-producing cyanobacteria. *BMC Genomics* **18**, 33, doi:10.1186/s12864-016-3389-4 (2017).
- 279 Maervoet, V. E. T., De Mey, M., Beauprez, J., De Maeseneire, S. & Soetaert, W. K. Enhancing the microbial conversion of glycerol to 1,3-propanediol using metabolic engineering. *Org Process Res & Dev* **15**, 189-202, doi:10.1021/op1001929 (2011).
- 280 Hao, J., Wang, W., Tian, J., Li, J. & Liu, D. Decrease of 3-hydroxypropionaldehyde accumulation in 1,3-propanediol production by over-expressing dhaT gene in *Klebsiella pneumoniae* TUAC01. *J Ind Microbiol Biotechnol* **35**, 735-741, doi:10.1007/s10295-008-0340-y (2008).
- 281 de Fouchecour, F., Sanchez-Castaneda, A. K., Saulou-Berion, C. & Spinnler, H. E. Process engineering for microbial production of 3-hydroxypropionic acid. *Biotechnol Adv* **36**, 1207-1222,

doi:10.1016/j.biotechadv.2018.03.020 (2018).

- 282 Ashok, S., Raj, S. M., Rathnasingh, C. & Park, S. Development of recombinant *Klebsiella pneumoniae* dhaT strain for the co-production of 3-hydroxypropionic acid and 1,3-propanediol from glycerol. *Appl Microbiol Biotechnol* **90**, 1253-1265, doi:10.1007/s00253-011-3148-z (2011).
- 283 Pan, G. et al. Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc Natl Acad Sci USA* **114**, E11131-E11140, doi:10.1073/pnas.1716245115 (2017).
- 284 D'Amico, S., Collins, T., Marx, J. C., Feller, G. & Gerday, C. Psychrophilic microorganisms: challenges for life. *EMBO Rep* **7**, 385-389, doi:10.1038/sj.embor.7400662 (2006).
- 285 Davidson, M. H. & Maki, K. C. Effects of dietary inulin on serum lipids. *J Nutr* **129**, 1474S-1477S, doi:10.1093/jn/129.7.1474S (1999).
- 286 Causey, J. L., Feirtag, J. M., Gallaher, D. D., Tungland, B. C. & Slavin, J. L. Effects of dietary inulin on serum lipids, blood glucose and the gastrointestinal environment in hypercholesterolemic men. *Nutr Res* **20**, 191-201, doi.org/10.1016/S0271-5317(99)00152-9 (2000).
- 287 Catry, E. et al. Targeting the gut microbiota with inulin-type fructans: preclinical demonstration of a novel approach in the management of endothelial dysfunction. *Gut* **67**, 271-283, doi:10.1136/gutjnl-2016-313316 (2018).
- 288 Ortiz-Soto, M. E., Olivares-Illana, V. & López-Munguía, A. Biochemical properties of inulosucrase from *Leuconostoc citreum* CW28 used for inulin synthesis. *Biocatal and Biotransformation* **22**, 275-281, doi:10.1080/10242420400014251 (2004).
- 289 Rozen, R. et al. Effect of chlorhexidine on molecular weight distribution of fructans produced by fructosyltransferase in solution and immobilized on surface. *Carbohydr Res* **338**, 571-575 (2003).

- 290 Bryant, J. E., Lecomte, J. T. J., Lee, A. L., Young, G. B. & Pielak, G. J. Protein Dynamics in Living Cells. *Biochemistry* **44**, 9275-9279, doi:10.1021/bi050786j (2005).

APPENDICES

ACKNOWLEDGEMENTS

Disclaimer - Long session

I am always fascinated from fantasy novels. Therefore, for me the best way to describe my PhD thesis is like a quest. You embark as a humble young scientist in this journey call PhD being all excited, ambitious and optimistic. But only if you knew better how naïve and fragile that is. PhD is made of merciless failed experiments, forsaken dark week-ends, heart-breaking revisions, ruthless students and hordes of disappointments. Of course, all these are tough challenges that require extensive stamina, a respected level of agility, signs of intelligence and an expertise in dexterity, and yes, eventually you may find your way in the misty way of science. Even so, I would recommend doing a PhD. I have gained so many experiences and I have evolved not only as a scientist, but also as a person. These four years (and something more) contained moments that extended from just work and became more part of me. However, in this glorious quest, I was not alone. I have met people (many) that stood by my side and helped me to overcome gigantic obstacles, fearsome issues and to create unforgettable moments of joy. Without their precious guidance and valuable support numerous things would have remained impossible. There are many people that contributed to my goal and I hope I managed to include them all here. Thank you all for making me complete my quest.

Prof. Fons (promoter), you are a definition of a Professor. Your way of thinking leads in simple advices that are able to close knowledge and extend scientific outcomes. My motivation for academia, is inspired by you, as your organized way to conduct science is something truly to be admired. The reason that you are Prof. Fons is that you have the ability to understand people and treat them as they deserve. You are always kind, eager to help and interested to listen. I would never forget that you shared my burden in that lake of Konstanz and I am really grateful for that. Finally, let us not forget that you can ran marathons (I do not think I can ever learn to do that). It was an honour working and having you as my promotor.

Peter (co-promoter), you were the first teacher to meet when I came in Wageningen at 2012. From that day you have always but always been

there for me. Of course you were the study advisor of Bioinformatics Master, but more than that you have seen the best in me and help me to survive in the world of Bioinformatics and PhD. I feel that I owe you quite some and I hope in the future I will be able to repay it (maybe collaborations or maybe Greek food?). I liked our communication (for all topics) and especially the way that you see the importance in things. Unfortunately, I will not promise to stop spamming the e-mail folder that you created just for me!

Diana (co-promoter), do you remember when I entered your office and told you that I want to change the world? Well, I have not given up yet and that is because of you believing in me. Even more, me being able to complete my quest is definitely because of you. You were there to guide me, work with me and solve my issues. Numerous hours for finding perfection, something that you seek and now me too. You have definitely shape me as a scientist. During these years though, you were able to make me consider you my friend! That is your charisma, you find the right balance for making science fun, interesting and significant. Talking about fun, I am happy that you always got entertained with my ambitious time of finishing my PhD! Lucky to have you as my supervisor, but even more lucky to have you as my friend. Thanks Diana for everything!

Members of the **thesis committee**, thank your critical assessment and evaluation of this thesis.

I would like to thank **Professor Guido Jenster** that I really respect, and his group in my Master thesis. I have learned so much in such a sort time that I literally do not know how it is possible. Definitely, one of the best times of my scientific career. I am happy that now we are both in EMC.

Divided between two groups, I would like to start with Micphys. I would like to thank **Ton van Gelder**, the man of the hour, the MVP in Micphys. So much help in the experiments and HPLC. Furthermore, you were one of the initiators of *Trichococcus* project and thus you are even more important. I am definitely going to miss working with you and also your cool BBQ parties (and your homemade beer). I would like to thank **Steven** one of my co-authors. Stevie, I think you are really

good on what you are doing, I believe you can go even further and stay cool! I would like to thank **Caroline** for being a great scientist and happy that I established a good communication with you in my early time of the PhD. I would like to thank a lot **Peer** for being a brilliant colleague but most of all a great friend. When I saw you in the lab, I know that would be a fun day. Thanks also for supporting me and especially in my unhealthy moments. **Ju!** Thanks a lot for being a good friend of mine and finally releasing me for trying to say correct your name and go with a more subtle way. I would like to thank my man **Rikkie**, for being one of the kindest people, a good friend and with somehow good jokes. **Monika**, thank you for all the collaboration in the lab and for the laughs. **Danny boy** thanks a lot for the nice talks and the nice jokes. **Martjin** thanks for all the fun moments and I hope we meet again for some pc gaming. **Anna F.** thanks a lot for the nice collaboration and fun times. Thanks **Lara** for the nice teaching times and fun moments. Thanks **Nam** for the nice teaching moments and the positive words. Thank you **Yuan** for the good times and for the fun trip in USA. Thanks **Lot, Denny, Monir, Vincente, Susakul, Nico, Nohemi, Iame, Petra and Michael** for making work fun.

Time to thank my other half group and that is CSB (and SSB). First of all **Maria**, thank you so much. I think you are the cornerstone of CSB and thank you for your precious input. **Benoito**, you are an important friend of mine, although "casual". I cannot imagine how many times did I tease you, no regrets. You made PhD and my life so much more fun and I hope we keep our cool friendship until scamZ arrives (so forever). **Edo** thanks a lot for all the statistical advices and the help with my Master thesis. **Jasper**, thank you a lot for the many interactions, as my Master thesis supervisor, officemate, bioinformatician colleague and co-author. It was really great working you in all these different levels. Thank you **Juanan**, for being my teacher in my Master but also a friend now, pity that Spain got you. **Milad**, my Persian friend, we had so good laughs, looking forward for more in the future! **Maarten**, you are special for me, I cannot deny it. Thanks for being around me you always made my day in real and gaming life. I am happy to have met a great gamer and an almost Nobel. **Javi** thanks for all the nice jokes and the party in Las Vegas. **Kal**, thanks a lot for being so cool, how can you be so cool? I wish you all the best. **Ruben**, we know each other

for quite some time, from our Masters and we keep interacting and gaming together. Well I hope that doesn't change! **Dorett**, thanks for the good funs and the proteomics questions! I would like to thank Niru, for making me listen more Bollywood songs in my office. I keep this habit even in my new work. Thanks **Jasper S** for the PhD trip cool times. I would like to thank people that made CSB awesome and these are Anna, Polish Anna, Wassin, Melanie, Erika, Nhung and Rik.

I would like to thank my office mates during my PhD. Old building: I would like to thank **Mark** as he improve a lot my bioinformatics skills and help me understand complicated issues. **Tjerko**, thanks for the good times! Thanks **Rob** for having really good fun and always nice conversations. Thanks a lot **Riek**, it was fun for talking about science. Thanks **Emma** the latest addition, for some interesting talks and songs. New building: **Bastian**, fun fact, you were my office mate even in my Master thesis and then also as a PhD. Basti, you are just awesome and I am a huge fun of yours (people know my nickname in Lan parties). I had fun with you and I miss the genomics conversations but mostly I miss your sounds of disappointment. **Niels**, you are a super kind person and I enjoyed so much the time with you in the office. We manage to calm each other many times and that matters. Looking forward in the future to eat a nougat with you. **Chris**, you are Greek and you are good friend of mine, but most important we had our office fun. I liked the help that we provided to each other and the nice comments for life. I will miss being asked for my USB cable.

I would like to thank my fellow gang of Greeks in the department. **Giannis K**, thank you so much you are from the people that you stood by my side and I am super grateful. I had laughs and good stories with you and I believe you will do great in your life, you deserve it. **Prokopis**, you are a scientist, hard worker and dedicated, finish that PhD in a year and come to work with me. Thank you for the Cypriote phrases stuck in my mind. **Giannis M**, another great scientist and from the early Greeks, thanks a lot for all the nice conversations and helps, we had some really good memories, especially in PhD trip! **Costas**, you came in the end of my PhD but definitely not late for having some serious fun and important conversations, keep rocking, especially in the dawn! **Stamatis**, you were the only Greek before I started my PhD

and I had some nice talks with you. You are the only one knowing real Synthetic Biology. **Menia**, you are sweet and kind, I am really happy to have meet you and confirmed why North Greeks are the best Greeks. **Despoina** was nice to meet you and keep doing great things. In the Greeks, I will also add **Sudy**, as he was proselytized by the Greeks. Sudy we had great times, interesting conversations and I am looking forward for our next Bollywood super movie (a.k.a. Robot 2).

I would like to thank all the technicians in our lab. Thank you **Philippe, Ineke, Merljin** and **Sjon**. Special thanks to **Tom W** and **Tom S** for the nice project of bioreactors. **Tom W**, we had some nice jokes and I want you back for online gaming. Special thanks to the lab secretaries **Anja** and **Caroline**. I would like to thank Mr. **Wim**, for giving me the autoclaving job before my PhD study, for helping me about everything during my PhD and for being the soul of microbiology.

I would like to thank my collaborators Dr. **Yuemei Lin, Michel Konnen**, Dr. **Jaap Sinninghe Damsté**, Dr. **Markus Göker** and Dr. **Jan Meier-Kolthoff**. Special thanks to Sjef for all the analysis of the proteomics and doing my samples with priority. Super special thanks to Dr. **Sofia**, a true amazing scientist that I definitely enjoyed your wisdom in the lab.

I would like to thank my five students. I would like to thank **Henri K** for being my first student, **Henry N**, for your critical performance (after changing your CV picture) resulting in one submitted publication and soon hosting me in Ecuador. I would like to thank **Lala A** for being one of the most hard worker students but mostly the kindest of all. I was proud to have the first (and only?) Azerbaijani student. **Laura N**, for being an amazing student with only positives and also being a good friend. I am expecting much from you and some nice drawings too! **Arianna**, thanks for the challenge, I don't know if I still like cats.

There are people in the lab that I would also like to thank. **Detmer**, you are special to me for being an important person in important moments. You are a really good friend of mine and I will really try to keep it that way. I hope I can visit your castle soon. **Jorrit**, brotha, we are related in so many fields and we had some pretty good laughs in events and games. I am looking forward for having more jorrit times! **Carrie**,

thanks for all the super cool moments and the really nice talks. You should know you speak so good Greek! **Jannie**, thanks for all the fun and nice conversations, I am looking forward for our dancing business potential. **Shaaron**, thanks for the good times and that you like specific ps4 games. I thank the three of you for being there when stress was a reality. **Jeroen**, thanks for being my volleyball mate. **Marcelle** and **Kees** (together officially now), thanks for being my friends and having some interesting talks. **Hugo**, thanks for knowing you from the designing a laboratory video for Africa and for the Lan parties. **Teunke**, thanks for all the nice conversations and support! I would like to thank **Lennart** for the nice Lennart jokes and for being a good friend. I would like to thank **Joyshree** for being a true Micphyser and a good friend. People in the lab that I would like also to thank are **Gerben** (good talks), **Catarina and Isma** (old friends), **Jo-hanna** (party fun and good time in PhD trip), **Enrique** (good times), **Ran** (super kind), **Caifang** (super kind), **Yifan** (whiskey fellow), **G** (Jie), **Alex** (the Russian), **Emmy** (good times), **Wen** (working fun), **Prartha** (good times), **Patrick** (go-kart mate), **Siavash** (teaching and nice talks), **Moleco Spanish Postdocs** (for teaching me correct Spannish), **Janneke O** (for the old good times), **Janneke E** (for the new good times), **Joan** (good talks) **Marnix** (fun talks and good times), **Nico C** (nice talks), **Alex K** (good times) **Dr. Clara** and **Prof. van der Oost**.

I would like to thank **Corné Klaassen** for the possibility of working in MMIZ of EMC and all the nice projects that we share. I would like to thank all the people in **MMIZ EMC** that have accepted me and make me feel more than welcome in this short time.

I would like to thank Carol Kitsou for the critical reviewing of my thesis.

I would like to thank my two paranymphs. **Irene**, the known pure awesomeness, I cannot think another person that have seen me as I truly I am. You have been from the beginning a landmark in my life guiding my way or generating some serious fun. For me it is so simple to crack a joke with you especially because we can understand each other so good. I am happy for our collaborations in science (a scientist of your level is nice to have as co-author), but mostly I am so happy that I have met you. **Bart-sama**, I am so grateful to have spent these years next to your office. Sometimes, I was looking forward to arrive in the office

just to show you some cool gadget. We talked for everything, I do not think we have not talk for something and that is because you listen, you are the best listener ever. More than that, without even knowing me you gifted me stuff, which shows what a gentle heart you have. You are always there for me and I can count on you and I hope you know that goes vice versa. Together with the kind **Arwa** and the best kids ever you have a great family.

I would like to thank **Stratos, Dimitris Ser** and **Pin** in Greece as a best friends of mine for 16 years now. I would like to thank also the gaming gang from **Samos**. I would like to thank **Sofia, Ioanna, Christos** for being really good friends of mine in Netherlands.

I would like to thank what makes my life interesting and that is my best friend **Argyris**. You are my brother and you know it. From our Bachelor till now nothing has changed. We still face this world as we like. Thank you so much.

I would like to thank the most important people in my life and that is my family. Starting with the best girls ever my sisters. My dear **Polina**, thanks for all my childhood memories and soon new even, my beloved **Konstantina**, thanks for being the most sensitive and charismatic sister, I am sure that you will do wonders in life. **Efi**, my little genius, I am sure that soon I will be reading your PhD acknowledgments. My precious **Stefania** just simple the best sister ever, a miracle of being. **Aunt Ioulia**, thanks for all the taking care, love in Greece and for your delicious cakes. **Theia Genofeva**, my favourite soprano thank you for being the kindest person that I know. Ευχαριστώ πολύ και σας αγαπώ!

Three more people needs special recognition. I would like to thank my **dad Leonidas** for all his sacrifices and efforts for me to reach in the position that I am now. You were denied part of life for me to have a chance in life. Now, one of the really important people, my **mom Eleni**. You always tell me that when I was born you wished that when you will talk to me, I would always understand and then was that for the first time I blinked my eyes. We are connected in this luck of our misfortune and I love you so much. Thanks mom and especially for being my best friend. Finally, I would like to thank my **uncle Vaggelis**. Without you, I would not have been who I am. Every Nikolas moment is because of

you, my first toys, my first room, my first MS-DOS PC, my first car and my first day in science. I do think I have your genes (especially the university part) and I am proud of that. Thank you so much.

I would like to thank two people that are not any longer with us and those are my **grandpa Ilias** and my **grandma Konstantina**. My grandpa was a person of honor and discipline. An example of values that I so much look upon. My grandma was the sweetest little being, she literally raised me. I love you so much grandma and I know that my happiness has its limit now that you are not around. I miss you both.

I want to end up with a phrase that always my grandma said and I follow. For love you have to jump the highest wall, even if it hurts you.

ABOUT THE AUTHOR



Nikolaos Strepis was born on the 8th of September in 1988 in Volos, Greece. He studied for a BSc degree in Biochemistry and Biotechnology at the High-University of Thessaly, Greece from where he graduated at 2011. During that time he worked for three months in the General Hospital of Greece, supervising the analysis of biochemical and cancer biomarkers. His BSc thesis was conducted at Medical School of Thessaly, where he designed simulations for preventing aneurysms in carotid artery using patient MRI scans. At 2012, he started his MSc Bioinformatics at Wageningen University, the Netherlands and finished at 2014. In his master, he was chosen to present the study in the European Evaluation Committee with positive feedback. He conducted his major thesis at Erasmus Medical Center, Rotterdam under the supervision of Prof. Guido Jenster in the experimental urological oncology. Directly after the Master degree, he performed the PhD work described in this thesis at Microbiology department, Wageningen University under the supervision of Dr. Diana Z. Sousa (Assistant Professor), Dr. Peter J. Schaap (Associate Professor) and Prof. Fons J.M. Stams (Chair). Currently, Nikolas has a Postdoc position in Erasmus Medical Center, Rotterdam at Medical Microbiology and Infections Diseases department. Positioned at Molecular Diagnostics and Research and Development, he is involved in introducing next generation sequencing in diagnostics, and in depth investigation of outbreaks and antimicrobial resistance.

PUBLICATIONS

Strepis N et al. Description of *Trichococcus ilyis* sp. nov. by combined physiological and in silico genome hybridization analyses. Int J Syst Evol Microbiol 66(10):3957-3963.

Published

Strepis N*, Palakawong Na Ayudthaya S*, Pristaš P, Plugge CM. 2017. Draft genome sequence of *Actinomyces glycerinitolerans* strain G10^T, isolated from sheep rumen fluid. Genome Announcement 5:e01589-16.

Published

Parshina S, Strepis N, Aalvink S, Nozhevnikova AN, Stams AJM, Sousa DZ. *Trichococcus shcherbakoviae* sp. nov. isolated from a laboratory-scale anaerobic EGSB bioreactor operated at low temperatures. Int J Syst Evol Microbiol.

Accepted

Sánchez-Andrea I, Florentino AP, Semerel J, Strepis N, Sousa DZ, Stams AJM. Co-culture of a novel fermentative bacterium, *Lucifera butyrica* gen. nov. sp. nov., with the sulfur reducer *Desulfurella amilsii* for enhanced sulfidogenesis. Front Microbiol.

Accepted

Strepis N, Benavides HN, Schaap JP, Stams AJM, Sousa DZ. Genome-guided identification of novel physiological traits of *Trichococcus species*.

Submitted

Strepis N*, Koehorst JJ*, Stams AJM, Sousa DZ, Schaap JP. Bioprospecting microbial omics data for the trait of 1,3-propanediol production and potential other traits.

Submitted

Strepis N, de Nies L, Boeren S, Schaap JP, Stams AJM, Sousa DZ. Cold survival mechanisms analysis using proteomics and physiological

analysis in the psychrotolerant *Trichococcus patagoniensis*.

In preparation

Strepis N, Akhundova L, Palma A, van de Weijer T, Schonewille T, Boeren S, Schaap JP, Stams AJM, Sousa DZ. Physiological and proteomics analysis of 1,3-propanediol metabolism in *Trichococcus pasteurii* and *Clostridium butyricum*.

In preparation

Strepis N, Salvador AF, del Rio TG, Smidt H, Alves MM, Schaap JP, Stams AJM, Sousa DZ. Comparative genomics of *Syntrophomonas zehnderi* and *Syntrophomonas wolfei* gives insights into the syntrophic degradation of fatty acids.

In preparation

* Equal contribution



*Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment*

| 225 |

D I P L O M A

For specialised PhD training

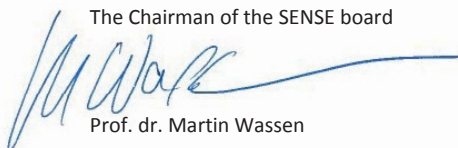
The Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Nikolaos Strepis

born on 8 September 1988 in Volos, Greece

has successfully fulfilled all requirements of the
Educational Programme of SENSE.

Wageningen, 16 January 2019

The Chairman of the SENSE board

Prof. dr. Martin Wassen

the SENSE Director of Education

Dr. Ad van Dommelen

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)



K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

**“BECAUSE LEARNING DOES NOT CONSIST ONLY
OF KNOWING WHAT WE MUST OR WE CAN DO,
BUT ALSO OF KNOWING WHAT WE COULD DO
AND PERHAPS SHOULD NOT DO.”**

Umberto Eco, *The Name of the Rose*

The research described in this thesis was financially supported by the European Research Council under the European Union's Seventh Framework Program (FP/2007-2013) / ERC Grant Agreement (323009) and by the Gravitation grant (024.002.002) of the Netherlands Ministry of Education, Culture and Science.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Printing of this thesis was financially supported by the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM).

Cover design and layout by Birte Ketting.

Printed by proefschriftmaken on FSC-certified paper.

