



universität
wien

MASTERARBEIT

Titel der Masterarbeit

Genomic insights into the evolution and unexpected
lifestyle of the mucus-dwelling bacterium
Mucispirillum schaedleri

verfasst von

Carina Pfann, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2015

Studienkennzahl lt. Studienblatt: A 066 833

Studienrichtung lt. Studienblatt: Masterstudium Ökologie

Betreut von: Assoz. Prof. Dipl.-Biol. Dr. Alexander Loy

Mitbetreut von: Ass.-Prof. David Berry, PhD

Table of content

Introduction.....	1
Material and methods.....	3
DNA sequencing, assembly and check for completeness	3
Genome annotation and analysis.....	3
Type VI secretion system (T6SS).....	3
Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system and mobile genetic elements identification	3
Proteins with eukaryotic-like domains.....	4
Identification of non-identical genes	4
Putative horizontally transferred genes.....	4
Results	5
Genome reconstruction and comparison	5
General genomic features	9
Central metabolism	9
Amino acid biosynthesis.....	10
Respiration and oxygen stress.....	10
Putative electron donors and carbon sources	11
Cofactors and vitamins.....	11
Other transporters	12
Storage compounds.....	12
Motility	12
Virus defense (CRISPR)	12
Mobile genetic elements.....	12
Proteins with eukaryotic-like domains.....	13
Type VI secretion system.....	14
Other secretion, virulence and resistance factors	14
Putative horizontally transferred genes (HGT).....	15
Discussion	22
Lifestyle – possible metabolic strategies.....	22
Secretion of proteins.....	23
Defense strategies and mechanisms of resistance to oxidative stress	24
Horizontal gene transfer	25

Conclusions and outlook	27
Abstract	28
Zusammenfassung.....	29
References.....	30
Supporting information.....	39
Comparing insertion sequences in both genomes.....	39
Supplementary tables	39
Supplementary figures	40
Acknowledgments.....	41
Curriculum vitae	42
Personal data.....	42
Education.....	42
Work experience	42

Introduction

Mucispirillum schaedleri is a Gram-negative, non-spore forming, non-colony forming bacterium with anaerobic growth at 37°C. It shows a curved to spiral-shaped rod morphology (3-4 µm x 0.4µm) with single, unsheathed, bipolar flagella (**Figure 1**)¹.

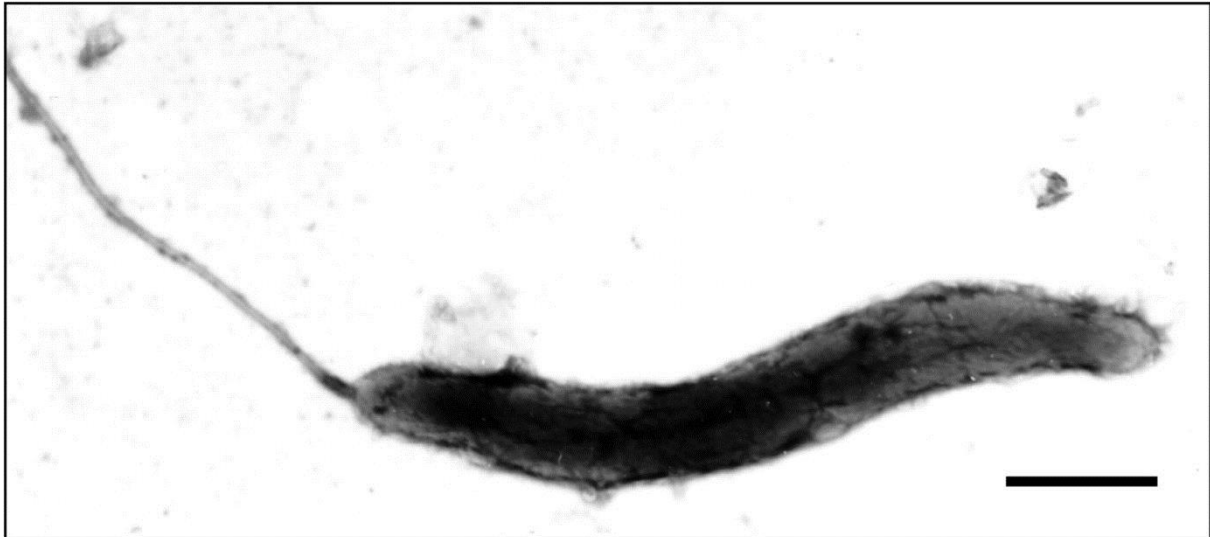


Figure 1: Transmission electron micrograph of *Mucispirillum schaedleri* (picture reproduced from reference¹).

Mucispirillum schaedleri inhabits the mammalian gastrointestinal (GI) tract and is physically associated with the mucus layer, where it is thought to degrade mucus¹. Mucus is primarily composed of mucins, high-molecular-weight O-linked glycoproteins, that stabilize the mucus barrier, regulate microbial community composition and influence metabolic functions of the microbiota². The reduction of the protective mucus layer by microorganisms is considered to be a pathogenicity factor³. However, an estimated number of 1% of microbiota in the intestine use mucin as carbon and energy source by degrading the oligosaccharide chains of mucins using enzymes like glycosidases⁴, or the protein backbone⁵. Like *Helicobacter* or *Campylobacter*, the majority of spiral-shaped bacteria identified in the mammalian GI tract¹, *Mucispirillum schaedleri* uses its spiral shape for moving through mucus and other viscous environments⁶. Its typical phenotypic trait is possibly a result of the selective pressure of living in viscous environments like the mucus layer in the gut. Taxonomically *Mucispirillum schaedleri* has no close relationship with the genus *Helicobacter* or *Campylobacter* or the phylum *Proteobacteria*, but is placed within the phylum *Deferribacteres*, a phylogenetically distinct lineage within the domain *Bacteria*¹.

The GI tract of mammals is densely colonized by microorganisms⁷. The metabolic activities of these microbiota include the efficient breakdown of host- and nutrient-derived carbohydrates to metabolites such as carbon dioxide, short-chain fatty acids (SCFAs) and hydrogen⁸. *Mucispirillum schaedleri* has been observed in the GI tract of termites, cockroaches, turkeys, pigs, goats, dogs and several common healthy laboratory mouse strains⁹⁻¹⁴, and is associated with inflammation in the murine gut¹⁵. Recently it has been detected as one of the few discriminating taxonomic groups between mice and humans¹⁶, that could not be detected in humans. However, it has been shown that it can inhabit the human GI tract in low abundance in the un-prepped colon and just disappeared after a 24-hour clear liquid diet and a standard polyethylene glycol-based bowel cleansing preparation¹⁷. It is not clear what factors cause this indiscernibility. The inference of a

taxa's exclusivity from the lack of observation in metagenomic studies is difficult as we are not yet able to capture the whole fraction of bacterial taxa present in the gut with current techniques. Due to the boundaries of limited sequencing depth, low-abundance taxa are often overlooked and it could be that *Mucispirillum schaedleri* is present in healthy humans in such a low abundance, that it is below the detection limit. Besides technical reasons there could clearly also be a biological reason like host-specificity or competition with or within the host we do not know about yet. Despite the possibility of a low-abundance occurrence of *Mucispirillum schaedleri* in humans it might play an important role. In mouse studies it has been shown that *Mucispirillum* was significantly increased in induced colitis and can be regarded as a biomarker for spontaneous colitis¹⁸. Besides members of *Helicobacteraceae*, *Mucispirillum* is one of a few discriminatory lineages for active colitis with significant enrichment in colitogenic gut microbiomes¹⁹. Despite its localization in the mucus layer it has not been identified as a significant degrader of host-derived compounds in vivo⁵, which could either be due to its use of non-host derived nutrients or due to the use of compounds not targeted in any study yet.

Mucispirillum schaedleri strain ASF457 is one of eight bacterial species of the altered Schaedler Flora (ASF)²⁰. The ASF is a standardized and defined low-complexity microbiota consortium for colonizing germfree rodents and for conducting studies with gnotobiotic mice. It consists of bacteria that had been isolated from the gut of conventional mice. Originally developed as the Schaedler flora (SF) in the mid-1960s by Russell W. Schaedler²¹, Roger Orcutt revised it and defined the ASF, which consists of four members of the original SF and four new strains including *Mucispirillum schaedleri* ASF457²². Recently draft genomes of all eight members of the ASF, including *Mucispirillum schaedleri* ASF457, have been announced²³.

In this study we compared the draft genomes of two *Mucispirillum schaedleri* ASF457 strains. This allowed us to gain insight into the unexpected lifestyle of this mucus-dwelling bacterium, its evolution with respect to laterally transferred genes and its survival strategies close to mucosal tissue blooms during inflammatory events.

Material and methods

DNA sequencing, assembly and check for completeness

For the *Mucispirillum schaedleri* ASF457 (genome MCS) shotgun-sequencing was performed on the Ion Torrent platform. The reads were then assembled using Newbler (software-version: 2.6), a proprietary assembler developed by 454 Life Sciences. The assembly of the *Mucispirillum schaedleri* ASF457 (genome AYGZ) was downloaded from the Broad Institute website²⁴. The completeness check for both genomes was done using tRNAscan-SE²⁵ (software version: 1.23), RNAmmer²⁶ (software version: 1.2), AMPHORA2²⁷ and CheckM²⁸ (lineage workflow).

Genome annotation and analysis

Genomes were annotated using the MicroScope²⁹ (originally named MaGe³⁰) annotation platform. The coding sequences (CDS) were automatically predicted and annotated. These annotations were used for analysis and interpretation of key metabolic functions. Metabolic pathways were reconstructed with the help of MicroCyc and the KEGG³¹ classification scheme within MicroScope.

Other features of the genome were separately identified using manual investigation on an annotated gene basis or other bioinformatics tools specified later.

Type VI secretion system (T6SS)

The type VI secretion system (T6SS) consists of a gene cluster containing 13 genes necessary for function³². These core components have distinct Clusters of Orthologous Groups (COG) IDs of which most are unique for T6SS function. For characterization of the T6SS these COG IDs were used to search for in MicroScope. Gene names were taken from UniProt. T6SS cluster genes with missing COG ID annotation in MicroScope were identified by blasting the protein sequence against the NCBI NR database.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system and mobile genetic elements identification

CRISPR repeats were identified using the CRISPR recognition tool (CRT)³³ as well as Piler-CR³⁴. Potential viruses corresponding to spacers within the CRISPR loci were searched for using CRISPRTarget³⁵ by predicting and analyzing crRNA targets. CRISPRTarget was run with CRT results with default parameters (database: Genbank-Phage, RefSeq-Plasmid; gap open: -10; gap extend: -2; nucleotide match: 1; nucleotide mismatch: -1; e-value: 1; word size: 7). Identification of the CRISPR array position in the genome and of the accompanying *csn* and *cas* genes was manually done in MicroScope.

Insertion sequences (IS) were identified using ISfinder³⁶. Protein sequences were blasted against the IS database, which provides a list of insertion sequences from bacteria and archaea, with BLASTP³⁷ (software version 2.2.31+) with a custom log e-value of -3. IS families of genes producing significant alignments were extracted from the BLAST results.

PHAST³⁸ was used for the similarity based analysis of putative prophages and phage-like proteins, whereby genomic regions that are enriched in protein coding genes with known phage homologs are detected.

Proteins with eukaryotic-like domains

For identification of proteins with Ankyrin repeat (ANK) domains the PROSITE^{39,40} motif for ANK_REP_REGION was used to scan it against our genomes using ScanProsite⁴¹ web service (option 3). Effective⁴² 4.0 (<http://www.effectors.org>) was used to predict other bacterial secreted proteins. To obtain domain names amino acid sequences of predicted loci were extracted and scanned against prosite.dat (release: 10.112 of Mar 04, 2015) using the Perl script ps_scan.pl (software version: 1.79). Both files were obtained from the PROSITE website <ftp://ftp.expasy.org/databases/prosite/>. The results were manually inspected, processed and sorted.

Identification of non-identical genes

Non-identical genes between the two strains were identified by subtracting almost identical genes (homology constraints were minLrap $\geq$ 0.9, maxLrap $\geq$ 0.9, identity $\geq$ 100%) from all genes (homology constraints were minLrap $\geq$ 0.9, maxLrap $\geq$ 0.9, identity $\geq$ 30%) using MicroScope's Gene Phyloprofile interface, whereby minLrap is defined as the quotient of the length of the match and the length of the shorter protein, whereas maxLrap is defined as the quotient of the length of the match and the length of the longer protein.

Putative horizontally transferred genes

For horizontal gene transfer (HGT) analysis phylogenetic trees for each gene were calculated using the software PhyloGenie⁴³. The trees that meet certain criteria were selected using PHAT (part of the PhyloGenie software package). For identifying genes putatively transferred from another phylogenetic group, any tree containing a node with *Mucispirillum* and with the phylogenetic group specified, but no other bacteria, was selected. The selection was done automatically, the list of annotated genes transferred from another phylum than *Deferribacteres* was sorted manually to overcome automatic sorting errors by PHAT. Nearest neighbors and sequence identity information were examined by blasting genes from the AYGZ genome using nucleotide BLAST³⁷ (software version: 2.2.31+) against the NCBI NR database and extracting the best hit. Phylogenetically closest species for inferring the donor of putatively transferred genes were examined using R⁴⁴ (software version: 3.2.1) and the R packages phytools⁴⁵ (software version: 0.4-56) and ape⁴⁶ by extracting species with the minimum phylogenetic distance within the node containing *Mucispirillum schaedleri*.

Results

Genome reconstruction and comparison

Recently, the genome of *Mucispirillum schaedleri* ASF457 (genome AYGZ) from a culture maintained in an American collection was announced²³. We had meanwhile sequenced the genome of *Mucispirillum schaedleri* ASF457 (genome MCS) using a culture maintained in a strain collection in Germany. While we believe that the two cultures are originally derived from the same stock from the Charles River laboratories, it is unclear how long the two cultures have been separated and how many bacterial generations may have occurred subsequently. Neither genome is closed, but estimates based on detection of tRNAs and conserved housekeeping genes indicate that the genomes are largely complete (**Table 1**). Both genomes have a GC content of 31% and a coding density of 88%. The number of detected coding sequences (CDS) without artifacts is 2,227 for the AYGZ and 2,187 for the MCS genome. The content of genomic objects in the two genomes is highly similar, with shared CDSs accounting for 92% of all CDSs in both genomes (shared CDSs are defined as having $\geq 80\%$ amino acid similarity and $\geq 80\%$ sequence length). As the genomes are not closed it is not possible to determine whether the differences in CDSs is due to absence from a genome or due to technical artifacts (incomplete sequencing and/or assembly). Of all shared CDSs, a bi-directional BLAST analysis identified only 49 CDSs that are not identical (**Table 2**). Non-identical CDSs, which generally still have very high sequence identity at the DNA level, include hydrogenase 2, transposases, transporters, and multiple proteins with unknown function.

Table 1: General features of the *Mucispirillum schaedleri* ASF457 genomes.

	AYGZ (BI, Broad Institute)	MCS (UV, University of Vienna)
Total size (Mb)	2.3	2.2
GC content	31.15%	31.17%
Number of contigs	39	145
Repeated Regions	10.54%	5.14%
Average CDS length (bp)	923.71	896.24
Average intergenic length (bp)	147.5	151.31
Protein coding density	88.21%	87.69%
# of genomic objects without artifacts	2,285	2,229
Total # of CDS	2,227	2,187
Protein coding genes	2,223	2,166
5S / 16S / 23S rRNA	2 / 3 / 2	1 / 1 / 1
tRNA	39 (all)	35 (all but Phe)

Housekeeping genes detected (estimate of completeness)	31 (of 31)	31 (of 31)
NCBI accession number	NZ_AYGZ000000000	not submitted
GOLD #	Gp0086862	Gi0042859

Table 2: Non-identical genes in both *Mucispirillum schaedleri* ASF457 genomes (AYGZ, MCS). Identity is given as the percentage of identity on the DNA level between both gene sequences based on the synton information in MicroScope.

Gene	Product	AYGZ (Locus BI)	MCS (Locus UV)	Identity
---	transposase (fragment)	AYGZv1_50084	MCSv1_10001	94.81%
gapA	glyceraldehyde-3-phosphate dehydrogenase A	AYGZv1_50095	MCSv1_10012	99.7%
hybC	hydrogenase 2, large subunit	AYGZv1_150009	MCSv1_1040005	99.82%
trbD	Conjugal transfer protein TrbD	AYGZv1_150077	MCSv1_1120007	98.96%
---	putative 4Fe-4S binding domain protein	AYGZv1_170017	MCSv1_1170008	99.58%
		AYGZv1_160085		35.85%
---	conserved membrane protein of unknown function	AYGZv1_160090	MCSv1_1190003	99.65%
glmU	Bifunctional protein GlmU [Includes: UDP-N-acetylglucosamine pyrophosphorylase , Glucosamine-1-phosphate N-acetyltransferase]	AYGZv1_160035	MCSv1_1190062	99.78%
---	protein of unknown function	AYGZv1_90119	MCSv1_130004	100%
		AYGZv1_210008		38.38%
		AYGZv1_260004		42.65%
		AYGZv1_90120		100%
		AYGZv1_100025		96%
---	protein of unknown function	AYGZv1_90127	MCSv1_130011	100%
		AYGZv1_230041		44.09%
---	transposase (fragment)	AYGZv1_50084	MCSv1_1320005	93.33%

---	protein of unknown function	AYGZv1_40006	MCSv1_1350008	99.56%
---	protein of unknown function	AYGZv1_210016	MCSv1_1440001	99.49%
---	protein of unknown function	AYGZv1_100025	MCSv1_170005	100%
---		AYGZv1_90120		96%
---		AYGZv1_210008		82%
---	exported protein of unknown function	AYGZv1_100039	MCSv1_190006	97.63%
---	putative enzyme	AYGZv1_100140	MCSv1_210032	99.65%
---		AYGZv1_150169		35.79%
sdC	Sodium-dependent dicarboxylate transporter SdcS	AYGZv1_120004	MCSv1_220011	99.8%
---	protein of unknown function	AYGZv1_120094	MCSv1_230053	99.61%
---	putative PAS/PAC sensor protein	AYGZv1_120095	MCSv1_230054	99.78%
---	exported protein of unknown function	AYGZv1_120176	MCSv1_260016	99.35%
---	transposase (fragment)	AYGZv1_50084	MCSv1_270001	94.07%
---	putative Glycosyltransferase	AYGZv1_130032	MCSv1_290004	99.38%
---		AYGZv1_130026		35.49%
---	putative Glycosyltransferase family protein	AYGZv1_130033	MCSv1_290005	99.68%
---	Acriflavin resistance protein	AYGZv1_140018	MCSv1_300002	99.91%
---	protein of unknown function	AYGZv1_140010	MCSv1_300010	98.61%
---	Extracellular solute-binding protein family 3	AYGZv1_180360	MCSv1_310010	99.62%
---	putative Trimethylamine-N-oxide reductase (cytochrome c)	AYGZv1_180345	MCSv1_310026	99.52%
lysC	aspartate kinase	AYGZv1_180213	MCSv1_420005	99.76%
---	polyamine transporter subunit , ATP-binding component of ABC superfamily (fragment)	AYGZv1_180199	MCSv1_420019	100%

		AYGZv1_180132		35.11%
		AYGZv1_100016		35.35%
copA	fragment of copper transporter (part 1)	AYGZv1_180198	MCSv1_430001	100%
pomA	Chemotaxis protein PomA	AYGZv1_80009	MCSv1_50007	99.61%
---	protein of unknown function	AYGZv1_180118	MCSv1_520008	99.62%
---	Long-chain fatty-acid-CoA ligase	AYGZv1_180110	MCSv1_520016	99.81%
		AYGZv1_180172		49.9%
		AYGZv1_180318		48.92%
---	putative Efflux pump, RND family, membrane fusion protein	AYGZv1_180064	MCSv1_570009	99.62%
---	protein of unknown function	AYGZv1_90120	MCSv1_680005	96.23%
		AYGZv1_100025		92.31%
		AYGZv1_210008		82%
---	protein of unknown function	AYGZv1_390001	MCSv1_690001	100%
		AYGZv1_360001		69.98%
---	protein of unknown function	AYGZv1_210016	MCSv1_750001	99.49%
hcpC	Major exported protein	AYGZv1_210014	MCSv1_750003	99.41%
---	protein of unknown function	AYGZv1_210008	MCSv1_760004	100%
		AYGZv1_260004		35.86%
---	protein of unknown function	AYGZv1_210008	MCSv1_760005	100%
		AYGZv1_100025		80.39%
		AYGZv1_90120		78.43%
---	protein of unknown function	AYGZv1_210004	MCSv1_770004	99.7%
		AYGZv1_210003		94.74%
---	conserved protein of unknown function	AYGZv1_200120	MCSv1_770022	95.24%
---	putative Histidine kinase	AYGZv1_230117	MCSv1_890004	99.87%

fabZ	(3R)-hydroxymyristol acyl carrier protein dehydratase	AYGZv1_230110	MCSv1_890011	99.32%
---	conserved protein of unknown function	AYGZv1_230068	MCSv1_890038	98.92%
		AYGZv1_170002		98.92%
		AYGZv1_160002		98.92%
		AYGZv1_110028		98.92%
		AYGZv1_90038		98.92%
		AYGZv1_70002		98.92%
		AYGZv1_50067		98.92%
		AYGZv1_160033		97.84%
		AYGZv1_150056		97.84%
		AYGZv1_150010		97.84%
		AYGZv1_140088		97.84%
---	protein of unknown function	AYGZv1_230080	MCSv1_900004	99.62%
rpoA	RNA polymerase, alpha subunit	AYGZv1_230050	MCSv1_930006	99.7%
---	protein of unknown function	AYGZv1_230041	MCSv1_930015	100%
		AYGZv1_90127		44.09%
rplF	50S ribosomal subunit protein L6	AYGZv1_230037	MCSv1_950003	99.48%
---	MCP methyltransferase, CheR-type	AYGZv1_140151	MCSv1_980006	99.64%

General genomic features

Central metabolism

A complete Embden-Meyerhof-Parnas (EMP) pathway was detected in the *Mucispirillum schaedleri* genome (**Table S1**). For the first step of this pathway, the phosphorylation of glucose can be performed by either a glucokinase (E.C. 2.7.1.2) or a phosphotransferase system (PTS, E.C. 2.7.1.69). The oxidative phase of the pentose phosphate pathway was not detected, but the non-oxidative phase is present (although the transaldolase, E.C. 2.2.1.2, could not be identified in either genome), which should allow for synthesis of five carbon sugars from glycolysis. The *Mucispirillum schaedleri* genome does not encode proteins for an Entner-Doudoroff pathway. Pyruvate:flavodoxin

oxidoreductase (E.C. 1.2.7.-), but not pyruvate:formate lyase or pyruvate dehydrogenase, is present, which is likely involved in converting pyruvate to acetyl-CoA. A complete tricarboxylic acid (TCA) cycle could be identified (**Table S1**), and features a *Helicobacter*-type succinyl-CoA:acetoacetate CoA transferase (SCOT, E.C. 2.8.3.8), which catalyzes the reversible conversion of succinyl-CoA to 2-oxoglutarate⁴⁷. A glyoxylate pathway (i.e. the genes for isocitrate lyase and malate synthase) was not detected. Oxaloacetate can be used to replenish the EMP pathway via the phosphoenolpyruvate (PEP) carboxykinase (E.C. 4.1.1.49). The EMP and TCA pathways can also be replenished from acetate or lactate utilization. Fructose can also be utilized via a fructokinase (E.C. 2.7.1.4).

Amino acid biosynthesis

Mucispirillum schaedleri has complete biosynthetic pathways for the amino acids alanine (from cysteine), β -alanine (from aspartate), arginine (acetyl cycle), cysteine, D-glutamate, D-glutamine, glycine, homocysteine, isoleucine, leucine, phenylalanine, proline, tyrosine, and valine (**Table S1**). Biosynthesis of aspartate from glutamate via an aspartate aminotransferase was not detected, but aspartate may be produced from fumarate via an aspartate ammonia-lyase (E.C. 4.3.1.1). Almost complete amino acid biosynthesis pathways include histidine, lysine, and ornithine. Other incomplete or not detected pathways include asparagine, homoserine, methionine, selenocysteine, serine, threonine, and tryptophan.

Respiration and oxygen stress

Mucispirillum schaedleri has genes for dissimilatory nitrate reduction to ammonia (DNRA), with a periplasmic nitrate reductase (NapA, E.C. 1.7.99.4) as well as a nitrite reductase (Nrf, E.C. 1.7.2.2) system (**Table S1**). The reduction of nitrate is coupled to the oxidation of ubiquinol to ubiquinone in the cytoplasmic membrane and the reduction of nitrite by electron transport from the menaquinol pool. Ammonia can be assimilated via a nicotinamide adenine dinucleotide phosphate (NADPH) specific glutamate dehydrogenase (GdhA, E.C. 1.4.1.4), which serves for glutamate synthesis⁴⁸, and a type 3 glutamine synthetase^{49,50} (GlnA, E.C. 6.3.1.2), which converts glutamate to glutamine. A glutamate synthetase was not detected. Glutamate and glutamine can then be used as biosynthesis donors for nitrogen-containing cell-components⁵¹.

Fumarate can also be used as a terminal electron acceptor for anaerobic respiration, which is coupled to the oxidation of membrane-bound menaquinol to menaquinone. The membrane-bound Rnf complex⁵², which is proposed to couple the electron transfer from reduced ferredoxin to NAD⁺ with the translocation of Na⁺ ions across the cytoplasmic membrane via a Na⁺-translocating ferredoxin:NAD⁺ oxidoreductase and thereby generating a sodium ion (Na⁺) gradient, was also detected. The resulting sodium ion gradient can then be used by an ATP synthase for ATP production (**Table S1**).

Mucispirillum schaedleri has a high-affinity cbb3-type cytochrome c oxidase (E.C. 1.9.3.1), which may be important for protection from O₂ stress^{53,54} or for micro-aerobic respiration^{55,56}. *Mucispirillum schaedleri* also has several genes for detecting and defending against oxidative stress including a superoxide reductase, catalase, cytochrome c peroxidase, and rubrerythrin. The genome also includes a nitroreductase as well as nitroreductase family proteins that may be important for scavenging nitrogen radicals formed during nitrate and nitrite reduction. A putative trimethylamine-N-oxide reductase (E.C. 1.7.2.3) has been detected, that can reduce trimethylamine N-oxide (TMAO) into trimethylamine (TMA), which could be a potential trophic link to methylotrophic methanogens

associated to gut environments, that use methylated amino compounds like TMA as substrate by reducing it by hydrogen to form methane⁵⁷. A thioredoxin reductase (E.C. 1.8.1.9), the only enzyme in bacteria known to reduce thioredoxin, a class of small redox proteins, could also be found.

Putative electron donors and carbon sources

The genome includes genes for hydrogenase 2 (E.C. 1.12.7.2), which is a membrane-bound Ni-Fe hydrogenase that is believed to reduce menaquinone to menaquinol in a reversible reaction⁵⁸. The genome encodes 15 proteases, of which 4 (AYGZ) respectively 5 (MCS) are predicted to be secreted, and 3 aminopeptidases. Catabolic pathways for glutamine, asparagine, and cysteine are present. The genome encodes multiple ABC transporters for amino acids in general, and in particular for leucine/isoleucine/valine, methionine and toluene. It has transporters for peptides (ABC-type), oligopeptides (appBCD), and a permease for oligopeptides is also present (**Table S2**). Also an ABC-type transporter for polyamines could be detected. The genome features an extremely reduced repertoire of polysaccharide degradation machinery with just 3 glycoside hydrolases belonging to family 57 (α -amylases) (**Table S1**).

Mucispirillum schaedleri has genes for degradation of glycerophosphodiester and glycerol utilization (**Table S1**), including a periplasmic glycerophosphodiester phosphodiesterase (E.C. 3.1.4.46), sn-glycerol-3-phosphate transporter, glycerol kinase, and both aerobic and anaerobic versions of the sn-glycerol-3-phosphate dehydrogenase (E.C. 1.1.1.94, respectively E.C. 1.1.5.3). The anaerobic flavin-dependent sn-glycerol-3-phosphate dehydrogenase is localized in the cytoplasmic membrane and couples oxidation of sn-glycerol-3-phosphate to glycerone phosphate with the reduction of quinone to quinol. Glycerol dehydrogenases are important for phospholipid biosynthesis as well as usage of glycerophosphodiesters as electron donor. The β -oxidation pathway for fatty acid degradation is incomplete. Only 1 out of 7 reactions are present (just a Long-chain fatty acid-CoA ligase E.C. 6.2.1.3 could be detected), which indicates that the pathway is likely absent. Dicarboxylate and C4-dicarboxylate as well as short-chain fatty acid transporters are encoded in the genome (**Table S2**). Additionally, a formate dehydrogenase was detected for reversible NAD-dependent interconversion of formate and CO₂. Because neither hydrogenase 3 nor 4 were detected there is no evidence for the presence of a formate-hydrogen lyase.

Cofactors and vitamins

Unsurprisingly, *Mucispirillum schaedleri* can produce coenzyme A (CoA), which plays a role in the oxidation and biosynthesis of fatty acids and in the TCA cycle. Biotin, a water-soluble B-vitamin that, as a coenzyme, is involved in gluconeogenesis and in the synthesis of isoleucine, valine and fatty acids, can also be produced. It can be synthesized from 7-keto-8-aminopelargonate, from riboflavin and flavin adenine dinucleotide (FAD), an essential flavin cofactor that is involved in a variety of redox reactions, and from thiamin diphosphate, which plays an essential role in energy metabolism as a cofactor of a variety of enzymes like pyruvate dehydrogenase or transketolase. The pathway for *de novo* biosynthesis of coenzyme B12 (cobalamin coenzyme) is incomplete, but *Mucispirillum schaedleri* can synthesize coenzyme B12, the biologically active form of cobalamin, from cobalamin (cyanocobalamin). Although no transporter for cobalamin was detected, coenzyme B12 can probably be synthesized *de novo* from cobinamide via an uncharacterized route^{59,60}.

Other transporters

SecD-SecE and SecY translocases from the Sec translocase-mediated pathway, and TatA and TatC translocases from the twin-arginine translocation (Tat) system for protein translocation across and insertion into membranes⁶¹ were detected (**Table S2**). *Mucispirillum schaedleri* has transporters for molybdate (ABC-type), peptide/nickel (ABC-type), nickel (ABC-type), iron (ABC-type), magnesium, cobalt (ABC-type), cadmium and zinc (**Table S2**). A sodium:proton antiporter was detected and the genome putatively also encodes a biotin transporter, a lipopolysaccharide transporter and a sulfate transporter (**Table S2**). A drug resistance MFS transport protein (drug:H⁺ antiporter-2 family), a putative multidrug-efflux transporter MexB and the multidrug efflux system protein SugE for exporting antibiotics and other cytotoxic substances are also present (**Table S2**).

Storage compounds

Mucispirillum schaedleri appears to be able to produce glycogen as a storage compound, as a glycogen synthase (E.C. 2.4.1.21) and a glycogen phosphorylase (E.C. 2.4.1.1) were detected in the genome. Adjacent to the glycogen synthase there are three glycosyl hydrolases (family 57) that may be involved in glycogen processing. A polyphosphate kinase (Ppk) was detected which enables the organism to synthesize polyphosphate (poly P) from inorganic phosphorus or from the terminal phosphate of ATP. Ppk is involved in growth, motility, quorum sensing, biofilm formation, and virulence^{62,63}. Despite the use of poly P as a potential energy source, it might also play a role in both stress response and pathogenicity. A poly P-AMP-phosphotransferase (PAP), which phosphorylates AMP to ADP using poly P as a substrate, could not be detected.

Motility

The genome encodes more than 80 proteins classified in the COG group Cell Motility (**Table S3**). Besides genes needed for biosynthesis of its flagellum, several chemotaxis-related proteins are present in the genome. The histidine kinase CheA^{64,65,66}, two coupling chemotaxis proteins CheW⁶⁷ – involved in the transmission of signals from the transmembrane chemotaxis receptor proteins to the flagellar motor⁶⁸ – and several putative methyl-accepting chemotaxis proteins (MCP)⁶⁹ could be detected. The PseB (E.C. 4.2.1.115) protein and the pseudaminic acid synthase PseI (E.C. 2.5.1.97), catalyzing the first respectively the last step in the biosynthesis of pseudaminic acid, were detected.

Virus defense (CRISPR)

Mucispirillum schaedleri has a CRISPR/Cas-System with a length of 694 nucleotides, which is identical between the two genomes (**Table S3**). The *cas1*, *cas2* and *cas9* genes were detected, indicating that it is a type II CRISPR/Cas-System⁷⁰. There are 10 spacers, all of them identical between both genomes, and a repeat consensus sequence with a length of 36 nucleotides. We searched for targets of the crRNA spacers, but of 10 spacers only one had a single match, which was *Bacillus thuringiensis* MC28 plasmid pMC189 (NC_018687).

Mobile genetic elements

One intact prophage, including putative head and tail proteins, was detected in the AYGZ genome. All of the predicted phage-like proteins are also present within a region in the MCS genome, but have not been predicted as intact prophage. The prophage region has a size of 134.6 kb and 32% of CDS (43) in this region encode phage-like proteins (**Figure 2, Table S3**).

Consistent with known quality of integration sites, multiple tRNAs and transposases were detected in this region, although an integrase could not be identified. The putative head and tail proteins are located in a region in close vicinity to the CRISPR/Cas-system region.

Putative prophage and phage-like proteins

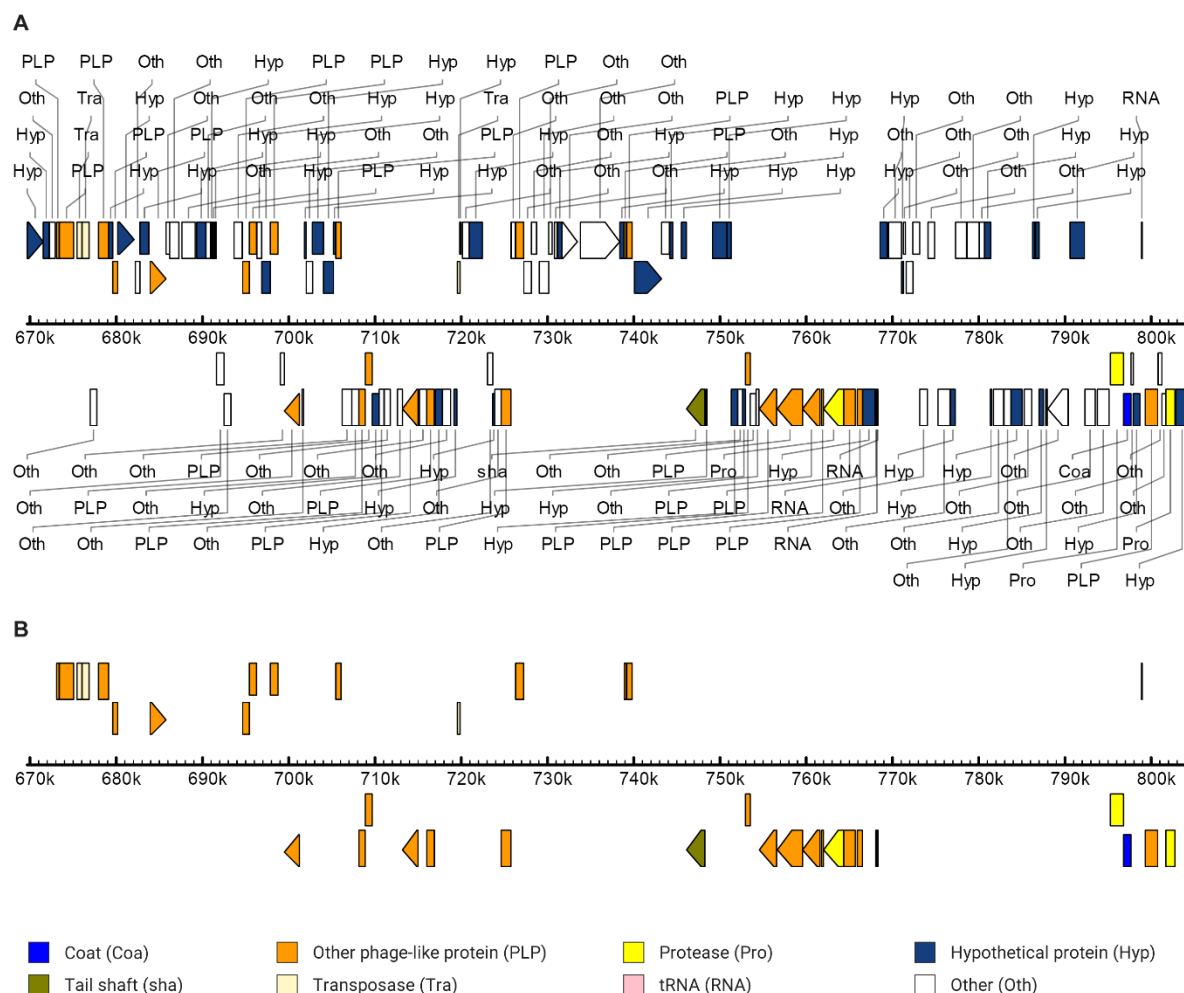


Figure 2: Prophage region predicted by PHAST. **(A)** 32% (43) of all CDSs in the predicted prophage region encode for phage-like proteins. Predicted CDSs of *Mucispirillum schaedleri* are indicated as hypothetical proteins (dark blue) and other proteins (white). **(B)** CDSs annotated as phage-like or phage-associated proteins are transposases (cornsilk), proteases (yellow), tRNAs (rose) and other phage-like proteins (orange). The putative head protein (middle blue on the right side) and tail protein (olive) are located in a region adjacent to the CRISPR/Cas-System region at around 745k.

Additionally, multiple insertion sequences (IS) were identified and were consistent between the two genomes, although the sequence identity was not always identical (**Table S3**). IS families detected include IS3, IS4, IS6, IS21, IS30, IS91, IS481, IS607, and ISL3.

Proteins with eukaryotic-like domains

Ankyrin repeat (ANK) is the most common protein-protein interaction motif and is primarily found in eukaryotic genomes, but proteins with ANK domains are also present in some symbiotic and pathogenic bacteria⁷¹. We detected 10 genes with ANK domains in *Mucispirillum schaedleri* (**Table**

S4). Eleven tetratricopeptide repeat (TPR) containing proteins, that are reported to be related to virulence mechanisms in bacterial pathogens⁷², could also be identified (**Table S4**).

Type VI secretion system

The type VI secretion system (T6SS) is known to consist of 13 conserved core proteins necessary for function³². *Mucispirillum schaedleri* has bacteriophage-like components of the T6SS including hemolysin co-regulated protein (Hcp), valine-glycine repeat protein G (VgrG, putatively), TssB/C, which complex in a way homologous to the T4 sheath⁷³, and TssE, which is homologous to the bacteriophage baseplate protein gp25⁷⁴ (**Table S4**). It also has membrane-associated proteins TssL, TssM, TssJ. Other conserved T6SS proteins with unknown function were also detected: TssA, TssF, TssG, and TssK. Additionally, homologs to other associated proteins were also present, such as ClpV. The T6SS enables *Mucispirillum schaedleri* to export effector proteins into other bacterial and/or eukaryotic cells⁷⁵ and can be used for antagonistic as well as for non-antagonistic purposes⁷⁶. A putative eukaryotic-like phospholipase D protein, which is involved in destabilization of the host cell membrane and in mediation of intra- and interspecies bacterial interactions as effectors of the T6SS⁷⁷, could be identified.

Other secretion, virulence and resistance factors

Hemolysin-3 and a putative colicin V production protein, a bacteriocin produced by many strains of *Escherichia coli*^{78,79}, were detected. A *yoeB* gene encoding YoeB toxin (EC 3.1.-.-) with an addiction module antitoxin RelB/DinJ family protein – DinJ antitoxin of YafQ-DinJ toxin-antitoxin (TA) system – next to it could be found. The YefM antitoxin, part of the YefM-YoeB TA system⁸⁰ and the YafQ toxin, part of the DinJ-YafQ TA system⁸¹, were not detected, indicating that those TA systems are not complete and thus not active in *Mucispirillum schaedleri*.

A virulence-associated protein D (VapD) was also found, but no other *vap* genes. The exact biological role of VapD has not yet been established, but it is known as a toxin in *Haemophilus influenza* and a typical prokaryotic toxin with the activity of an mRNA interferase⁸². *Helicobacter pylori*, an important pathogen, also has solely VapD⁸². In the genome of *Mucispirillum schaedleri* VapD is located between the Tra and Trb loci of a putative type IVA secretion system (T4ASS)⁸³ and several transposases, integrases and tRNA genes. Five Tra conjugal transfer proteins (TraC, TraD, TraF, TraG, TraL), three Trb conjugal transfer proteins (TrbB, TrbD, TrbG) and a VirB complex with a type IV secretion/conjugal transfer ATPase VirB4 family protein and a VirB8 family protein belonging to the putative T4ASS⁸⁴ were detected, which may play a role in horizontal gene transfer⁸³ via conjugation or in pathogenicity^{85,86}.

We also detected parts of a type II secretion system (T2SS) or type IV pilus biogenesis machinery (T4P). The T2SS is widely distributed especially among *Proteobacteria*, most of them extracellular pathogens⁸⁷. Usually it is encoded by at least 12 genes in a single operon⁸⁸. In *Mucispirillum schaedleri* only 5 proteins, the major prepilins T2SC-T2SG, could be detected. Seven of the T2SS core proteins⁸⁹ (T2SH-T2SO) seem to be missing indicating either a non-functioning T2SS system or by contrast the presence of a T4P or DNA uptake⁹⁰ machinery, which consists of homologues of the T2SS protein constituents. Several *Chlamydiae* also have homologues of only T2SC-T2SG genes⁹⁰. Type IV pili are implicated in motility as well as virulence and horizontal gene transfer⁹¹. A putative T4P assembly protein PilZ is located in closer vicinity to the T2SS/T4P genes. A PilT protein proposed to be a force-generating protein for pili retraction⁹¹ was also detected.

Seven genes encoding putative β -lactamases (EC 3.5.2.6) were identified (**Table S4**), which could provide *Mucispirillum schaedleri* with resistance to β -lactam antibiotics.

Putative horizontally transferred genes (HGT)

Phylogenetic trees containing *Mucispirillum schaedleri* could be calculated for most of the genes in both genomes (AYGZ: 1,599 / MCS: 1,715). Normally the amount of newly acquired genes in a bacterial genome is below 15%^{92,93}. Based on automatic selection of trees using PHAT, more than half of the genes have putatively been laterally transferred between *Mucispirillum schaedleri* and bacteria not belonging to the phylum *Deferribacteres* (AYGZ: 841 / MCS: 1,142). During an HGT event new DNA is introduced into a single lineage, whereby the new DNA segment or gene is limited to the recipient strain or species and absent from any other closely related taxa⁹³. In this case, the most recent connecting ancestral node is shared by two genes from different species which are spaced apart in the species tree⁹⁴. This produces a dispersed phylogenetic distribution in the gene tree, the same as we could see for the gene trees of putatively horizontally transferred genes in our *Mucispirillum schaedleri* genomes. Therefore we assume that *Mucispirillum schaedleri* was the recipient of those genes rather than the donor. According to our analysis, many of those putatively horizontally transferred genes originate from *Proteobacteria* (AYGZ: 261 / MCS: 286) and *Firmicutes* (AYGZ: 168 / MCS: 173) (**Table 3, Figure 3A**). *Proteobacteria* are a major phylum within Gram-negative bacteria and include a variety of pathogens like *Helicobacter*. Among *Proteobacteria*, *Epsilonproteobacteria* putatively contributed the majority of genes (AYGZ: 97 / MCS: 110) to the genome (**Table 3, Figure 3B**), with the largest fraction coming from *Campylobacter* and *Helicobacter* (**Figure 3C**). *Firmicutes* are mostly Gram-positive with low GC content. Among *Firmicutes*, *Clostridia* contribute the largest fraction of genes (AYGZ: 108 / MCS: 98), nearly as much as *Epsilonproteobacteria*. Based on analysis of the nearest neighbor within the gene trees, most of the genes within *Clostridia* apparently come from the genera *Eubacterium* and *Clostridium*.

Sources of putative horizontal gene transfer

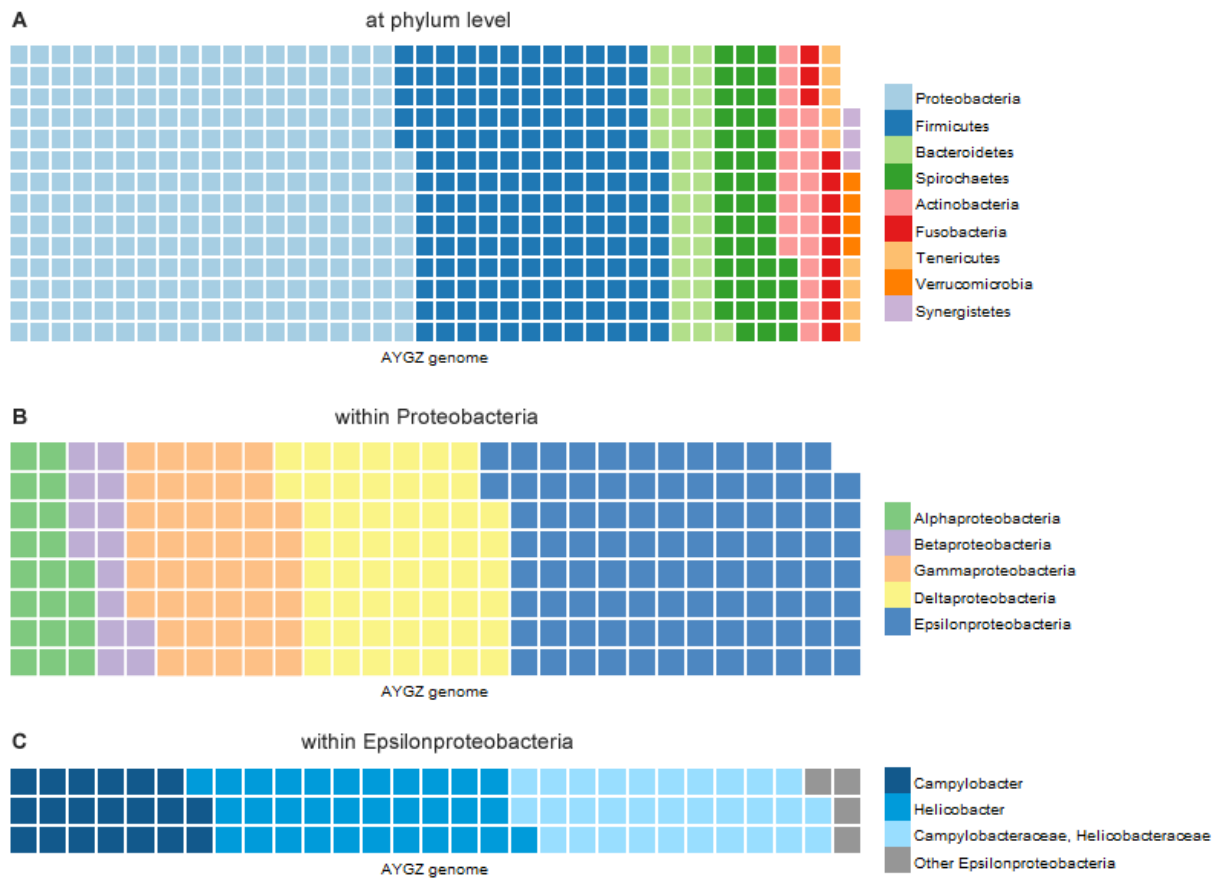


Figure 3: Putative horizontally transferred genes in the *Mucispirillum schaedleri* AYGZ genome automatically selected by PHAT. Each square indicates one gene. **(A)** *Proteobacteria* and *Firmicutes* appear to be the largest sources of most laterally transferred genes in the *Mucispirillum schaedleri* genome. **(B)** Most of the genes in the phylum *Proteobacteria* appear to originate from *Epsilonproteobacteria*. **(C)** Within the *Epsilonproteobacteria* most genes have a nearest neighbor from *Helicobacteraceae* and/or *Campylobacteraceae*, with more than half of the genes being clearly assigned to *Helicobacter* or *Campylobacter*. Many *Mucispirillum schaedleri* genes have sisters from both families and can therefore not clearly be assigned to one or the other.

Table 3: Putative horizontally transferred genes in the *Mucispirillum schaedleri* genomes automatically selected by PHAT. Listed are the number of genes which trees contain nodes with only *Mucispirillum* and a defined taxonomic group. Often a branch contains also other different taxa than the ones specified. Those trees are listed as either being from other taxonomic groups or ambiguous. A large fraction of genes in the genomes putatively have their origin in HGT events, with *Epsilonproteobacteria* and *Clostridia* being the main contributors.

	AYGZ (BI, Broad Institute)	MCS (UV, University of Vienna)
Total # of trees	1,599	1,715
Not Deferribacteres	1,142	841
Proteobacteria	261	286
Firmicutes	168	173
Bacteroidetes	34	48
Spirochaetes	45	50
Actinobacteria	21	8
Fusobacteria	12	13
Tenericutes	9	3
Verrucomicrobia	4	3
Synergistetes	3	5
Other phyla or ambiguous	585	252
<i>Proteobacteria and Firmicutes</i>		
Proteobacteria		
Alphaproteobacteria	20	13
Betaproteobacteria	14	7
Gammaproteobacteria	44	55
Deltaproteobacteria	56	64
Epsilonproteobacteria	97	110
Other Proteobacteria or ambiguous	30	37

Firmicutes

Clostridia	108	98
Bacilli	15	19
Lachnospiraceae	7	7
Ruminococcaceae	5	8
Other Firmicutes or ambiguous	33	41

The T6SS appears to have been laterally transferred from *Epsilonproteobacteria* (*Helicobacter* or *Campylobacter*). Nearly all of the genes belonging to the T6SS share a node with just *Mucispirillum schaedleri* and *Campylobacter* and/or *Helicobacter* in their phylogenetic trees (**Figure S1**). The T6SS synton (minimum 30% identity, matching at least 80% of the gene's length) between *Mucispirillum schaedleri* and *Helicobacter hepaticus* shows an identical order except for two genes fragmented in the MCS genome as compared to the AYGZ genome (**Figure 4**).

Type VI secretion system - synton between *Mucispirillum schaedleri* and *Helicobacter hepaticus*

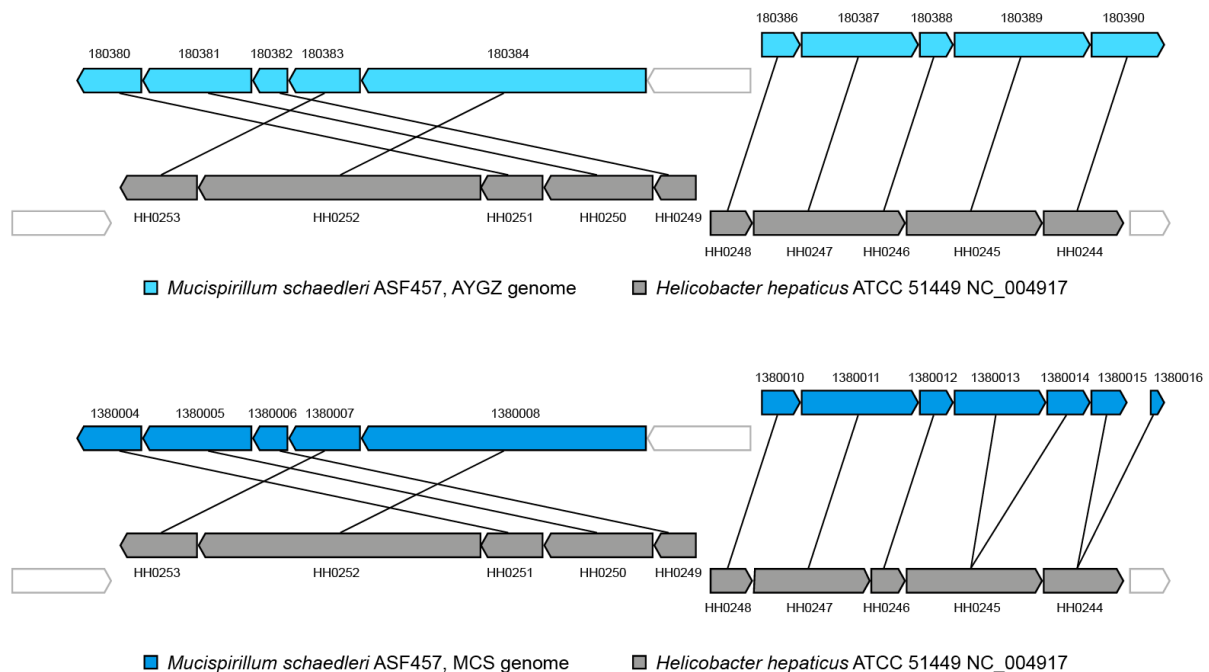


Figure 4: The synton of the type VI secretion system in *Mucispirillum schaedleri* and *Helicobacter hepaticus* ATCC 51449 (NC_004917) shows a perfect alignment and order. The MCS genome has two fragmented genes (AYGZv1_180389 = MCSv1_1380013 + MCSv1_1380014 and AYGZv1_180390 = MCSv1_1380015 + MCSv1_1380016). Most of the genes have between 40% and 50% identity.

Other *Proteobacteria* are probably the source for sn-glycerol-3-phosphate dehydrogenase (*Desulfovibrio* from *Deltaproteobacteria*) and sn-glycerol-3-phosphate transporter (*Proteus* from *Gammaproteobacteria*). Besides these *Mucispirillum schaedleri* putatively acquired several genes involved in virulence, resistance and defense, and mobile genetic elements from other bacteria. The gene of the HlyD family secretion protein, which is involved in the transport of hemolysin A shares a node with *Helicobacter*. Parts of the CRISPR/Cas-system were putatively acquired from *Bacilli* and *Epsilonproteobacteria*, with the CRISPR-associated Csn1 (Cas9) family protein coming from *Staphylococcus* (*Bacilli*) and the CRISPR-associated endonuclease Cas1 protein coming from either *Campylobacter* or *Helicobacter* (*Epsilonproteobacteria*), which also appear to be the origin of the gene for the VapD protein in the *Mucispirillum schaedleri* genome. The Tra conjugal transfer proteins and the VirB complex from a putative T4ASS are related to genes from *Proteobacteria*. Genes involved in resistance appear to have its origin in a wider range of phylogenetic groups with a drug resistance MFS transporter (drug:H⁺ antiporter-2 family) coming from *Bifidobacterium*, a drug/metabolite transporter (DMT family) coming from *Staphylococcus*, a putative multidrug resistance protein MexB coming from *Desulfovibrionaceae*, and a putative β -lactamase coming from *Campylobacter*. The putative methyl-accepting chemotaxis protein is related to genes from *Campylobacter*.

Clostridia appear to be the source of several proteins involved in transport, cofactor biosynthesis, respiration and oxygen stress response. Transport proteins for cobalt are related to those from *Clostridium* as well as the nitroreductase and ruberythrin. A cobalamin synthase putatively comes from *Eubacterium*, the hydrogenase 2 from *Geobacter* and the catalase from *Sphaerochaeta*.

Based on a BLAST best hit analysis, the nearest neighbors of annotated putative horizontally transferred genes from organisms not belonging to *Deferribacteres*, not including proteins of unknown function, belong to the genera *Clostridium* (171), *Campylobacter* (41), *Brachyspira* (36), *Bacillus* (24), *Fusobacterium* (26), *Thermodesulfovibrio* (22), *Streptobacillus* (22), *Fusobacterium* (22), *Arcobacter* (19), *Desulfurobacterium* (18), *Megamonas* (14), *Lawsonia* (14), *Sulfurihydrogenibium* (13), *Desulfurella* (13), *Thermoanaerobacter* (12), *Cyprinus* (12), *Helicobacter* (11), *Thermodesulfobacterium* (10), *Halanaerobium* (10) (**Table S6**). This BLAST analysis does not take into account information about sister clades of *Mucispirillum schaedleri* in the phylogenetic trees as calculated by PhyloGenie, and random manual inspection showed that the nearest neighbors are not always consistent with those in the phylogenetic trees.

Most of the putative HGT genes are classified in the COG classification scheme as replication, recombination and repair (COG category L) with a large fraction coming from *Firmicutes* (**Figure 5**). *Firmicutes* are also by far the largest group putatively contributing to coenzyme transport and metabolism (COG category H) and inorganic iron transport and metabolism (COG category P), whereas *Proteobacteria* appear to be an important source for laterally transferred genes in most of the other COG categories.

Sources of putative HGT grouped by COG category (AYGZ)

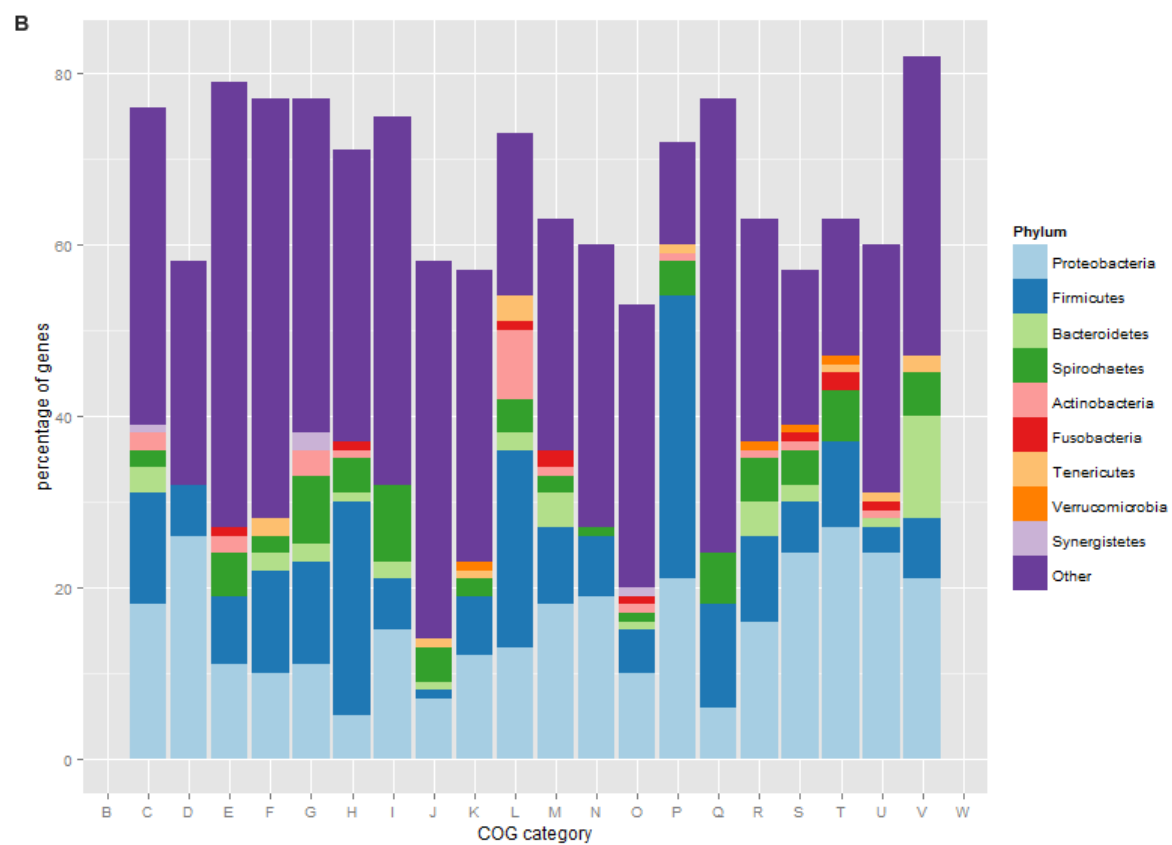
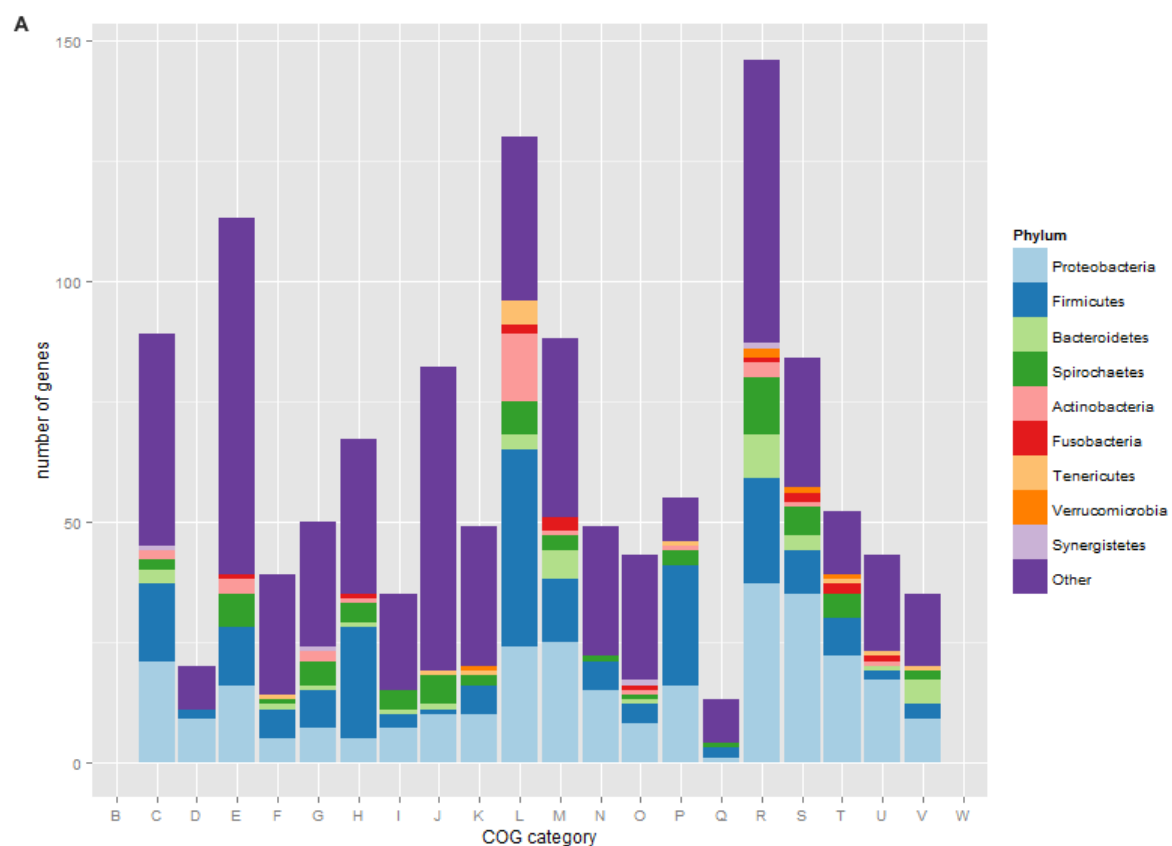


Figure 5: This figure is based on the AYGZ genome and all genes putatively acquired from phyla other than *Deferribacteres*. Colors for phyla are the same as in figure 3A. 'Other' contain all phyla except *Deferribacteres* not explicitly listed. **(A)** In most COG categories *Proteobacteria* form the largest group, followed by *Firmicutes*. Only in category F, G, H, L and P *Firmicutes* contribute most of the genes. **(B)** Putatively transferred genes are normalized for the different number of genes in each COG category and the percentage of genes per category for each phylum is shown. Defense mechanisms (COG category V) appear to be functions mostly influenced by HGT, followed by amino acid transport and metabolism (COG category E).

COG categories are: B = Chromatin structure and dynamics, C = Energy production and conversion, D = Cell cycle control, cell division, chromosome partitioning, E = Amino acid transport and metabolism, F = Nucleotide transport and metabolism, G = Carbohydrate transport and metabolism, H = Coenzyme transport and metabolism, I = Lipid transport and metabolism, J = Translation, ribosomal structure and biogenesis, K = Transcription, L = Replication, recombination and repair, M = Cell wall/membrane/envelope biogenesis, N = Cell motility, O = Posttranslational modification, protein turnover, chaperones, P = Inorganic ion transport and metabolism, Q = Secondary metabolites biosynthesis, transport and catabolism, R = General function prediction only, S = Function unknown, T = Signal transduction mechanisms, U = Intracellular trafficking, secretion, and vesicular transport, V = Defense mechanisms.

Discussion

Neither of the two genomes is closed. The exact number is not known, but the separation of the two strains – *Mucispirillum schaedleri* ASF 457 MCS and *Mucispirillum schaedleri* ASF 457 AYGZ – is believed to have happened many generations in the past. Despite their independent evolution the two genomes are very similar. This indicates that, despite different sequencing technologies, technical idiosyncrasies of the sequencing platforms did not influence the assemblies and our annotation results. Additionally the evolutionary pressure during separate cultivation was obviously similar. It is also mentionable that the number of contigs had no substantial effect on the annotation, although for some single genes in the AYGZ genome only fragments were predicted for the MCS genome. This is likely due to minor sequencing errors like homopolymer errors on the Ion Torrent platform, or due to the location of a single gene on different contigs in the assembly. Most strikingly the spacers in the CRISPR/Cas-system are exactly the same in both organisms, which indicates that there has been no exposure to any novel bacteriophage targeted by the CRISPR/Cas-system during this time period. Most of the shared genes that are not identical are annotated as transposases, transporters, putative proteins or proteins of unknown function.

Lifestyle – possible metabolic strategies

Compared to the number of genes for polysaccharide utilization in the genomes of other human gut commensals⁸ like *Bacteroides thetaiotamicron*, which can also use glycans when polysaccharides are absent from the diet⁹⁵, *Mucispirillum schaedleri* has a very limited repertoire for breaking down polysaccharides with only 3 glycoside hydrolases family 57 (α -amylase). This is a totally unexpected finding for a microorganism that inhabits the mucus layer that consists of the abundant and complex glycoprotein mucin. The genome of *Bacteroides thetaiotaomicron*, which has the largest repertoire of genes involved in metabolism of polysaccharides known so far⁹⁶, has an estimate of more than 270 glycoside hydrolase family genes according to the [CAZY](#) (carbohydrate-active enzymes) database. *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, still has 35 predicted glycoside hydrolases in its genome⁹⁷. As an inhabitant of the mucus layer in the gut it is an unexpected finding that complex compounds like host-derived mucus or dietary polysaccharides that pass the gut cannot be used as the primary energy source by the organism. With complete pathways for the degradation of several amino acids, of acetoacetate to acetate via acetyl-CoA, and of glycerol, the genome predicts that *Mucispirillum schaedleri* rather uses peptides, amino acids, glycerol, and SCFAs as substrates for its energy metabolism, as additionally transporters for these compounds have been detected. It is thus rather a secondary consumer of breakdown products produced by primary hydrolytic/fermentative microorganisms. The uptake of peptides and amino acids can also serve as substrates for anaplerotic reactions as well as for protein synthesis, where energy costs can be reduced.

The genome of *Mucispirillum schaedleri* predicts a fumarate reductase for converting fumarate to succinate. The C4-dicarboxylate transport/antiport system (DcuA/DcuB), which is necessary for the anaerobic respiration with fumarate⁹⁸, is present. It has yet to be shown in physiological experiments, if fumarate is actually used as terminal electron acceptor. This also applies to the function of the formate dehydrogenase that catalyzes the oxidation of formate to carbon dioxide and opens the possibility to use formate as an electron donor.

Additionally, *Mucispirillum schaedleri* can reduce nitrate to ammonia via DNRA⁹⁹. The direct step from nitrite to ammonia via the *nrfA* gene is used for detoxification and/or for energy production¹⁰⁰. *Mucispirillum schaedleri* also has a hydrogenase (*hyb*) and can use hydrogen as electron donor. The nickel-specific ABC transporter from *Escherichia coli*¹⁰¹, that is highly specific and provides Ni²⁺ ions for anaerobic biosynthesis of the hydrogenase, is also present. By metabolizing four H₂ molecules per reaction in the second step of DNRA, which is the reduction of nitrite to ammonia and takes place preferably under nitrate-limited conditions usually predominant in the human large intestine, the overall gas volume in the gut is also reduced¹⁰². During the DNRA process metabolites including nitrite and ammonia are produced that are potentially harmful or toxic to mammalian cells. The biological impact of DNRA on the host has yet to be determined. For *Mucispirillum schaedleri* it is nevertheless a favorable process for energy production under anaerobic conditions and an effective mechanism for scavenging of H₂. As no hydrogenase 3 or 4 were found it is unlikely that *Mucispirillum schaedleri* can produce H₂ itself for use in DNRA. Cross-feeding of hydrogen has a key role in anaerobic ecosystems⁸ and rather than degrading polysaccharides available in the mucus layer itself, *Mucispirillum schaedleri* uses hydrogen that is produced by other polysaccharide-degrading species during fermentation.

As *Mucispirillum schaedleri* can not only use ammonia nitrogen but also non-ammonia nitrogen like amino acids and peptides to satisfy its nitrogen requirements, colocalization studies could provide insight into whether *Mucispirillum schaedleri* is preferably associated with certain polysaccharide-degrading species in the mucus layer, as these species could provide it with nutrients like monosaccharides or amino acids.

The GI tract is a competitive environment for microorganisms. For *Mucispirillum schaedleri*, as an inhabitant of the mucus layer without the genetic potential to degrade mucus, motility and chemotaxis play a crucial role for tracing its energy sources, which are mainly breakdown products produced by other bacterial species. This implies that *Mucispirillum schaedleri* should be able to fight for energy sources using strategies targeted at other microorganisms. The fight for nutrients is a reason for which microorganisms in the colon suppress competing bacteria like *Escherichia coli*^{103,104} and *Clostridium difficile*¹⁰⁵. It could be that *Mucispirillum schaedleri* also uses antagonistic strategies via secreted proteins or toxins targeted at other microorganisms competing for the same nutrients in its niche, although physiological experiments have not been conducted yet. In diverse *Bacteroidetes* for example, a T6SS-like pathway is used for exporting antibacterial effectors targeted against competitor bacteria in the gut⁷⁶. The genome of *Mucispirillum schaedleri* predicts that the organism has the potential for secretion of effector proteins via its T6SS and that it can also produce toxins that could be targeted at other bacteria, although their actual use, targets and functions have yet to be proven in physiological experiments. *Mucispirillum schaedleri* is also capable of chemotaxis and motility, skills that provide it with an advantage in nutrient acquisition. Overall, *Mucispirillum schaedleri* seems to be very well adapted to its environment.

Secretion of proteins

Protein secretion is commonly used by bacterial pathogens for mediating interactions with their hosts¹⁰⁶. The T6SS was originally identified in *Vibrio cholera*¹⁰⁷ and *Pseudomonas aeruginosa*¹⁰⁶, where these genes are required for cytotoxicity against amoeba and mammalian macrophages, whereas mutations in homologs of these genes in other bacterial species are reported to attenuate virulence¹⁰⁷. Proteins secreted by T6SS can also be targeted at other bacteria as antimicrobial

effectors mediating inter-bacterial antagonism. Many bacterial species harboring a T6SS are anaerobic, host-associated commensals or pathogens in densely populated poly-microbial sites like the GI tract. The eukaryotic-like phospholipase D protein we detected, a member of the type VI lipase effector superfamily, primarily targets bacterial membranes and membranes of host cells in the mucosa by degrading its major component phosphatidylethanolamine⁷⁷. It is thought that each T6SS assumes a different role in interactions with other organisms and it is not known yet whether the T6SS can target both prokaryotes and eukaryotes¹⁰⁸. Experimental prove has to be given whether *Mucispirillum schaedleri* uses its T6SS for pathogenicity against other bacteria or for promoting the establishment of a mutualistic, commensal, or pathogenic relationship with its eukaryotic host. The T6SS pathway is distributed throughout all classes of *Proteobacteria* and *Helicobacter hepaticus* has a unique predicted T6SS, that itself has probably been acquired from *Gammaproteobacteria* via HGT³². Its major way of dissemination among bacteria is probably horizontal gene transfer, as phylogenetic analyses have revealed³². Subclasses of the T6SS differ in regulatory and accessory protein content, indicating that it is used for adaption to different functions, hosts and/or environments. The T6SS of *Mucispirillum schaedleri* has probably been laterally transferred from either *Helicobacter* or *Campylobacter*. The gene order is the same as in *Helicobacter hepaticus*¹⁰⁹, a spiral-shaped pathogen with single, bipolar flagella, which also inhabits the mucus layer in the murine intestine and is known to play an important role in the development of severe inflammatory bowel disease (IBD)¹¹⁰. The genome of *Mucispirillum schaedleri* also encodes several putative effector proteins with eukaryotic-like domains, namely ANK repeats and TPR containing proteins that can be used for interaction with the host and may play a role during inflammation.

Defense strategies and mechanisms of resistance to oxidative stress

During intestinal inflammation oxidative stress increases in the mucosa^{111,112}. *Mucispirillum schaedleri* has a system for scavenging oxygen and reactive oxygen species that enables it to survive during inflammation and can explain its persistence and increased relative abundance in the inflamed gut. Besides a superoxide reductase, catalase and a cytochrome c peroxidase, the genome also encodes for rubrerythrin, an oxidative stress response protein that has been detected in the gut microbiome of the piglet gut¹¹³ and functions as a hydrogen peroxidase reductase like it is also used by the gut bacterium *Bacteroides thetaiotamicron* for scavenging hydrogen peroxide¹¹⁴. The genome of *Mucispirillum schaedleri* also encodes multidrug efflux systems and transporters, which provide the organism with resistance to a wide range of chemotherapeutical agents, and several putative β -lactamases for providing resistance to β -lactam antibiotics like ampicillin through hydrolysis. *Mucispirillum schaedleri* is well adapted to the unfavorable conditions during inflammation as it can survive under oxidative stress conditions and has mechanisms for resistance when the host is treated with chemotherapeutical agents or β -lactam antibiotics.

But *Mucispirillum schaedleri* not only has resistance mechanisms to host-derived stressors, it has also adapted to viral attacks. The genome analysis predicts a type II CRISPR/Cas-system, which probably evolved through two transposon insertion events and therefore consists entirely of transposon-derived genes¹¹⁵. Only 1 out of 10 spacers could be identified as signature for a known phage, possibly owing to the limited number of sequenced and documented viruses. Interestingly the CRISPR/Cas-system is located within a region predicted as a prophage. Virion structural proteins, tRNAs and several phage associated proteins were found in this region supporting the evidence that the region has been integrated, although no integrase could be detected. The taxonomy of the

prophage is not known, as several known phages instead of a single one are predicted for the prophage region. This might be due to an unknown phage that has not been sequenced and documented yet or due to a prophage in a defective state. Although the prophage was only predicted for the AYGZ genome, all of the predicted phage-like proteins were also found within a region in the MCS genome, with 100% identity for all but 3 genes that are annotated as transposases. In *Vibrio cholerae*, a bacteriophage was detected that encodes its own CRISPR/Cas-system¹¹⁶ targeting a phage inhibitory chromosomal island on the host. But this seems not to be the case for *Mucispirillum schaedleri* and its prophage as none of the spacers are identical to other locations in the genome. Recently CRISPR sequences in a potential prophage have been identified in *Clostridium difficile*¹¹⁷. It is speculated that the system might be used to prevent superinfection by other phages. Nevertheless, the cause for the location of the CRISPR/Cas-system within the prophage region remains unclear. One possible explanation could be that the prophage had been integrated before the acquisition of the CRISPR/Cas-system and that the prophage has been inactivated and resides now inside the genome in a defective state¹¹⁸. Usually integrated phages are major contributors to the bacterial gene repertoire providing the bacterial host with molecular systems¹¹⁸ involved in secretion¹¹⁹, defense^{120,121} or gene transfer^{122,123}. In line with this hypothesis, several transposases and a putative ANK repeat protein were detected within this region.

Little is known about the viromes in the mammalian gut. It is not yet clear whether and how Lotka-Volterra predator-prey dynamics, where viral and bacterial abundances of interacting populations noticeably oscillate over a period of time with the oscillation peak of viruses lagging slightly behind the oscillation peak of bacteria, operate in the intestine, but different viromes in the human gut may be one explanation for different presence patterns of *Mucispirillum schaedleri* in the intestines of mice and humans. However, analyses of the fecal viromes in humans and gnotobiotic mice suggest that temperate phages dominate and – in contrast to other ecosystems – lytic life cycles and Red Queen dynamics, a constant arms race where bacterial prey populations constantly develop escape strategies to gain reproductive advantage and predating viral populations counter with adaptations, play a minor role¹²⁴. Also a large interpersonal diversity of human gut viruses¹²⁵ subject to rapid evolution¹²⁶ exists, and it is unclear if a distinct core human gut virome exists and if it differs from a core mouse gut virome. As phages are highly selective for different bacterial species, humans might harbor a special virus that keeps *Mucispirillum schaedleri* under control. Additionally diet has an influence on not only the bacterial but also the virus-like particles (VLP) community structure¹²⁵.

Horizontal gene transfer

Horizontal gene transfer is a major source of phenotypic innovation and a way to facilitate niche adaptation. As has been shown recently, inter-phylum HGT occurs more frequently in anaerobic than in aerobic bacteria and it is a major driver of metabolic diversity¹²⁷. In *Mucispirillum schaedleri* it is not the metabolic capacity that got primarily enhanced through lateral gene transfer, but main features favoring its growth and survival in a competitive mammalian gut environment. Genes subject to horizontal gene transfer are involved in interactions with other bacteria or the host (T6SS), resistance and defense (CRISPR, drug resistance), and mobile genetic elements. Laterally transferred pathways like T6SS and glycerol-3-phosphate utilization might be especially important and have facilitated its survival and establishment in the mammalian GI tract. Interestingly also genes involved in chemotaxis, motility and horizontal gene transfer (T4P, T4ASS, Tra conjugal transfer proteins) have putatively been acquired via HGT. T4ASS and T6SS can translocate effector substrates directly into

eukaryotic cells. *Proteobacteria*, one of the phyla common in the human gut¹²⁸, have been the largest phylogenetic group contributing to the gene pool of *Mucispirillum schaedleri*. *Helicobacter* and *Campylobacter* are known gut pathogens. The large fraction of genes acquired from *Helicobacteraceae* and *Campylobacteraceae* and putatively involved in pathogenicity and/or host-interaction indicates a gene repertoire potentially harmful for the host. The role of *Mucispirillum schaedleri* during inflammation has to be studied further. Recent studies have shown that microbiota perturbations and changes in the environment can lead to blooms of low-abundance bacteria and boost HGT¹²⁹. Interestingly the presence of the T6SS in *Helicobacter hepaticus* limits intestinal inflammation¹³⁰. Different subclasses of T6SSs exist fulfilling different functions. If the presence in *Mucispirillum schaedleri*, whose abundance is increased during inflammation and whose T6SS is related to the T6SS in *Helicobacter*, also could have mitigating effects on inflammation has yet to be shown.

It is tempting to say that without features acquired via HGT, *Mucispirillum schaedleri* would not be able to live and survive in the mucus layer of the mammalian GI tract.

Conclusions and outlook

The annotation of the *Mucispirillum schaedleri* genome revealed some unexpected insights like its inability to degrade polysaccharides despite inhabiting the mucus layer of the mammalian GI tract. Besides being able to reduce nitrate to ammonium via DNRA, cross-feeding by using nutrients and substrates produced by other organisms is a part of its metabolic strategy. This leads to the necessity of being able to sense nutrients via chemotaxis and of being motile – features acquired via HGT. Mechanisms of resistance to oxidative stress facilitate its survival during inflammation, while the role of TPR and ANK repeat containing proteins and the effect of the T6SS on the host is unclear. Laterally transferred genes mainly from *Proteobacteria* clearly played a major role in the evolution and lifestyle of this organism.

The reason for the limited indiscernibility in the human GI tract as compared to the murine GI tract – be it dietary impacts, bacterial community structure, VLP community structure, cross-talk with the host or other yet unknown factors – and its role during inflammation have to be studied further and investigated experimentally.

Abstract

Mucispirillum schaedleri inhabits the mucus layer of the mammalian gastrointestinal tract, where it was thought to degrade mucus. We analyzed the genome of *Mucispirillum schaedleri* ASF457 to gain insights into the metabolic capacity and general lifestyle of this Gram-negative bacterium belonging to the phylum *Deferribacteres*. Most surprisingly for an inhabitant of the mucus layer, the genome lacks the potential for degrading polysaccharides. *Mucispirillum schaedleri* rather uses amino acids, peptides, glycerol and short-chain fatty acids for its energy metabolism. Additionally, it can reduce nitrate via dissimilatory nitrate reduction to ammonia, and it has systems for scavenging oxygen and reactive oxygen species that facilitate survival during inflammatory events. *Mucispirillum schaedleri* is motile and is capable of chemotaxis. It can secrete various eukaryotic-like proteins as effectors and harbors a type VI secretion system for interaction with other bacteria and/or the host. It also has mechanisms for defense against other bacteria and viruses, an integrated prophage that may now reside in a defective state inside the genome, and several antibiotic and drug resistance genes. To detect lateral gene transfer, we used a phylogenetic approach on a gene basis. Interestingly, more than half of the genes in the genome had putatively been horizontally transferred from other bacteria, with the majority coming from *Proteobacteria*, especially *Deltaproteobacteria* and *Epsilonproteobacteria*, and from *Firmicutes*. Laterally transferred pathways from other gut bacteria like the type VI secretion system, secretion of effector proteins, and the glycerol-3-phosphate utilization may have facilitated its survival and establishment in the mammalian gastrointestinal tract.

Zusammenfassung

Mucispirillum schaedleri bewohnt die Darmschleimschicht im Verdauungstrakt von Säugetieren und es wurde angenommen, dass es dort Schleim abbaut. Wir haben das Genom von *Mucispirillum schaedleri* ASF457 analysiert um Einblicke in die Stoffwechselkapazität und in den potenziellen Lebensstil dieses Gram-negativen Bakteriums, das zum Phylum *Deferribacteres* gehört, zu gewinnen. Für einen Bewohner der Darmschleimhaut ist es überraschend, dass dem Genom das Potenzial zum Abbau von Polysacchariden fehlt. *Mucispirillum schaedleri* verwendet vielmehr Aminosäuren, Peptide, Glycerol und kurzkettige Fettsäuren für seinen Energiestoffwechsel. Zusätzlich kann es Nitrat über die dissimilatorische Nitratreduktion zu Ammonium reduzieren, und es verfügt über Mechanismen zum Abfangen von Sauerstoff und reaktiven Sauerstoffspezies, welche während Entzündungsprozessen im Darm das Überleben von *Mucispirillum schaedleri* ermöglichen. *Mucispirillum schaedleri* ist beweglich und zur Chemotaxis befähigt. Es kann verschiedene Eukaryoten-ähnliche Proteine als Effektoren sekretieren und besitzt ein Typ VI Sekretionssystem um mit anderen Bakterien und/oder dem Wirt zu interagieren. Auch hat es Mechanismen zur Verteidigung gegen andere Bakterien und Viren, einen integrierten Prophagen, der sich nun in einem unvollkommenen Zustand im Genom befinden dürfte, und es hat verschiedene Gene für Resistenzen gegen Antibiotika und Chemotherapeutika. Um horizontalen Gentransfer zu untersuchen haben wir einen phylogenetischen Ansatz auf Basis der einzelnen Gene gewählt. Interessanterweise wurde mehr als die Hälfte der Gene im Genom horizontal von anderen Bakterien übertragen, die meisten kamen dabei von *Proteobacteria*, insbesondere *Deltaproteobacteria* und *Epsilonproteobacteria*, und *Firmicutes*. Die horizontal übertragenen Gensätze für z. B. das Typ VI Sekretionssystem, die Sekretion von Effektor-Proteinen, und die Nutzung von Glycerol könnten sein Überleben und die Etablierung im Darmtrakt von Säugetieren ermöglicht haben.

References

1. Robertson, B. R. *et al.* Mucispirillum schaedleri gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *International Journal of Systematic and Evolutionary Microbiology* **55**, 1199–1204 (2005).
2. Bergstrom, K. S. B. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* **23**, 1026–1037 (2013).
3. Zhou, J. S., Gopal, P. K. & Gill, H. S. Potential probiotic lactic acid bacteria Lactobacillus rhamnosus (HN001), Lactobacillus acidophilus (HN017) and Bifidobacterium lactis (HN019) do not degrade gastric mucin in vitro. *International Journal of Food Microbiology* **63**, 81–90 (2001).
4. Hoskins, L. C. & Boulding, E. T. Mucin Degradation in Human Colon Ecosystems. *Journal of Clinical Investigation* **67**, 163–172 (1981).
5. Berry, D. *et al.* Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4720–5 (2013).
6. Young, K. D. The selective value of bacterial shape. *Microbiology and molecular biology reviews : MMBR* **70**, 660–703 (2006).
7. Ley, R. E., Lozupone, C. a, Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature reviews Microbiology* **6**, 776–788 (2008).
8. Flint, H. J., Bayer, E. a, Rincon, M. T., Lamed, R. & White, B. a. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature reviews Microbiology* **6**, 121–131 (2008).
9. Scupham, A. J., Patton, T. G., Bent, E. & Bayles, D. O. Comparison of the cecal microbiota of domestic and wild turkeys. *Microbial Ecology* **56**, 322–331 (2008).
10. Li, R. W. *et al.* Alterations in the Porcine Colon Microbiota Induced by the Gastrointestinal Nematode Trichuris suis. *Infection and Immunity* **80**, 2150–2157 (2012).
11. Hildebrand, F. *et al.* Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome biology* **14**, R4 (2013).
12. Schauer, C., Thompson, C. L. & Brune, A. The bacterial community in the gut of the cockroach Shelfordella lateralis reflects the close evolutionary relatedness of cockroaches and termites. *Applied and Environmental Microbiology* **78**, 2758–2767 (2012).
13. Liu, J., Xu, T., Zhu, W. & Mao, S. High-grain feeding alters caecal bacterial microbiota composition and fermentation and results in caecal mucosal injury in goats. *British Journal of Nutrition* **112**, 416–427 (2014).

14. Suchodolski, J. & Simpson, K. Canine gastrointestinal microbiome in health and disease. *Veterinary Focus* **23**, 22–28 (2013).
15. Berry, D., Schwab, C. & Milinovich, G. Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *The ISME Journal* 2091–2106 (2012). doi:10.1038/ismej.2012.39
16. Krych, L., Hansen, C. H. F., Hansen, A. K., van den Berg, F. W. J. & Nielsen, D. S. Quantitatively Different, yet Qualitatively Alike: A Meta-Analysis of the Mouse Core Gut Microbiome with a View towards the Human Gut Microbiome. *PLoS ONE* **8**, (2013).
17. Harrell, L. *et al.* Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. *PLoS ONE* **7**, (2012).
18. Vereecke, L. *et al.* A20 controls intestinal homeostasis through cell-specific activities. *Nature Communications* **5**, 5103 (2014).
19. Rooks, M. G. *et al.* Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *The ISME journal* **8**, 1403–17 (2014).
20. Dewhirst, F. E. *et al.* Phylogeny of the defined murine microbiota: Altered Schaedler flora. *Applied and Environmental Microbiology* **65**, 3287–3292 (1999).
21. Schaedler, R. W., Dubos, R. & Costello, R. The Development of the Bacterial Flora in the Gastrointestinal Tract of Mice. *The Journal of experimental medicine* **122**, 59–66 (1965).
22. Orcutt, R. P., Gianni, F. J. & Judge, R. J. Development of an ‘Altered Schaedler Flora’ for NCI gnotobiotic rodents. *Microecol. Ther* **17**, 59 (1987).
23. Wannemuehler, M. J., Overstreet, A., Ward, D. V & Phillips, J. Draft Genome Sequences of the Altered Schaedler Flora , a Defined Bacterial Community from Gnotobiotic Mice. *Genome Announcements* **2**, 1–2 (2014).
24. Broad Institute (broadinstitute.org). Genomic approaches to characterize a highly defined microbial community in gastrointestinal health and disease initiative. (2013). at <https://olive.broadinstitute.org/strains/muci_scha_asf457.1>
25. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research* **25**, 0955–964 (1997).
26. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic acids research* **35**, 3100–8 (2007).
27. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).
28. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **3**, 1–4 (2007).

29. Vallenet, D. *et al.* MicroScope - An integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research* **41**, 636–647 (2013).
30. Vallenet, D. *et al.* MaGe: A microbial genome annotation system supported by synteny results. *Nucleic Acids Research* **34**, 53–65 (2006).
31. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **27**, 29–34 (1999).
32. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y. & Attree, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC genomics* **10**, 104 (2009).
33. Bland, C. *et al.* CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC bioinformatics* **8**, 209 (2007).
34. Edgar, R. C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC bioinformatics* **8**, 18 (2007).
35. Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. & Brown, C. M. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology* **10**, 817–27 (2013).
36. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research* **34**, D32–D36 (2006).
37. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC bioinformatics* **10**, 421 (2009).
38. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST: A Fast Phage Search Tool. *Nucleic Acids Research* **39**, 1–6 (2011).
39. Sigrist, C. J. a *et al.* New and continuing developments at PROSITE. *Nucleic Acids Research* **41**, 1–4 (2013).
40. Sigrist, C. J. a *et al.* PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in bioinformatics* **3**, 265–274 (2002).
41. De Castro, E. *et al.* ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34**, 362–365 (2006).
42. Jehl, M. A., Arnold, R. & Rattei, T. Effective-A database of predicted secreted bacterial proteins. *Nucleic Acids Research* **39**, 591–595 (2011).
43. Frickey, T. & Lupas, A. N. PhyloGenie: Automated phylome generation and analysis. *Nucleic Acids Research* **32**, 5231–5238 (2004).
44. The R Core Team. R: A language and environment for statistical computing. *R: A language and environment for statistical computing*. (2015). at <<http://www.r-project.org>>
45. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).

46. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
47. Bergonzelli, G. E. *et al.* Cloning and Characterization of *Helicobacter pylori* Succinyl CoA : Acetoacetate CoA-transferase , a Novel Prokaryotic Member of the CoA-transferase Family *. **272**, 25659–25667 (1997).
48. Phibbs, P. V. & Bernlohr, R. W. Purification, properties, and regulation of glutamic dehydrogenase of *Bacillus licheniformis*. *Journal of Bacteriology* **106**, 375–385 (1971).
49. García-Domínguez, M., Reyes, J. C. & Florencio, F. J. Purification and characterization of a new type of glutamine synthetase from cyanobacteria. *European journal of biochemistry / FEBS* **244**, 258–264 (1997).
50. Wen, Z. T., Peng, L. & Morrison, M. The glutamine synthetase of *Prevotella bryantii* B14 is a family III enzyme (GlnN) and glutamine supports growth of mutants lacking glutamate dehydrogenase activity. *FEMS Microbiology Letters* **229**, 15–21 (2003).
51. Pengpeng, W. & Tan, Z. Ammonia assimilation in rumen bacteria: a review. *Animal biotechnology* **24**, 107–28 (2013).
52. Biegel, E., Schmidt, S., González, J. M. & Müller, V. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cellular and Molecular Life Sciences* **68**, 613–634 (2011).
53. Pitcher, R. S. & Watmough, N. J. The bacterial cytochrome cbb3 oxidases. *Biochimica et Biophysica Acta - Bioenergetics* **1655**, 388–399 (2004).
54. Bertini, I., Cavallaro, G. & Rosato, A. Cytochrome c: Occurrence and functions. *Chemical Reviews* **106**, 90–115 (2006).
55. Colburn-Clifford, J. & Allen, C. A cbb(3)-type cytochrome C oxidase contributes to *Ralstonia solanacearum* R3bv2 growth in microaerobic environments and to bacterial wilt disease development in tomato. *Molecular plant-microbe interactions : MPMI* **23**, 1042–1052 (2010).
56. Preisig, O., Anthamatten, D. & Hennecke, H. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3309–3313 (1993).
57. Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W. & Brugère, J.-F. Archaea and the human gut: New beginning of an old story. *World journal of gastroenterology : WJG* **20**, 16062–16078 (2014).
58. Pinske, C. *et al.* Physiology and bioenergetics of [NiFe]-hydrogenase 2-catalyzed H₂-consuming and H₂-producing reactions in *Escherichia coli*. *Journal of bacteriology* **197**, 296–306 (2015).
59. Ford, J. E., Holdsworth, E. S. & Kon, S. K. The biosynthesis of vitamin B12-like compounds*. *Biochemical Journal* **59**, 86–93 (1955).

60. Lawrence, J. G. & Roth, J. R. Evolution of coenzyme B12 synthesis among enteric bacteria: Evidence for loss and reacquisition of a multigene complex. *Genetics* **142**, 11–24 (1996).
61. Yuan, J., Zweers, J. C., Van Dijk, J. M. & Dalbey, R. E. Protein transport across and into cell membranes in bacteria and archaea. *Cellular and Molecular Life Sciences* **67**, 179–199 (2010).
62. Brown, M. R. W. & Kornberg, A. The long and short of it - polyphosphate, PPK and bacterial survival. *Trends in Biochemical Sciences* **33**, 284–290 (2008).
63. Rashid, M. H. *et al.* Polyphosphate kinase is essential for biofilm development, quorum sensing, and virulence of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9636–9641 (2000).
64. Liu, Y., Levit, M., Lurz, R., Surette, M. G. & Stock, J. B. Receptor-mediated protein kinase activation and the mechanism of transmembrane signaling in bacterial chemotaxis. *The EMBO journal* **16**, 7231–7240 (1997).
65. Borkovich, K. a, Kaplan, N., Hess, J. F. & Simon, M. I. Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 1208–1212 (1989).
66. Ninfa, E. G., Stocke, A., Mowbray, S. & Stock, J. Reconstitution of the bacterial chemotaxis signal transduction system from purified components. *Journal of Biological Chemistry* **266**, 9764–9770 (1991).
67. Boukhvalova, M. S., Dahlquist, F. W. & Stewart, R. C. CheW binding interactions with CheA and Tar. Importance for chemotaxis signaling in *Escherichia coli*. *Journal of Biological Chemistry* **277**, 22251–22259 (2002).
68. Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nature Reviews Molecular cell biology* **5**, 1024–1037 (2004).
69. Gestwicki, J. E. & Kiessling, L. L. Inter-receptor communication through arrays of bacterial chemoreceptors. *Nature* **415**, 81–84 (2002).
70. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature reviews Microbiology* **9**, 467–477 (2011).
71. Al-Khodor, S., Price, C. T., Kalia, A. & Abu Kwaik, Y. Functional diversity of ankyrin repeats in microbial proteins. *Trends in Microbiology* **18**, 132–139 (2010).
72. Cervený, L. *et al.* Tetratricopeptide repeat motifs in the world of bacterial pathogens: Role in virulence mechanisms. *Infection and Immunity* **81**, 629–635 (2013).
73. Cascales, E. & Cambillau, C. Structural biology of type VI secretion systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**, 1102–1111 (2012).
74. Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).

75. Silverman, J. M., Brunet, Y. R., Cascales, E. & Mougous, J. D. Structure and Regulation of the Type VI Secretion System. *Annual Review of Microbiology* **66**, 453–472 (2012).
76. Russell, A. B. *et al.* A Type VI Secretion-Related Pathway in Bacteroidetes Mediates Interbacterial Antagonism. *Cell Host and Microbe* **16**, 227–236 (2014).
77. Russell, A. B. *et al.* Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**, 508–12 (2013).
78. Smith, H. W. A search for transmissible pathogenic characters in invasive strains of *Escherichia coli*: the discovery of a plasmid-controlled toxin and a plasmid-controlled lethal character closely associated, or identical, with colicine V. *Journal of general microbiology* **83**, 95–111 (1974).
79. Smith, H. W. & Huggins, M. B. Further observations on the association of the colicine V plasmid of *Escherichia coli* with pathogenicity and with survival in the alimentary tract. *Journal of general microbiology* **92**, 335–350 (1976).
80. Nieto, C. *et al.* The yefM-yoeB toxin-antitoxin systems of *Escherichia coli* and *Streptococcus pneumoniae*: Functional and structural correlation. *Journal of Bacteriology* **189**, 1266–1278 (2007).
81. Motiejunaite, R., Armalyte, J., Markuckas, A. & Sužiedeliene, E. *Escherichia coli* dinJ-yafQ genes act as a toxin-antitoxin module. *FEMS Microbiology Letters* **268**, 112–119 (2007).
82. Kwon, A. R. *et al.* Structural and biochemical characterization of HP0315 from *Helicobacter pylori* as a VapD protein with an endoribonuclease activity. *Nucleic Acids Research* **40**, 4216–4228 (2012).
83. Glöckner, G. *et al.* Identification and characterization of a new conjugation/type IVA secretion system (trb/tra) of *Legionella pneumophila* Corby localized on two mobile genomic islands. *International Journal of Medical Microbiology* **298**, 411–428 (2008).
84. Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S. & Cascales, E. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annual review of microbiology* **59**, 451–485 (2005).
85. O’Cellaghan, D. *et al.* A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion systems is essential for intracellular survival of *Brucella suis*. *Molecular Microbiology* **33**, 1210–1220 (1999).
86. Cascales, E. & Christie, P. J. The versatile bacterial type IV secretion systems. *Nature reviews Microbiology* **1**, 137–149 (2003).
87. Sandkvist, M. Type II Secretion and Pathogenesis MINIREVIEW Type II Secretion and Pathogenesis. **69**, 3523–3535 (2001).
88. Korotkov, K. V., Sandkvist, M. & Hol, W. G. J. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature Reviews Microbiology* **10**, 336–351 (2012).
89. Cianciotto, N. P. Type II secretion: A protein secretion system for all seasons. *Trends in Microbiology* **13**, 581–588 (2005).

90. Peabody, C. R. *et al.* Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* **149**, 3051–3072 (2003).
91. Merz, a J., So, M. & Sheetz, M. P. Pilus retraction powers bacterial twitching motility. *Nature* **407**, 98–102 (2000).
92. Koonin, E. V, Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes. *Annual review of microbiology* **55**, 709–742 (2001).
93. Ochman, H., Lawrence, J. G. & Groisman, E. a. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
94. Ravenhall, M., Škunca, N., Lassalle, F. & Dessimoz, C. Inferring Horizontal Gene Transfer. *PLOS Computational Biology* **11**, e1004095 (2015).
95. Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science (New York, N.Y.)* **299**, 2074–2076 (2003).
96. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science (New York, N.Y.)* **307**, 1955–1959 (2005).
97. Van Passel, M. W. J. *et al.* The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS ONE* **6**, (2011).
98. Simon, R. G. *et al.* Transport of C 4 -Dicarboxylates in Wolinella succinogenes. *Society* **182**, 5757–5764 (2000).
99. An, S. & Gardner, W. S. Dissimilatory nitrate reduction to ammonium (DNRA) as a nitrogen link, versus denitrification as a sink in a shallow estuary (Laguna Madre/Baffin Bay, Texas). *Marine Ecology Progress Series* **237**, 41–50 (2002).
100. Tiso, M. & Schechter, A. N. Nitrate Reduction to Nitrite, Nitric Oxide and Ammonia by Gut Bacteria under Physiological Conditions. *Plos One* **10**, e0119712 (2015).
101. Eitinger, T. & Mandrand-Berthelot, M. a. Nickel transport systems in microorganisms. *Archives of microbiology* **173**, 1–9 (2000).
102. Parham, N. J. & Gibson, G. R. Microbes involved in dissimilatory nitrate reduction in the human large intestine. *FEMS Microbiology Ecology* **31**, 21–28 (2000).
103. Freter, R., Brickner, H., Botney, M., Cleven, D. & Aranki, A. Mechanisms That Control Bacterial Populations in Continuous-Flow Culture Models of Mouse Large Intestinal Flora. *Infection and Immunity* **39**, 676–685 (1983).
104. Guiot, H. F. Role of competition for substrate in bacterial antagonism in the gut. *Infection and Immunity* **38**, 887–892 (1982).
105. Wilson, K. H. & Perini, F. Role of competition for nutrients in suppression of Clostridium difficile by the colonic microflora. *Infection and Immunity* **56**, 2610–2614 (1988).
106. Mougous, J. D. *et al.* A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science (New York, N.Y.)* **312**, 1526–1530 (2006).

107. Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1528–1533 (2006).
108. Sarris, P. & Trantas, E. Phytobacterial Type VI secretion system-gene distribution, phylogeny, structure and biological functions. *Plant Pathology* (2012). at http://cdn.intechopen.com/pdfs/34840/InTech-Phytobacterial_type_vi_secretion_system_gene_distribution_phylogeny_structure_and_biological_functions.pdf
109. Fox, J. G. *et al.* *Helicobacter hepaticus* sp. nov., a microaerophilic bacterium isolated from livers and intestinal mucosal scrapings from mice. *Journal of Clinical Microbiology* **32**, 1238–1245 (1994).
110. Cahill, R. J. *et al.* Inflammatory bowel disease : an immunity-mediated condition triggered by bacterial infection with *Helicobacter hepaticus* . Inflammatory Bowel Disease : an Immunity-Mediated Condition Triggered by Bacterial Infection with *Helicobacter hepaticus*. **65**, 3126–3131 (1997).
111. Keshavarzian, a *et al.* Excessive production of reactive oxygen metabolites by inflamed colon: analysis by chemiluminescence probe. *Gastroenterology* **103**, 177–185 (1992).
112. Kruidenier, L. & Verspaget, H. W. Review article: oxidative stress as a pathogenic factor in inflammatory bowel disease--radicals or ridiculous? *Alimentary pharmacology & therapeutics* **16**, 1997–2015 (2002).
113. Poroyko, V. *et al.* Gut microbial gene expression in mother-fed and formula-fed piglets. *PLoS ONE* **5**, (2010).
114. Mishra, S. & Imlay, J. a. An anaerobic bacterium, *Bacteroides thetaiotaomicron*, uses a consortium of enzymes to scavenge hydrogen peroxide. *Molecular Microbiology* **90**, 1356–1371 (2013).
115. Koonin, E. V & Krupovic, M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nature Publishing Group* 1–9 (2014). doi:10.1038/nrg3859
116. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–91 (2013).
117. Sebaihia, M. *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nature genetics* **38**, 779–786 (2006).
118. Bobay, L.-M., Touchon, M. & Rocha, E. P. C. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1405336111– (2014).
119. Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 4154–4159 (2009).

120. Brüssow, H., Canchaya, C., Hardt, W. & Bru, H. Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and molecular biology reviews* **68**, 560–602 (2004).
121. Waldor, M. K. & Friedman, D. I. Phage regulatory circuits and virulence gene expression. *Current Opinion in Microbiology* **8**, 459–465 (2005).
122. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L. & Brüssow, H. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**, 417–424 (2003).
123. Lang, A. S., Zhaxybayeva, O. & Beatty, J. T. Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology* **10**, 472–482 (2012).
124. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
125. Minot, S. *et al.* The human gut virome : Inter-individual variation and dynamic response to diet The human gut virome : Inter-individual variation and dynamic response to diet. 1616–1625 (2011). doi:10.1101/gr.122705.111
126. Minot, S. & Bryson, A. Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 12450–12455 (2013).
127. Caro-quintero, A. & Konstantinidis, K. T. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. *The ISME Journal* **9**, 1–10 (2014).
128. Sekirov, I. & Russell, S. Gut microbiota in health and disease. *Physiological Review* 859–904 (2010). doi:10.1152/physrev.00045.2009.
129. Stecher, B., Maier, L. & Hardt, W.-D. ‘Blooming’ in the gut: how dysbiosis might contribute to pathogen evolution. *Nature Reviews Microbiology* **11**, 277–284 (2013).
130. Chow, J. & Mazmanian, S. K. A pathobiont of the microbiota balances host colonization and intestinal inflammation. *Cell Host and Microbe* **7**, 265–276 (2010).

Supporting information

Comparing insertion sequences in both genomes

After performing a BLAST search (BLASTP, E-value 0.001) of all 94 IS found in AYGZ against all 39 IS found in MCS and extracting the best hits where the query length was equal to the alignment length, it appears that one IS in AYGZ has multiple hits for the same IS in MCS with a relatively high identity (> 98% for most IS). This seems to be the reason for the different number of genes in the IS Finder results.

Supplementary tables

Table S1. Genes encoding proteins of diverse metabolic pathways (carbohydrate metabolism, energy metabolism, lipid metabolism, respiration, fermentation, glycan biosynthesis and metabolism, cofactors and vitamins, amino acid metabolism, nucleotide metabolism).

Table S2. Genes encoding transport proteins.

Table S3. Virus defense (CRISPR), mobile genetic elements (prophage, insertion sequences), and cell motility.

Table S4. Genes encoding proteins for type VI secretion system, β -lactamases, and eukaryotic-like proteins.

Table S5. Selected genes encoding information processing proteins.

Table S6. Putative horizontally transferred genes in the AYGZ genome, listed with evolutionary closest species based on the gene trees and with nearest neighbor including sequence identity based on BLAST best hit analysis.

Supplementary tables can be downloaded from figshare:

<http://dx.doi.org/10.6084/m9.figshare.1517102>

Supplementary figures

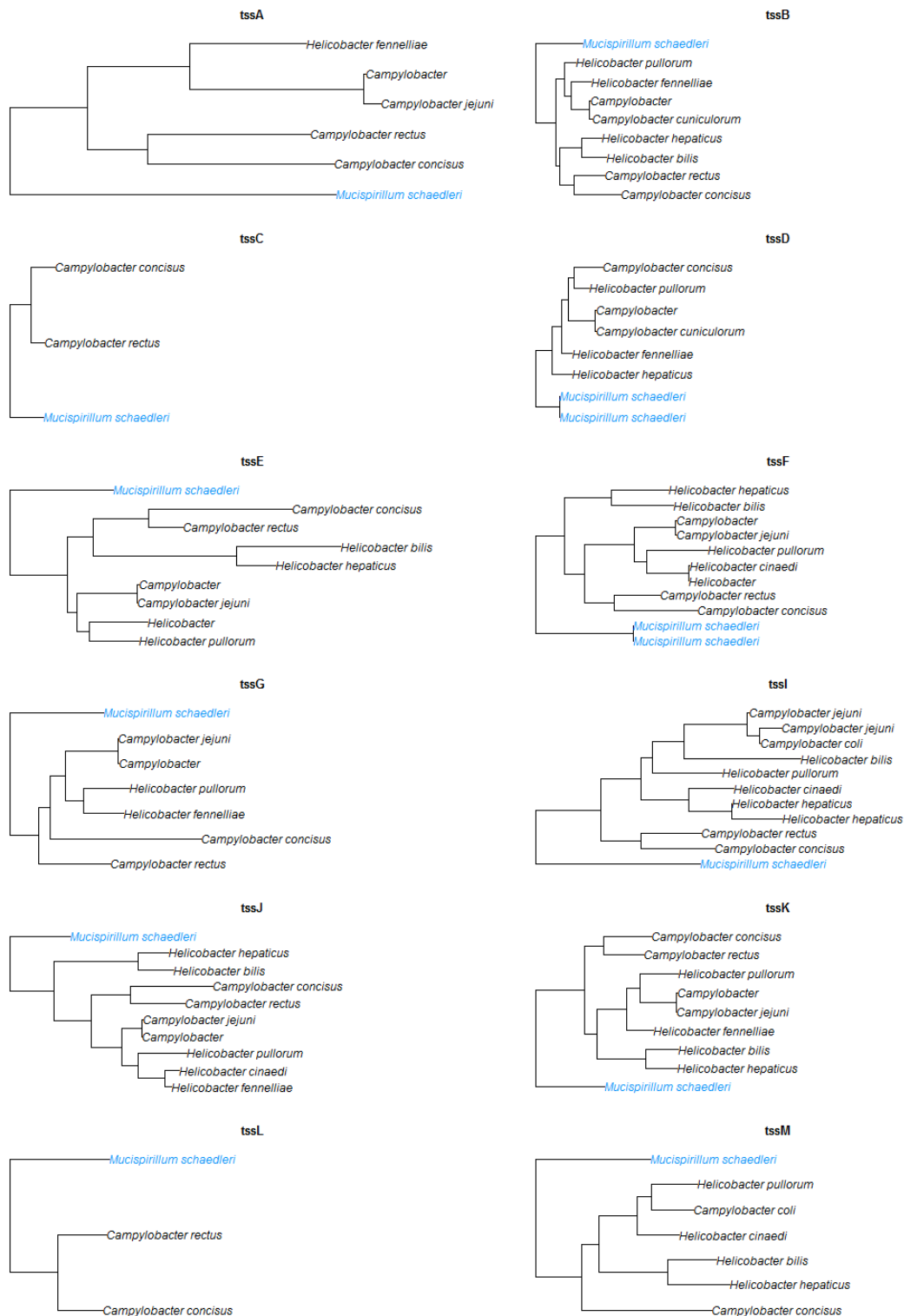


Figure S1: For 12 out of 13 T6SS core genes (all but *tssH*) of the AYGZ genome phylogenetic trees were calculated and nodes containing *Mucispirillum schaedleri* (marked in blue) were extracted. It appears that the T6SS has been laterally transferred from either *Helicobacter* or *Campylobacter*. As some branches contain just *Campylobacter*, but some others also *Helicobacter* it cannot clearly be stated where the T6SS in *Mucispirillum schaedleri* was transferred from.

Acknowledgments

This master thesis was performed at the Department of Microbiology and Ecosystem Science, Division of Microbial Ecology of the Vienna Ecology Center (University of Vienna).

I kindly want to thank:

- David Berry and Alexander Loy for giving me the opportunity to do my master thesis, for offering the flexibility that allowed me to do my thesis while continuing my professional career, and also for their great support and for placing reliance on me.
- David Berry for introducing me into exciting topics during my Großpraktikum and for giving me the opportunity to dig deeper into programming for bioinformatics and into doing reproducible analyses and building pipelines.
- Buck, Orest, Fátima and all the other members of the gut group for helpful and very interesting discussions and for letting me be part of an amazing group.
- All the other members of DOME for their support and for giving me a community feeling although not being together on an everyday basis.

Last but not least I want to thank my family for their great support and for their understanding of my not available spare time. I also want to thank my friends who motivated me and who stayed with me although I could not share a lot of time with them during the last months.

And I want to thank the gut bacteria inside me, whose cells outnumber mine. It was a real team effort ...

Curriculum vitae

Personal data

Carina Pfann BSc	
Date of birth	November 16 th , 1976
Place of birth	Mödling, Austria
Nationality	Austria

Education

1987-1995	Gymnasium Keimgasse, Mödling
1995-1997	Werbe Akademie, Wien
since 2009	Study of biology with focus on microbial ecology and bioinformatics techniques, University of Vienna
since 2014	Master thesis at the Division of Microbial Ecology, University of Vienna

Work experience

1998-2004	Project manager, senior consultant, creative director and founding member at several New Media agencies
since 2004	Self-employed interaction designer and UI & UX designer