universität wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Establishing a Workflow for Exploring the Diversity and Environmental Distribution of *Chlamydiae* "


verfasst von / submitted by

Tamara Halter BSc


angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

| | |
|---|---|
| Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet: | A 066 830 |
| Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet: | Masterstudium Molekulare Mikrobiologie, Mikrobielle Ökologie und Immunbiologie |
| Betreut von / Supervisor: | Univ.-Prof. Dr. Matthias Horn |

“We can only see a short distance ahead,

but we can see plenty there that needs to be done.”

— *Alan Turing*

# ACKNOWLEDGEMENTS

First of all, I would like to thank Michael Wagner for giving me the opportunity to work in his lab and Matthias Horn for welcoming me in his group. A special thanks goes to Astrid Horn for a really great supervision. Thank you for always being patient and encouraging me whenever I felt lost. Of course, I also would like to thank all DOMiES for the nice and familiar working atmosphere and especially the symbiosis group (Allen, Stephan, Paul, Flo W., Vincent, Cecilia, Jasmin, Lena, Gabi) for supporting me and all the nice and productive discussions. Further, I would particularly like to thank Stefano Fazi, Rok Kostanjsek, Julia Vierheilig and Petra Pjevac for providing samples for the project. Finally, I also would like to thank Nadia for a really great friendship and for all the fun we had in the lab and my mom and sister for their unconditional support.

# TABLE OF CONTENTS

# 1 INTRODUCTION

## 1.1 The phylum *Chlamydiae*

The phylum *Chlamydiae* comprises diverse gram-negative, obligate intracellular bacteria that show a characteristic biphasic life cycle (Kuo et al., 2011). In the course of this, elementary bodies (EBs), which are the infective form and only carry out basic transcription and biosynthesis (Omsland et al., 2014) are used to infect host cells (AbdelRahman and Belland, 2005). Once, they have entered the host cells they transform into reticulate bodies (RB), which are the metabolically active stage and start cell division (AbdelRahman and Belland, 2005). Inside the host cells, they form so called inclusions, which are packed with dividing RBs (AbdelRahman and Belland, 2005). Finally, they differentiate into EBs again, lyse their host cells and start to infect new ones (figure 1) (AbdelRahman and Belland, 2005).

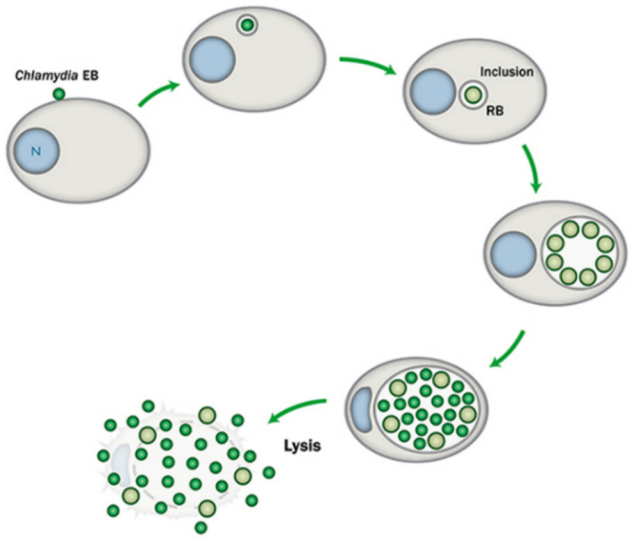



Figure 1: Biphasic life cycle of *Chlamydiae*. Figure adapted from hefty.faculty.ku.edu.

For a long time, the *Chlamydiaceae* were the only known members of the phylum *Chlamydiae* and are so far the most intensively studied group due to their influence towards human health (Horn, 2008). Members of the *Chlamydiaceae* include major human pathogens like *Chlamydia trachomatis*, the causative agent of trachoma and *Chlamydia pneumoniae,* which causes pneumonia (Horn, 2008). However, in the last decade new isolates and molecular data were gained and it became clear that the phylum comprises a huge diversity, currently reflected in ten described chlamydial families including *Rhabdochlamydiaceae*, *Criblamydiaceae* and *Piscichlamydiaceae* (Lagkouvardos et al., 2014) (figure 2). Nevertheless, there are indications that most of the diversity is still undiscovered (Lagkouvardos et al., 2014). In contrast to the

*Chlamydiaceae*, members of those families also referred to as environmental chlamydiae not only infect vertebrates but have a much broader host spectrum including protists and arthropods. Furthermore, they can be found in many different environments including marine sediments, drinking water, wastewater treatment plants and soils (Lagkouvardos et al., 2014). However, *Chlamydiae* are known to be rare in most environments (Viana and Buchrieser, 2016; Pizetti 2016) and thus belong to the rare biosphere (Pedrós-Alió, 2012). The term rare biosphere is used to describe groups of organisms that account for less than 0.1 % of the total community (Pedrós-Alió, 2012). For a long time, those organisms were overlooked and first detected with the rising of high-throughput sequencing techniques (Sogin et al., 2006). Despite their low abundance, members of the rare biosphere have important ecological roles and serve as reservoirs of diversity (Lynch and Neufeld, 2015).

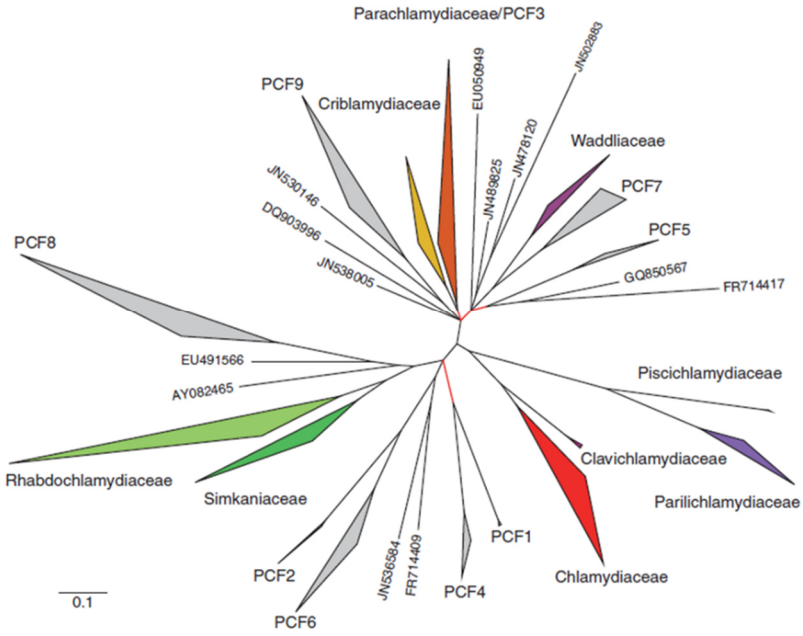



Figure 2: Phylogenetic tree of the phylum *Chlamydiae*. Figure taken from Lagkouvardos, 2014.

Regardless of their ability to infect invertebrate hosts, there are indications that environmental chlamydiae have the potential to infect mammalian cells (Corsaro and Venditti, 2004). But so far there is no consensus in the literature and deeper investigations are necessary to clarify if those members are emerging human pathogens or use mammalian cells only as facultative hosts (Horn, 2008). In addition to their phylogenetic distance and difference in host spectrum, the genomes of members of this largely poorly understood families include numerous genes associated with host-adaptation and ecology (Collingro et al., 2011) compared to the *Chlamydiaceae*.

Apart from the diversity on family level, there are indications that there is even a huge undiscovered diversity within species level (Lagkouvardos et al., 2014). One species suitable to investigate chlamydial microdiversity is *Rhabdochlamydia porcellionis,* which infects the hepatopancreas of the rough woodlouse *Porcellio scaber* (Kostanjsek, 2004). *Porcellio scaber* is commonly found in forests and meadows all over the world and is a very important leaf litter degrader especially in temperate climates (animaldiversity.org). Woodlice mainly feed on litter associated with high microbial evidence (Gunnarsson, 1987). It was previously shown that the microbes serve as a nutrient source (Gunnarsson, 1987) and source of cellulolytic enzymes (Hassall and Jennings, 1975). Like other chlamydiae*, Rhabdochlamydia porcellionis* seems to have a harmful rather than a beneficial effect on their hosts (Kostanjsek et al., 2015). As shown by Kostanjsek et al., 2015, they cause severe tissue damage leading to the death of the hosts in a late stage of infection.

## 1.2 Aims of the project

The main aim of the current study was to establish a molecular workflow to reveal the undiscovered diversity within the phylum *Chlamydiae* and to determine their occurrence in different environments. Using this workflow, environmental samples were screened in order to analyze chlamydial populations with respect to composition and dynamics. The key challenge in establishing the workflow was the low abundance of *Chlamydiae* in the environment. Furthermore, standard bacterial primers could not be used as they have mismatches to some members of the *Chlamydiae*. Thus, the current study is the first one targeting chlamydiae directly in their natural environments.

In order to gain deeper insights into the genomic diversity and evolution of single chlamydial species, an additional workflow was established for a detailed population genetics analysis. The workflow was used to reveal distinctions in the genomes of different *Rhabdochlamydia porcellionis* populations and to investigate the role of *Chlamydiae* as pathogens of isopods and potential vertebrate pathogens (Corsaro and Venditti, 2004). For that purpose, it was necessary to establish protocols for a quick and reliable screening of host populations as well as for the efficient isolation of genomic DNA and cultivation of the *Chlamydiae*. As a model system, the woodlouse *Porcellio scaber* which harbors *Rhabdochlamydia porcellionis* was chosen as the animals can be handled easily in the lab and can be screened for infection by eye (Kostanjsek et al., 2015) in case the animals are highly infected (figure 3). Further, *Rhabdochlamydia porcellionis* can be cultivated in

Sf9 insect cells (Sixt et al., 2013) and *Rhabdochlamydiaceae* is one of the most diverse and largest chlamydial families (Lagkouvardos et al., 2014). However, the presented workflow could also be adapted to other *Chlamydiae* found in invertebrates.


Figure 3: Hepatopancreas observed through the sternites of a highly infected animal. The white nodules (arrows) are a characteristic of infection and a consequence of severe tissue damage. Figure taken from Kostanjsek, 2015.

# 2 MATERIAL AND METHODS

## 2.1 Material

### 2.1.1 Chemicals

| Chemicals | Manufacturer |
|---|---|
| 4',6-Diamidino-2-phenylindole (DAPI) | Lactan GmbH, Graz, Austria |
| 96% Ethanol (extra pure) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| 96% Ethanol, denatured | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Boric acid > 99.8 % p.a., ACS, ISO | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Calcium chloride dihydrate ($CaCl_2$*2 $H_2O$) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Chloroform/Isoamyl alcohol (24:1) | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Citifluor AF1, Glycerol/PBS solution | Agar Scientific Ltd., Stansted, UK |
| DanKlorix | Colgate-Palmolive, Vienna, Austria |
| Di-Sodiumhydrogen phosphate dihydrate ($Na_2HPO_4$*2 $H_2O$) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Ethylenediamine tetra acetic acid disodium salt dihydrate ($Na_2EDTA$*$H_2O$) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Formaldehyde 37% (w/w) Rotipuran® | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Formamide (FA) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Hydrochloric acid 37% (w/w) (HCl) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| LE Agarose | Biozym, Hessisch Oldendorf, Germany |
| Magnesium chloride hexahydrate ($MgCl_2$) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Phenol:Chloroform:Isoamyl alcohol (25:24:1) | Sigma-Aldrich Chemie GmbH, Steinheim, Germany |
| Potassium chloride (KCl) | Sigma-Aldrich Chemie GmbH, Steinheim, Germany |
| Potassium dihydrogen phosphate ($KH_2PO_4$) | Sigma-Aldrich Chemie GmbH, Steinheim, Germany |
| Sodium acetate | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Sodium chloride (NaCl) | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Sodium dodecyl sulfate (SDS) | Carl Roth GmbH & Co.KG, Karlsruhe, Germany |
| Sodium hydrogen carbonate ($NaHCO_3$) | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Sodium hydroxide (NaOH) | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Sucrose | Sigma-Aldrich Chemie GmbH, Steinheim, Germany |
| Tris | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |

### 2.1.2 Consumables

| Consumables | Manufacturer |
|---|---|
| Cell culture flask, 25 $cm^2$ | Thermo Fisher scientific, New York, USA |
| Cover glasses 24 x 50 mm | Carl Roth GmbH & Co. KG, Karlsruhe, Germany |
| Dstroy sticks | Biozym, Hessisch Oldendorf, Germany |
| Eppendorf® LoBind microcentrifuge tubes (1.5 ml, 2 ml) | Eppendorf AG, Hamburg, Germany |
| Greiner 96 well plates, black | Greiner Bio-One GmbH, Frickenhausen, Germany |
| Greiner tubes (50 ml) | Greiner Bio-One GmbH, Frickenhausen, Germany |
| Lysing Matrix A tubes | MP Biomedicals, LLC, CA, USA |
| Manual Filter Tips (various sizes) | Biotix, San Diego, CA, USA |
| Microscope slides, 10 reaction wells | Paul Marienfeld GmbH & Co KG, Lauda-Königshofen, Germany |
| MILLIPORE Express® PLUS (0.22 µm) | Merck KGaA, Darmstadt, Germany |
| Minisart® Syringe Filters (5 µm, 1.2 µm) | Sartorius Stedim Biotech GmbH, Göttingen, Germany |
| Nunclon® Multidishes (various sizes) | Thermo Fisher scientific, New York, USA |
| Omnifix®-F, 1 ml | B. Braun Melsungen AG, Melsungen, Germany |
| PCR 8er SoftStrips (0.2 ml) | Biozym, Hessisch Oldendorf, Germany |
| PCR SoftTubes (0.1 mL, 0.5 ml) | Biozym, Hessisch Oldendorf, Germany |
| Plastic tips (1000 µl) | Biozym, Hessisch Oldendorf, Germany |
| Plastic tips (various sizes) | Carl Roth GmbH & Co KG, Karlsruhe, Germany |
| Reaction tubes (1.5 ml, 2 ml) | Greiner Bio-One GmbH, Frickenhausen, Germany |
| Sealing Mats for 96-Well PCR Plates | Bio-Rad Laboratories GesmbH, Vienna, Austria |
| Sterican® Standardkanülen, 0.90x40 mm | B. Braun Melsungen AG, Melsungen, Germany |
| VWR® 96-Well PCR Plates | VWR International, Vienna, Austria |

### 2.1.3 Media, buffers and solutions

PBS (10x)

| | |
|---|---|
| NaCl | 40.00 g |
| KCl | 1.00 g |
| $Na_2HPO_4$-Dihydrate | 9.00 g |
| $KH_2PO_4$ | 1.20 g |
| $ddH_2O$ | ad 1000 ml |
| pH 7.2-7.4 | |

PBS (1x)

| | |
|---|---|
| 10x PBS | 100 ml |
| $ddH_2O$ | ad 1000 ml |
| pH 7.2-7.4 | |

NaOH [2 N]

| | |
|---|---|
| NaOH | 80.00 g |
| $ddH_2O$ | ad 1000 ml |

NaCl [5 M]

| | |
|---|---|
| NaCl | 292.20 g |
| $ddH_2O$ | ad 1000 ml |

SDS [10% w/v]

| | |
|---|---|
| SDS | 5.00 g |
| $ddH_2O$ | ad 50 ml |

PFA [4 %]

| | |
|---|---|
| Formalin [37%] | 21.60 ml |
| 1x PBS | 178.40 ml |

EDTA [0.5 M]

| | |
|---|---|
| EDTA | 186.10 g |
| $ddH_2O$ | ad 1000 ml |

HCl [1 N]

| | |
|---|---|
| HCl [37 %] | 4.10 ml |
| $ddH_2O$ | ad 50 ml |

Tris/HCl [1 M]

| | |
|---|---|
| Tris | 121.10 g |
| $ddH_2O$ | ad 1000 ml |
| pH 8.0 | |

Cell culture medium

| | |
|---|---|
| L-Glutamine | 2 ml |
| FBS | 20 ml |
| Grace's insect medium | 180 ml |

TBE buffer (10x)

| | |
|---|---|
| Tris [0.9 M] | 107.80 g |
| Boric acid [0.9 M] | 55.00 g |
| $Na_2$-EDTA-Dihydrate | 7.40 g |
| $ddH_2O$ | ad 1000 ml |

TBE buffer (1x)

| | |
|---|---|
| TBE (10x) | 100 ml |
| $ddH_2O$ | ad 1000 ml |

Buffer A

| | |
|---|---|
| Tris-HCl | 0.22 g |
| Sucrose | 4.28 g |
| KCl | 0.09 g |
| $MgCl_2$ | 0.10 g |
| $ddH_2O$ | 50 ml |
| pH 7.5 | |

Buffer A with EDTA

| | |
|---|---|
| Tris-HCl | 0.44 g |
| Sucrose | 8.56 g |
| EDTA | 9.30 g |
| KCl | 0.18 g |
| $MgCl_2$ | 0.20 g |
| $ddH_2O$ | ad 100 ml |
| pH 7.5 | |

Sodium acetate [3 M]

| | |
|---|---|
| Sodium acetate | 24.61 g |
| $ddH_2O$ | ad 100 ml |
| pH 5.0 | |

TE buffer

| | |
|---|---|
| Tris-HCl [1 M] | 1.00 ml |
| EDTA [0.5 M] | 0.20 ml |
| $ddH_2O$ | ad 100 ml |

Lysis buffer

| | |
|---|---|
| Tris-HCl [1 M] | 1.00 ml |
| EDTA [0.5 M] | 0.20 ml |
| NaCl [5 M] | 2.00 ml |
| SDS | 0.50 ml |
| Proteinase K | 5.00 ml |
| $ddH_2O$ | add 100 ml |

DAPI work solution

| | |
|---|---|
| DAPI stock solution | 1 mg/ml |
| 1xPBS | 1:10,000 dilution |

Hybridization buffer [25 %]

| | |
|---|---|
| NaCl [5 M] | 180.00 µl |
| Tris [1 M] | 20.00 µl |
| Formamide | 250.00 µl |
| SDS [10 % v/w] | 1.00 µl |
| $ddH_2O$ | 550 ml |

Washing buffer [25 %]

| | |
|---|---|
| NaCl [5 M] | 1.49 ml |
| Tris [1 M] | 1.00 ml |
| EDTA [0.5 M] | 500 µl |
| $ddH_2O$ | ad 50 ml |

### 2.1.4 Enzymes and Reagents

| Enzymes/Reagents | Manufacturer |
|---|---|
| Bovine Serum Albumin (BSA) | Thermo Fisher scientific, New York, USA |
| DNaseI (1 U/µL) | Thermo Fisher scientific, New York, USA |
| dNTP Mix (10 mM, 2 mM each) | Thermo Fisher scientific, New York, USA |
| DreamTaq Green buffer | Thermo Fisher scientific, New York, USA |
| DreamTaq Green PCR Polymerase | Thermo Fisher scientific, New York, USA |
| EDTA (0.5 M) | Thermo Fisher scientific, New York, USA |
| Fetal bovine serum | PAA Laboratories GmbH, Cölbe, Germany |
| Glycogen | Thermo Fisher scientific, New York, USA |
| Grace's Insect Medium | Thermo Fisher scientific, New York, USA |
| L-glutamine | Sigma-Aldrich Chemie GmbH, Steinheim, Germany |

| | |
|---|---|
| Proteinase K | QIAGEN, Hilden, Germany |
| RNase A (10 mg/mL) | Thermo Fisher scientific, New York, USA |

### 2.1.5 Kits

| Kits | Manufacturer |
|---|---|
| DNeasy Blood and tissue kit | QIAGEN, Hilden, Germany |
| QIAquick PCR Purification Kit | QIAGEN, Hilden, Germany |
| Quant-iT™ PicoGreen® dsDNA Assay Kit | Invitrogen Corporation, Carlsbad, CA, USA |
| ZR-96 DNA Sequencing Clean-up Kit | Zymo Research Corporation, Irvine, CA, USA |

### 2.1.6 Organisms

| Organisms | Reference |
|---|---|
| *Spodoptera frugiperda* Sf9 cells | Vaughn et al., 1977 |

### 2.1.7 FISH probes

| Probe | FA [%] | Target | Sequence 5'-3' | Reference |
|---|---|---|---|---|
| EUB338 I<br>EUB338 II<br>EUB338 III | 0-50 | Most *Bacteria*<br>*Planctomycetales*<br>*Verrucomicrobiales* | GCT GCC TCC CGT AGG AGT<br>GCA GCC ACC CGT AGG TGT<br>GCT GCC ACC CGT AGG TGT | Amann, 1990<br>Daims, 1999<br>Daims, 1999 |
| CHLS-0523 | 25 | *Chlamydiales* | CCT CCG TAT TAC CGC AGC | Poppert, 2002 |
| EUK516 | 0-50 | Most *Eukarya* | ACC AGA CTT GCC CTC C | Amann, 1990 |

### 2.1.8 PCR primers

| Name | *E. coli* position | Target | Sequence 5'- 3' | Reference |
|---|---|---|---|---|
| 16SF/<br>16SR | 42<br>1522 | 16S rDNA<br>*Chlamydiales* | GCG TGG ATG AGG CAT GCA A<br>GGA GGT GAT CCA GCC CCA | chlamydiae.com |
| PanF/<br>PanR | 35<br>1481 | 16S rDNA<br>*Chlamydiales* | CGT GGA TGA GGC ATG CRA GTC G<br>GTC ATC RGC CYY ACC TTV SRC RYY TCT | Corsaro, 2002 |
| 616V/<br>1492R | 8<br>1492 | 16S rDNA *Bacteria* | AGA GTT TGA TYM TGG CTC AG<br>RGY TAC CTT GTT ACG ACT T | Kim, 2009<br>McAllister, 2011 |
| 341F/<br>785R | 341<br>785 | 16S rDNA *Bacteria* | CCT ACG GGA GGC AGC AG<br>CTA CCA GGG TAT CTA ATC C | Juck, 2000 |
| 351F/<br>805R | 351<br>805 | 16S rDNA<br>*Chlamydiae* | GCW GCA GTC GAG RAT YWT TSG C<br>GTR TRC ATM GTT TAM RGC WWG G | This study |

| | | | | |
|---|---|---|---|---|
| 351F/ 805R (Barcoding) | 351<br>805 | 16S rDNA *Chlamydiae* | GCT ATG CGC GAG CTG CGC WGC AGT CGA GRA TYW TTSGC<br>GCT ATG CGC GAG CTG CG TRT RCA TMG TTT AMR GCW WGG | This study |

### 2.1.9 Technical Equipment

| Instrument | Manufacturer |
|---|---|
| Accu-jet® pro pipette aid | BRAND GMBH + CO KG, Wertheim, Germany |
| Avantgarde Incubator | Binder GmbH, Tuttlingen, Germany |
| Biorad T100™ | BIO RAD Laboratories, Inc., UK |
| Centrifuge 5430 R | Eppendorf AG, Hamburg, Germany |
| Centrifuge 5804 R | Eppendorf AG, Hamburg, Germany |
| Confocal Laser Scanning Microscope LSM 510 Met | Carl Zeiss MicroImaging GmbH, Jena, Germany |
| CRUMA-P1 Weighing Cabinet | CRUMA Material de Laboratorio, S.A., Barcelona |
| Epifluorescence microscope Axioplan 2 imaging | Carl Zeiss MicroImaging GmbH, Jena, Germany |
| Eppendorf Research® plus (various sizes) | Eppendorf AG, Hamburg, Germany |
| Holten LaminAir Safety Cabinet (model 1.8, 1.2) | Thermo Fisher scientific, New York, USA |
| Hybridization oven | Memmert GmbH + Co.KG, Germany |
| IKA Genius 3 | IKA®-Werke GmbH & CO. KG, Staufen, Germany |
| Incubation bath GFL 1004 | Gesellschaft für Labortechnik GmbH, Burgwedel, Germany |
| Infinite® F500 microplate reader | Tecan Group AG, Männedorf, Switzerland |
| KERN ABT 120-5DM | KERN & SOHN GmbH, Balingen, Germany |
| Laminar flow hood | |
| Mikro 20 benchtop centrifuge | Andreas Hettich GmbH & Co KG, Tuttlingen, Germany |
| Milli-Q Biocel System Ultrapure Water (MQ) | Merck Millipore, Darmstadt, Germany |
| Nanodrop 1000 Spectrophotometer | Thermo Fisher scientific, New York, USA |
| OHAUS® Analytical Plus balance | Ohaus Corporation, New York, USA |
| pH meter, pH 3110 | Wissenschaftlich-Technische Werkstätten GmbH, Weilheim, Germany |
| RCT Basic IKAMAG® | IKA®-Werke GmbH & CO. KG, Staufen, Germany |
| Sartorius BL 6100 | Sartorius AG, Göttingen, Germany |
| Ultraviolet Sterilizing PCR Workstation | Peqlab Biotechnologie GmbH, Erlangen, Germany |
| UVP UV2 PCR Workstation | Peqlab Biotechnologie GmbH, Erlangen, Germany |
| Vortex-Genie 2 | Scientific Industries Inc., New York, USA |

| | |
|---|---|
| VWR™ Galaxy Mini Microcentrifuge | VWR International, Vienna, Austria |
| VWR™ MiniStar silverline | VWR International, Vienna, Austria |

### 2.1.10 Software

| Software | URL | Reference |
|---|---|---|
| ARB | http://www.arb-home.de/ | Ludwig et al., 2004 |
| Basic Local Alignment Search Tool | http://www.ncbi.nlm.nih.gov/BLAST/ | Altschul et al., 1990 |
| ChromasPro | http://www.technelysium.com.au/chromas.html | Scientific & Educational Software |
| CIPRES | https://www.phylo.org/ | Miller et al., 2010 |
| FigTree | http://tree.bio.ed.ac.uk/software/figtree/ | Institute of Evolutionary Biology, Edinburgh, UK |
| i-controlTM-Microplate Reader Software | http://www.tecan.com/i-control | Tecan Group Ltd., Männedorf, Switzerland |
| IQ-Tree | http://iqtree.cibiv.univie.ac.at/ | Trifinopoulos et al., 2016 |
| Microsoft Office 2016 | https://www.microsoftstore.com/ | Microsoft Corporation, Redmond, USA |
| Ribosomal Database Project | http://rdp.cme.msu.edu/ | Wang et al., 2007<br>Cole et al., 2014 |
| RStudio | https://www.rstudio.com/ | RStudio, Inc., Boston, USA |
| SeaView | http://doua.prabi.fr/software/seaview | Gouy et al., 2010 |

## 2.2 Methods

### 2.2.1 Biogeography of *Chlamydiae*

#### *2.2.1.1 Primer design and evaluation*

Specific rRNA gene targeted primers were designed to be able to detect at least all known members of the phylum *Chlamydiae* and possibly also unknown members in environmental samples. The primers were designed manually using the alignment tool of ARB (Ludwig et al., 2004) and their specificity and sensitivity were checked with RDP Probe Match (Cole et al., 2014). The primers amplify the highly variable regions of the 16S rDNA V3 and V4 (Methé et al., 2012) leading to an amplicon of about 450 bp (figure 4).

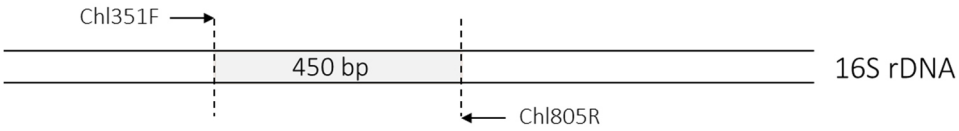



Figure 4: Schematic illustration of the binding positions of the designed primers.

To increase the coverage, the primers are highly wobbled and contain 32 forward- and 64 reverse primer versions. Further, a head sequence was added to the primers that serves as a binding region for barcodes (figure 5) to be able to correctly assign reads to the appropriate samples after multiplexing and MiSeq sequencing.

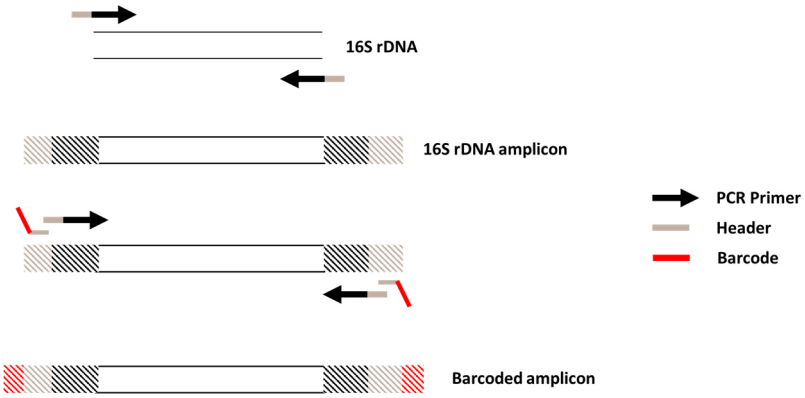



Figure 5: Barcoding for MiSeq amplicon sequencing.

In order to find the best PCR conditions for the primers, they were evaluated using different samples with increasing complexity ranging from pure culture DNA and mock communities to environmental samples. First, the primers were tested with DNA isolated from a *Rhabdochlamydia porcellionis* culture using different annealing temperatures ranging from 65 °C to 50 °C. In addition, all annealing temperatures were tested with DNA of *Verrucomicrobium spinosum* to

exclude unspecific amplification. In a next step, the primers were applied on a genomic DNA and 16S rDNA mock community (table 1). Finally, the primers were tested with an activated sludge sample (VetMed) as well as some lake water ($T5_5$, $T4_2$) and sediment (A, C, E) samples containing relatively low DNA concentrations (table 4). Based on the results of the evaluation, the PCR settings were set as follows: denaturation at 94 °C for 3 minutes, 30 cycles of 30 s at 95 °C, 30 s at 50 °C and 90 s at 72 °C, followed by a terminal extension of 7 minutes at 72 °C. Further, the primer (0.8 µM) and polymerase (2.5 U per 50 µl) concentrations were increased in order to increase sensitivity. To investigate the functionality and applicability of the primers for MiSeq amplicon sequencing, they were applied on a set of 33 environmental samples. For the PCR, the DreamTaq Green PCR Polymerase was used. To minimize the influence of possible inhibitory substances BSA (0.2 µl per 50 µl) was added to the PCR master mixes.

#### *2.2.1.2 MiSeq amplicon sequencing*

##### 2.2.1.2.1 Samples

The samples used for the evaluation were largely provided by Stefano Fazi, Italian Water Research Institute (IRSA-CNR) and included samples from a marine coastal lake in central Italy (Lago di Paola) as well as from the River Po delta. All sampling points were known to be associated with a relatively high abundance of *Chlamydiae* (Pizzetti et al., 2012). The Lago di Paola samples included 20 filters with filtrates of 50–150 ml lake water and isolated DNA from sediment. All samples originated from sampling station SAB 2 (figure 6) and were taken at three different time points (table 2).


Figure 6: Map of Lago di Paola and sampling station SAB2.

The River Po delta samples included isolated DNA from marine sediments of two different time points (June 2013 and October 2014) and 10 different sampling points. The sampling points differ in their geographic location and are associated with different chlamydial abundances (figure 7) (table2). The DNA was isolated with the PowerSoil® DNA Isolation Kit (MO BIO) as specified.

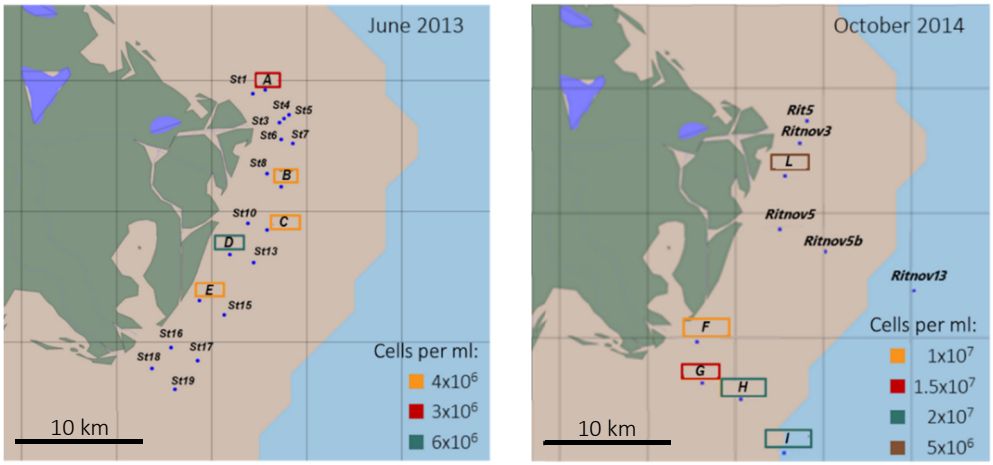



Figure 7: Map of River Po delta. The investigated sampling points are marked. The colors correspond to the chlamydial abundance, which was estimated using Fluorescence In Situ Hybridization.

Furthermore, two activated sludge samples were included in the study. They originated from the University of Veterinary Medicine in Vienna and the wastewater treatment plant of Ingolstadt (Bavaria, Germany). For both sludge samples isolated DNA was provided for the study (table 2). The DNA was extracted as described in Daims et al., 2015. The DNA concentrations of the environmental samples were measured using the Nanodrop 1000 Spectrophotometer.

In addition to the environmental samples, two mock communities (table 1) were included to be able to evaluate the quality of the workflow. For the genomic DNA mock community, isolated DNA from chlamydial EBs and a pure culture of *E. coli* was used. The 16S rDNA mock community consisted of 16S rDNA amplicons, which were amplified using the PANF/PANR and 616V/1492R primers. All samples were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit and pooled accordingly.

Table 1: Composition of the 16S rDNA- and genomic DNA mock community.

| | Percentage | |
|---|---|---|
| Organism | 16Sr DNA mock community | DNA mock community |
| *Escherichia coli* | 89.45 % | 93.95 % |
| *Waddlia chondrophila* | - | 5.00 % |
| *Simkania negevensis* | 10.00 % | 0.50 % |
| *Parachlamydia acanthamoeba* UV7 | 0.50 % | 0.50 % |
| *Protochlamydia amoebophila* E25 | 0.05 % | 0.05 %. |

Table 2: Detailed information about the environmental samples. The cell numbers were quantified by Stefano Fazi using Fluorescence In Situ Hybridization.

| Identifier | Sampling date | Sample type | Sampling point | Cells $ml^{-1}$ | Sampling volume | Barcodes |
|---|---|---|---|---|---|---|
| $T4_1$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 150 ml | GAGTCACT |
| $T4_2$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | AGAGACTG |
| $T4_3$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | ATATGCCG |
| $T4_4$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | CCTACGAA |
| $T4_5$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | CGTAGGAA |
| $T4_6$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | CTCTGACT |
| $T4_7$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | GTCATCAG |
| $T4_8$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | TCACTCTG |
| $T4_9$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 50 ml | CTGACAGT |
| $T5_1$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 115 ml | AGTGACAG |
| $T5_2$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 100 ml | ACCTACCT |
| $T5_3$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 100 ml | CATGTGGT |
| $T5_4$ | 25.09.2013 | water filter | Lago di Paola | n.a. | 100 ml | TTCGTAGG |
| $T5_5$ | 30.10.2013 | water filter | Lago di Paola | n.a. | 50 ml | GTCTTGAG |
| $T5_6$ | 30.10.2013 | water filter | Lago di Paola | n.a. | 50 ml | ACAGAGTG |
| $T5_7$ | 30.10.2013 | water filter | Lago di Paola | n.a. | 50 ml | TTCGAACG |
| $T5_9$ | 30.10.2013 | water filter | Lago di Paola | n.a. | 65 ml | GATGAGGT |
| $T6_1$ | 27.11.2013 | water filter | Lago di Paola | n.a. | 150 ml | CCTACGTT |
| $T6_2$ | 27.11.2013 | water filter | Lago di Paola | n.a. | 150 ml | GAGTTGAG |
| $T6_3$ | 27.11.2013 | water filter | Lago di Paola | n.a. | 150 ml | GGTTCGAT |
| M | 30.10.2013 | sediment | Lago di Paola | n.a. | 1.00 g | ATCGTTGG |
| A | June 2013 | sediment | River Po Delta | $3x10^6$ | 1.00 g | CATCCAAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| B | June 2013 | sediment | River Po Delta | $4x10^6$ | 1.00 g | CGTTGGAT |
| C | June 2013 | sediment | River Po Delta | $4x10^6$ | 1.00 g | GATCTGGT |
| D | June 2013 | sediment | River Po Delta | $6x10^6$ | 1.00 g | GTTCCTTG |
| E | June 2013 | sediment | River Po Delta | $4x10^6$ | 1.00 g | CTAGTGGT |
| F | October 2014 | sediment | River Po Delta | $1x10^7$ | 1.00 g | CAGTCAGT |
| G | October 2014 | sediment | River Po Delta | $1.5x10^7$ | 1.00 g | TAGGAACG |
| H | October 2014 | sediment | River Po Delta | $2x10^7$ | 1.00 g | TCTCCAGT |
| I | October 2014 | sediment | River Po Delta | $2x10^7$ | 1.00 g | AGTCGACT |
| L | October 2014 | sediment | River Po Delta | $5x10^6$ | 1.00 g | AACGTAGG |
| VetMed | 21.01.2015 | activated sludge | Vienna | n.a. | 0.25 g | CTACGTAG |
| Ingo | n.a. | activated sludge | Ingolstadt | n.a. | n.a. | TTCCGGTT |

#### *2.2.1.3 DNA Isolation of Lago di Paola filters*

For the DNA isolation, polycarbonate filters were cut with sterile scissors and transferred into Lysing Matrix A tubes. Afterwards, the filters were incubated with 440 µl lysis buffer at room temperature for 15 minutes. Subsequently, 500 µl phenol:chloroform:isoamyl alcohol (P:C:I) were added, followed by a vortexing step of 2 minutes. The samples were then centrifuged at 13,000 rpm at 4 °C for 2 minutes. The supernatant was treated again with 500 µl P:C:I. Afterwards, the supernatant was recovered with 2.5x volume of ice cold ethanol absolute and incubated for 2 hours with 1 µl sodium acetate and 1 µl glycogen at -20 °C. The samples were centrifuged at 13,000 rpm and 4 °C for 10 minutes. Finally, the pellets were washed with 500 µl 70 % ethanol, centrifuged and resuspended in 100 µl 1x TE buffer.

#### *2.2.1.4 MiSeq amplicon sequencing*

In order to amplify 16S rDNA and to barcode the samples, a two-step PCR approach was carried out for MiSeq amplicon sequencing. For that purpose, the PCR settings described in section 2.2.1.1 were adapted as follows: the first PCR step comprised 25 cycles and in order to decrease the influence of PCR biases triplicates were used for each sample. In the second PCR, 4 µl of the pooled products of the first PCR step were used as a template. Further, the number of cycles was decreased to 5 and per sample duplicates were set up. After the second PCR step the gained products were cleaned-up using the ZR-96 DNA Sequencing Clean-up Kit™ as recommended by the manufacturer. Afterwards, the samples were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit. Finally, the cleaned PCR products were pooled in equimolar amounts in order to

reach 20 × $10^9$ copies of each sample and sent for lllumina MiSeq amplicon sequencing to Microsynth GmbH (Baglach, Switzerland). The sequencing data were processed using the pipeline described by Herbold et al., 2015. For a small subset of samples (VetMed, Lago di Paola $T4_4$ and River Po delta H), the general bacterial primers 785R/341F were applied in addition to the chlamydiae specific primers in order to get an overview of the overall bacterial community. The settings used for PCR were the same as described above and in section 2.2.1.1 but the primer concentration was decreased to 0.05 µM.

*2.2.1.5 Statistical and phylogenetic analysis*

The statistical analyses were carried out using the vegan package in R (Oksanen et al., 2016). In order to visualize a possible influence of time on the chlamydial communities, a NMDS (Non-metric multidimensional scaling) was carried out for the Lago di Paola and River Po delta samples, respectively. For hypothesis tests, perMANOVA (permutational multivariate analysis of variance) was used and the homogeneity of variances was checked using the betadisper function. To evaluate which operational taxonomic units (OTUs) are associated with the time response of the communities, a similarity percentages analysis was carried out using the SIMPER function. An OTU is here defined as a group of closely related sequences that share a 16S rDNA identity of ≥ 97% (Lynch and Neufeld, 2015). For the taxonomic affiliation, the RDP classifier was used (Wang et al., 2007). Additionally, three phylogenetic trees were calculated using different methods. All trees were rooted using *Planctomycetes*, *Verrucomicrobia* and *Lentisphaerae* as outgroup. First, the MiSeq sequences were added to a Bayesian inference tree including 229 full-length chlamydial sequences using the ARB Parsimony (Quick add marked) tool (Ludwig et al., 2004). For that purpose, the sequences obtained in the current study were aligned to the full-length sequences from the Silva and GenBank databases (Benson et al., 2013; Quast et al., 2013) using the alignment tool of ARB (Ludwig et al., 2004). Then, a maximum likelihood tree was calculated using the IQ-Tree webserver and 1,000 ultrafast bootstraps (Trifinopoulos et al., 2016). For the maximum likelihood tree, the multiple sequence alignment was trimmed to the length of the MiSeq sequences. Finally, a Bayesian tree was calculated from the full-length sequences and the short MiSeq sequences using MrBayes at the CIPRES cluster (Miller et al., 2010).

*2.2.1.6 Adaptation and re-evaluation of PCR settings*

Based on the results of the first MiSeq run, the PCR settings and the primers were further adapted, tested on a small subset of samples to keep the effort feasible and sent for MiSeq sequencing

again. In order to enhance the specificity of the primers the annealing temperature was increased. In total, three different annealing temperatures were tested using the River Po delta A sample and the Lago di Paola sample $T6_3$. These samples were chosen as they were associated with a high number and diversity of PVC (*Planctomycetes*, *Verrucomicrobia* and *Chlamydiae*) superphylum reads in the first MiSeq analysis. The samples were tested using an annealing temperature of 55 °C, 58 °C and 61 °C in the first PCR, respectively. In the second PCR the annealing temperature was set to 55° C for all samples. As low PCR product concentrations were expected the number of replicates was increased to four replicates per sample in the second cycle in order to achieve enough product for sequencing. All other settings were used as described in section 2.2.1.4.

In order to increase the PCR product concentration, the number of used cycles was increased and tested using the Lago di Paola sample $T4_5$ and $T6_2$. These samples were chosen as they represent the samples with the highest ($T4_5$) and lowest ($T6_2$) observed read numbers. In total, two different cycle numbers were tested using an annealing temperature of 50 °C and 58 °C, respectively. An overview of the whole setup is shown in table 3. To reduce the influence of possible PCR biases three replicates were pooled in the second cycle. The other settings were used as described in section 2.2.1.4.

Table 3: Settings used to evaluate increased cycle number. The declared number of cycles represent the total number of cycles. In the first cycle 30 and 35 cycles were used, respectively.

| | 1. Cycle | 2. Cycle | 1. Cycle | 2. Cycle |
|---|---|---|---|---|
| Temperature / No. of cycles | 50 °C | 50 °C | 58 °C | 55 °C |
| 35x | $T4_5$, $T6_2$ | | $T4_5$, $T6_2$ | |
| 40x | $T4_5$, $T6_2$ | | $T4_5$, $T6_2$ | |

### 2.2.2 Microdiversity of *Rhabdochlamydia porcellionis*

#### *2.2.2.1 Sampling and animals*

The *Porcellio scaber* individuals investigated in this study were sampled from 15 different locations in North Italy, Bavaria, Upper Austria, Lower Austria and Vienna (table 9). The woodlice were collected in summer/autumn 2015 and spring/summer 2016 and stored in plastic boxes with soil and litter at relatively high humidity at room temperature. For the study individuals of both sexes and different developmental stages were used. The Slovenian population was known to be infected and provided for the study by Rok Kostanjsek, University of Ljubljana. This population was used as a positive control.

#### *2.2.2.2 Preparation of the animals and hepatopancreases*

For the molecular approaches, the animals had to be surface sterilized. For that purpose, the animals were put in DanKlorix and sterile 1x PBS for 1 minute, respectively. Afterwards, the hepatopancreases were removed with fine-tipped forceps and homogenized in 1x PBS with Dstroy sticks.

#### *2.2.2.3 Screening of woodlice populations*

In order to quickly screen the woodlice populations for infection with *Rhabdochlamydia,* two different PCR approaches have been tested: standard PCR and semi-nested PCR of 16S rDNA. In total, two individuals per population were used. For both approaches the homogenized hepatopancreases were filtered through a 5 µm filter and washed with 1x PBS. Afterwards, the solution was exposed to three freeze-thaw cycles (-72 °C and 37 °C), filtered through a 1.2 µm filter and washed with 1x PBS. The received solution was directly used for the PCRs. The PCRs were carried out using the DreamTaq Green PCR Polymerase as specified. To minimize the influence of possible inhibitory substances BSA (0.2 µl per 50 µl) was added. For the amplification of the 16S rRNA gene using standard PCR, the PANF/R primer set was applied (denaturation at 94 °C for 7 minutes, 30 cycles of 30 s at 95 °C, 30 s at 65 °C and 90 s at 72 °C, followed by a terminal extension of 7 min at 72 °C).

For the semi-nested PCR, two separate PCR reactions were carried out. In the first reaction, the 16S1/PANR primers were used. The settings of the first PCR reaction were the same as with the standard PCR except for the annealing temperature, which was set to 56 °C, and the number of cycles, which was decreased to 25. In the second reaction the PANF/R primers were applied. The

PCR was carried out as described for the standard PCR except for the first long denaturation step, which was set to 3 minutes. As a template, a 1:50 dilution of the end product of the first reaction was used. Both the standard and the semi-nested PCR resulted in amplicons of approximately 1500 bp. The amplified DNA was checked by agarose gel electrophoresis.

In order to evaluate the sensitivity of the two different PCR approaches a dilution series of the infected hepatopancreases from the Slovenian population was carried out. For that purpose, the homogenized hepatopancreas sample was diluted 1:10. In total, seven dilution steps ($10^1$, $10^2$, $10^3$, $10^4$, $10^5$, $10^6$) were conducted. The dilutions were used as a template for the standard and semi-nested PCR. The settings were the same as outlined above, except for the primers of the standard PCR, which were replaced by the 16S1/16S2 primers. Two additional controls were used in order to exclude the influence of potential PCR inhibitors present in the hepatopancreas that would cause false negatives. These controls consisted of *Rhabdochlamydia*-DNA isolated from an insect cell culture with and without the supplement of a hepatopancreas extract of an uninfected *Porcellio scaber* population.

#### *2.2.2.4 Cultivation of Rhabdochlamydia porcellionis*

Sf9 insect cells were used for the cultivation of *Rhabdochlamydia porcellionis* from the Slovenian population. The insect cell line was maintained at 27 °C in Grace's insect medium supplemented with 10 % fetal bovine serum and 2 mM L-glutamine and sub-cultured once per week. Overall, four woodlice were used for cultivation. The homogenized hepatopancreases from surface-sterilized individuals were filtered through 1.2 µm filters. Afterwards, 50 µl of each solution were transferred to the first column of a 24-well plate containing Sf9 cells. Then, a dilution series was carried out in order to reduce the probability of contaminations by transferring 100 µl of the previous well into the next one (figure 8). In total, five dilution steps were conducted for each of the four hepatopancreases. Finally, the plate was centrifuged 15 minutes at 1,000 rpm to increase the probability of infection. After an incubation time of 7 days the content of not overgrown wells was transferred to a 12-well plate and after 14 days the content of the wells from the lowest dilution steps ($10^{-6}$, $10^{-8}$, $10^{-10}$) was transferred to culture flasks.

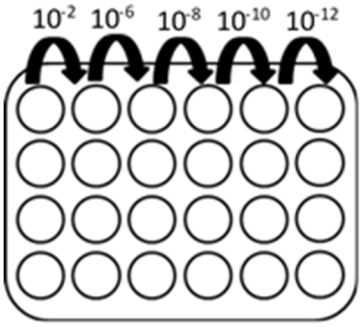



Figure 8: Scheme of the dilution series carried out on a 24-well plate containing Sf9 insect cells.

*2.2.2.5 Screening of Sf9 cultures*

Two of the received cell cultures were checked for infection with *Rhabdochlamydia* via PCR. For that purpose, DNA was isolated from the cell cultures using the DNeasy Blood and tissue kit according to the manufacturer's recommendations. The PCR of the 16S rDNA was carried out using the DreamTaq Green PCR Polymerase as specified and the 16S1/16S2 primers (denaturation at 95 °C for 3 minutes, 30 cycles of 30 s at 95 °C, 30 s at 56 °C and 90 s at 72 °C, followed by a terminal extension of 7 min at 72 °C). The PCR product was quality checked by gel electrophoresis and purified using the QIAquick PCR Purification Kit according to the manufacturer's recommendations. The purified PCR products were sent for Sanger sequencing to Microsynth GmbH (Baglach, Switzerland). The sequences were analyzed using BLASTN (Altschul et al., 1990).

As an alternative method for screening the cell cultures for *Rhabdochlamydia* Fluorescence In Situ Hybridization (FISH) was used. For that purpose, the cell cultures were harvested and washed with 1x PBS. The cells were then applied on a glass slide and incubated at room temperature for 30 minutes to allow the cells to attach. Afterwards, the cells were washed with 1x PBS and fixed with 4% paraformaldehyde at room temperature for 1 hour. Subsequently, the cells were dehydrated in a gradient ethanol series (50 %, 80 % and 96 %). The hybridization buffer (25 % FA) and 1 µl of each of the labeled probes (EUK516, CHLS-0523, EUBI/II/III) were added and the slide was incubated at 46 °C for 3 hours in a humid hybridization chamber. Then, the slide was transferred to a pre-warmed washing buffer and left there for 10 minutes at 48 °C. Finally, the slide was dipped into cold distilled water and quickly dried by an airstream. In addition to the fluorescence dyes, the cells were stained with DAPI using a 1:10,000 PBS dilution. After an incubation period of 2 minutes at room temperature, the slide was washed in 96% ethanol for 2 minutes. Prior to microscopy the slide was mounted with CitiFluor. The slide was analyzed under a CLSM microscope (Leica CTR 6500).

*2.2.2.6 High molecular weight DNA isolation for genome sequencing*

In total, 20 animals were used for HMW-DNA isolation (5 for each replicate). For that purpose, the animals were surface sterilized by putting them in DanKlorix and PBS for 2 minutes, respectively. Then, the hepatopancreases were removed with fine-tipped forceps and homogenized in buffer A with Dstroy sticks. The received solution was filtered through a 5 µm filter and centrifuged at 6,000 rpm for 10 minutes. Subsequently, the pellet was resuspended in 8 µl distilled water, 1 µl DNase I buffer and 1 µl DNase I and incubated on ice for 1 hour. The mixture was then incubated with 1 µl 0.5 M EDTA at 65 °C for 10 minutes and centrifuged at 6,000 rpm for 10 minutes. Afterwards, the pellet was washed with 20 µl buffer A + EDTA and 0.5 M EDTA and incubated with 200 µl lysis buffer for 30 minutes at 37 °C. The DNA isolation was then carried out using the DNeasy Blood and tissue kit as specified. In a last step, RNase A (1 µl per 100 µl) was added and the solution was incubated for 20 minutes at 37 °C. The quality of the DNA was checked with gel electrophoresis and the DNA concentrations were evaluated using the Quant-iT™ PicoGreen® dsDNA Assay Kit according to the manufacturer's recommendations. The measurement was carried out using the Infinite® F500 microplate reader.

# 3 RESULTS

## 3.1 Biogeography of *Chlamydiae*

### 3.1.1 DNA Isolation

The sediment samples from the River Po delta contained about 27 ng/µl and the water filters from Lago di Paola about 53 ng/µl DNA on average (table 4). The DNA extraction of the activated sludge samples resulted in a mean concentration of 30 ng/µl, where the VetMed sample was used in a 1:100 and the Ingolstadt sample in a 1:50 dilution of the original sample (table 4).

Table 4: DNA concentrations of the environmental samples in ng/µl measured with Nanodrop 1000.

| Identifier | DNA conc. [ng/µl] |
|---|---|
| $T4_1$ | 139.0 |
| $T4_2$ | 22.7 |
| $T4_3$ | 33.4 |
| $T4_4$ | 57.0 |
| $T4_5$ | 81.2 |
| $T4_6$ | 46.9 |
| $T4_7$ | 38.2 |
| $T4_8$ | 66.9 |
| $T4_9$ | 98.9 |
| $T5_1$ | 34.7 |
| $T5_2$ | 18.0 |
| $T5_3$ | 27.2 |
| $T5_4$ | 62.7 |
| $T5_5$ | 20.0 |
| $T5_6$ | 28.7 |
| $T5_7$ | 46.3 |
| $T5_9$ | 40.7 |
| $T6_1$ | 45.3 |
| $T6_2$ | 35.6 |
| $T6_3$ | 70.3 |

| Identifier | DNA conc. [ng/µl] |
|---|---|
| M | 27.1 |
| A | 5.0 |
| B | 26.6 |
| C | 8.7 |
| D | 10.9 |
| E | 0.0 |
| F | 37.1 |
| G | 36.3 |
| H | 43.0 |
| I | 25.4 |
| L | 81.9 |
| VetMed | 26.0 (original: 2,600) |
| Ingo | 27.5 (original: 1,375) |

### 3.1.2 PCR and specificity of the PCR primers

The PCR resulted in sufficient but relatively low product concentrations of 6.0 ng/µl on average for the environmental samples. Furthermore, using the described settings the primers were not specific for *Chlamydiae* but amplified also 16S rDNA of other members of the PVC superphylum. Most of the unspecific reads belonged to *Planctomycetes*, followed by *Verrucomicrobia*,

*Lentisphaerae* and candidate division OP3. The fraction of chlamydial reads was less than 50 % in most of the samples (figure 9). In total, the filter samples contained about 850 reads, the sediment samples about 780 reads and the sludge samples about 870 reads on average. Only the read counts of filter sample $T6_2$ and sediment sample E deviated significantly from the average amounting to 30 and 10 reads, respectively. For comparison, the average read count in the positive controls amounted to 2,500 reads.

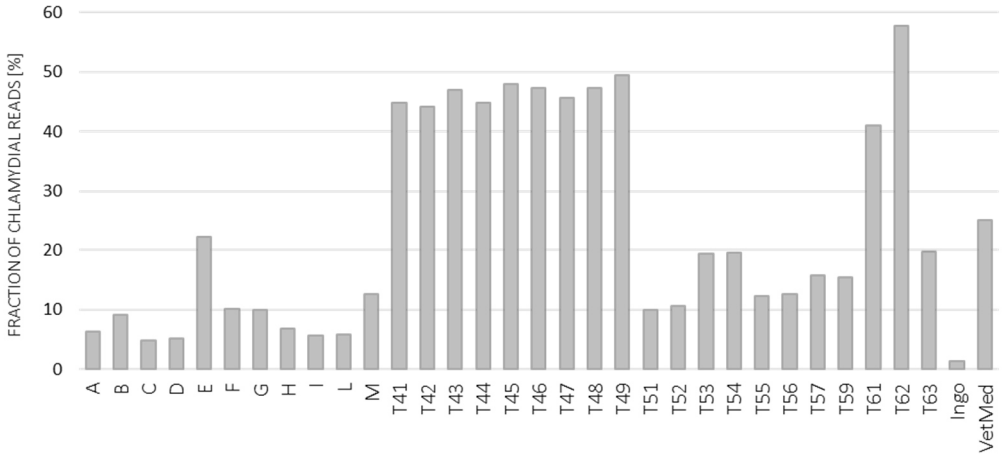



Figure 9: Fraction of chlamydial reads in percentage for environmental samples.

### 3.1.3 Evaluation of the mock communities

For the genomic DNA mock community 0.2 % and for the 16S rDNA mock community 0.3 % of all reads did not belong to any of the pooled organisms. In both cases, *Simkania negevensis* was underrepresented, whereas *Parachlamydia acanthamoeba UV7* was present in a higher abundance than expected (figure 10).

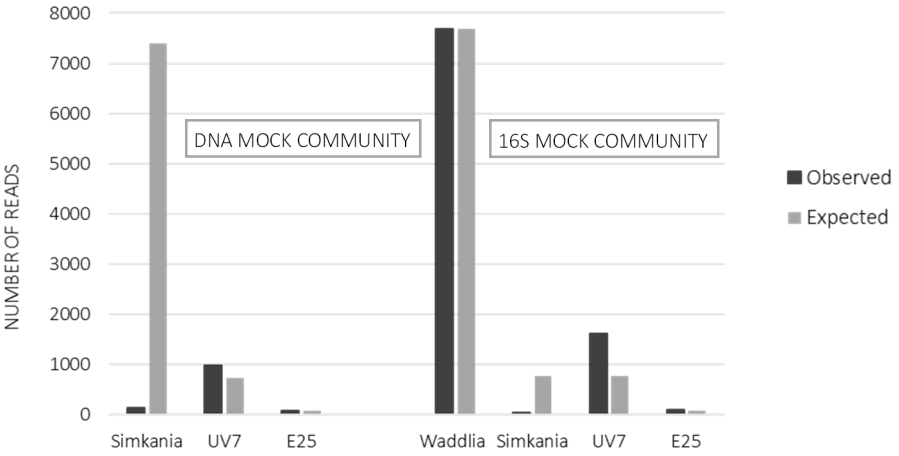



Figure 10: Read-based composition of the genomic DNA and 16S rDNA mock community.

A detailed analysis of the primer sequences revealed a T-C and A-C mismatch to *Simkania negevensis* in the reverse primer (figure 11).


Figure 11: Mismatch positions of *Simkania negevensis*-16S rDNA and the reverse primer.

### 3.1.4 Statistical analysis

All statistical analyses were carried out using a threshold read count of 50 reads, which means that all samples with a read count lower than 50 were excluded. The relatively low threshold was used in order to keep as many samples as possible. Further, due to unequal read counts the samples were rarefied to the lowest observed read number for the NMDS and SIMPER analysis. To retain as much information as possible, OTUs that occurred in the negative control were not strictly removed. Instead, an OTU was only removed from a sample if it had less than 10 reads assigned to it. An overview of all OTUs and assigned reads is given in supplement 3.

#### *3.1.4.1 Community composition*

The chlamydial community composition of the environmental samples on the genus level can be seen in figure 12. The communities associated with the Lago di Paola samples were dominated by the genus *Fritschea,* followed by unclassified *Chlamydiales* and *Parachlamydiaceae.* Those groups accounted for more than 80 % of the whole community in all water samples. In comparison, the communities of the sediment samples showed a more even distribution among the chlamydial groups. In sample B to D, unclassified *Parachlamydiaceae* were most abundant, whereas in sample A unclassified *Chlamydiales* were more abundant. In sample F to H, members of the genus *Fritschea* showed the highest abundance. Sample L was dominated by members of *Fritschea*, unclassified *Chlamydiales* and *Criblamydia*. The sediment sample from Lago di Paola (M) showed a high abundance of *Neochlamydia*. The activated sludge sample (VetMed) was dominated by unclassified *Chlamydiales* followed by *Protochlamydia* and *Simkania*. In total, those *Chlamydiae* accounted for about 98 % of the whole community.

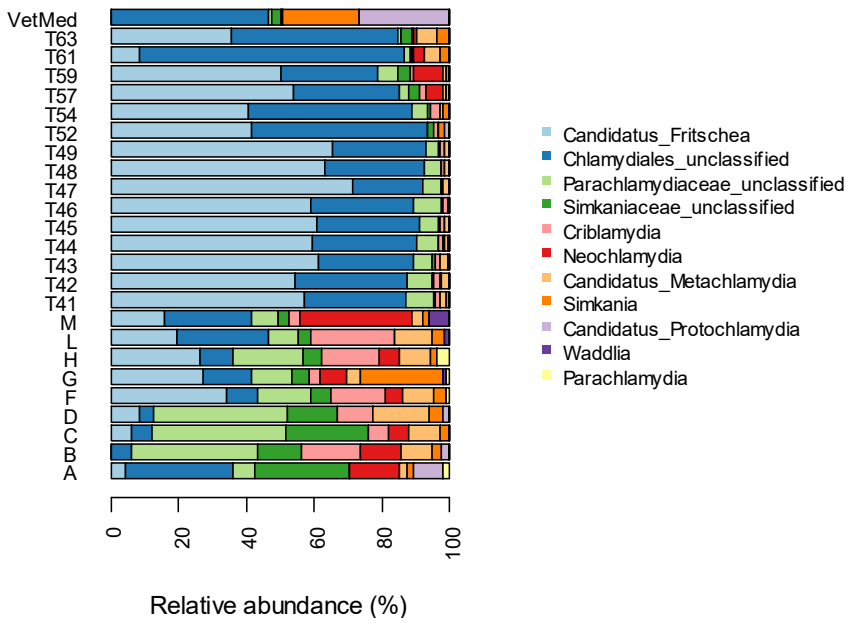



Figure 12: Chlamydial community composition on genus level. The classification is based on RDP.

The overall bacterial community amplified with general bacterial primers for the River Po Delta sample H, the Lago di Paola sample $T4_4$ and the activated sludge sample on phylum level are shown in figure 13. The sediment sample was dominated by *Proteobacteria*, which accounted for about 80 % of the total community. The water sample showed a more even distribution where *Proteobacteria*, *Bacteroidetes*, *Cyanobacteria* and *Actinobacteria* were the dominant phyla. The activated sludge sample was as the sediment sample dominated by *Proteobacteria*. Further, in the sediment sample H and the VetMed sample *Chlamydiae* accounted for about 0.9 % of the total community.

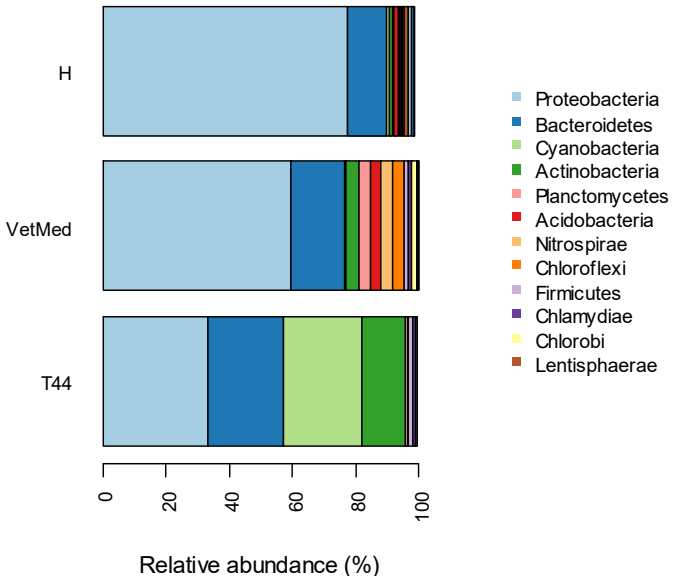



Figure 13: General bacterial community composition on phylum level. In the figure only phyla that account for at least 0.5 % of the community on average are shown. The classification is based on RDP.

### *3.1.4.2 Non-metric multidimensional scaling and perMANOVA*

As can be seen in figure 14, samples from the same time point were situated more closely to each other than to samples from the other time points. This indicates that samples of the same time point were more similar to each other with respect to OTU composition than to the other time points. The stress value accounted for 0.1, representing a fair fit of the ordination to the data. The results of the perMANOVA are given in table 5. As the variances were homogenous (p= 0.11), the usage of perMANOVA was valid. As already indicated by the NMDS also the results of the perMANOVA showed a highly significant influence of time, explaining about 80 % of the variance in the data (table 5).

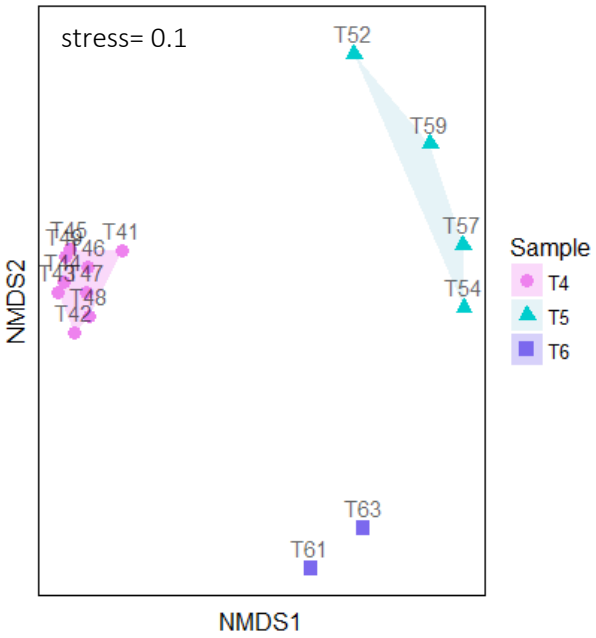



Figure 14: NMDS plot of the water samples from Lago di Paola. The stress value accounts for 0.1.

Table 5: Results of the perMANOVA using the factor time. The P-values are based on 999 permutations.

| | Df | SS | MS | F | $R^2$ | P |
|---|---|---|---|---|---|---|
| time | 2 | 1.5165 | 0.75823 | 27.25 | 0.81955 | 0.001 |
| Residuals | 12 | 0.3339 | 0.02782 | | 0.18045 | |
| Total | 14 | 1.8503 | | | 1.00000 | |

The results of the NMDS analysis of the River Po Delta sediment samples are shown in figure 15. As for the water samples, the sediment samples seem to cluster with time. The stress value accounted for 0.08, representing a fair fit of the ordination to the data. The perMANOVA results

can be seen in table 6. In addition, to time also the geographic location was tested as influencing factor. For both factors, the variances were homogenous (p= 0.58 (time), p=0.08 (location)). The perMANOVA showed a slight influence of the factor time despite the highly significant p-value as only 33% of the variance in the data could be explained. For the factor geographic location, both the p-value and the $R^2$ showed no significant influence.

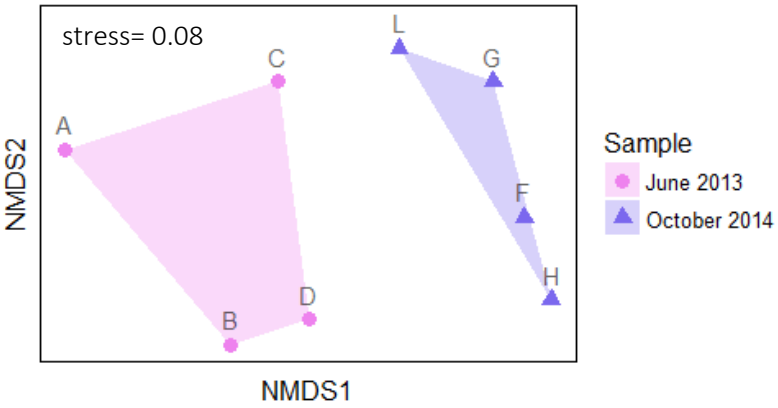



Figure 15: NMDS plot of the sediment samples from River Po Delta. The stress value accounts for 0.08.

Table 6: Results of the perMANOVA using the factor time. The P-values are based on 999 permutations.

| | Df | SS | MS | F | $R^2$ | P |
|---|---|---|---|---|---|---|
| time | 1 | 0.66357 | 0.66357 | 3.4234 | 0.32834 | 0.003 |
| geo | 2 | 0.58211 | 0.29105 | 1.5016 | 0.28803 | 0.100 |
| Residuals | 4 | 0.77533 | 0.19383 | | 0.38364 | |
| Total | 7 | 2.02101 | | | 1.00000 | |

### *3.1.4.3 Similarity percentages analysis*

Table 7 shows the results of the similarity percentages analysis for the Lago di Paola samples. OTU_3 and OTU_19 were significantly different abundant among time point T4 and T5. Both OTUs were assigned to unclassified *Chlamydiales* according to the RDP classifier (Wang et al., 2007). The same OTUs varied also significantly in abundance among time point T4 and T6. Whereby, in this case also OTU_5, which was assigned to *Candidatus Fritschea*, was significantly different. Time points T5 and T6 were only associated with a significant difference in abundance of OTU_97, which was also affiliated to *Candidatus Fritschea*. For the River Po Delta samples OTU_91 and OTU_207 were significantly different in abundance on a significance level of p= 0.05. Both OTUs were assigned to unclassified *Chlamydiales*.

Table 7: SIMPER results for the water samples of Lago di Paola.

| | OTU | P |
|---|---|---|
| T4 vs T5 | OTU_3 | 0.001 |
| | OTU_19 | 0.05 |
| T4 vs T6 | OTU_5 | 0.001 |
| | OTU_3 | 0.05 |
| | OTU_19 | 0.05 |
| T5 vs T6 | OTU_97 | 0.05 |

### 3.1.5 Phylogenetic analysis

OTU_1 and OTU_286 were excluded from the multiple sequence alignment, as they did not fit well in the overall alignment. The Bayesian inference tree including the MiSeq sequences is shown in figure 16. The maximum likelihood tree and the Bayesian inference tree of the trimmed sequences are given in the supplement. For both Bayesian inference and maximum likelihood, it was not possible to resolve the relationships between the genera leading to a highly multifurcating Bayesian inference tree and low UF-bootstrap support values for the maximum likelihood tree. However, within clades the relationships could be resolved leading to a quite similar pattern for all trees. In general, most of the OTUs were either assigned to uncultured or unknown members of the *Chlamydiae*. Further, some OTUs clustered and formed distinct clades in the tree. Overall, the OTUs were spread throughout the tree. Most OTUs were found in only one type of sample, only few were found in all of them.

### 3.1.6 Re-evaluation of the PCR primers

#### *3.1.6.1 Specificity of the primers*

In all tested settings, the specificity of the primers could be increased (figure 17). For sample $T6_3$ the fraction of chlamydial reads could be increased from 20 % up to 77 % and for sample A, from 7 % up to 70 % using an annealing temperature of 61 °C, respectively. In sample A, the specificity increased continuously with higher annealing temperatures, whereas in sample $T6_3$ an annealing temperature of 58 °C led to a slight decrease in specificity. Despite the higher specificity, the increased annealing temperatures also affected the PCR product concentration, as it was relatively low for all samples accounting for about 2 ng/µl on average, especially for 61 °C.

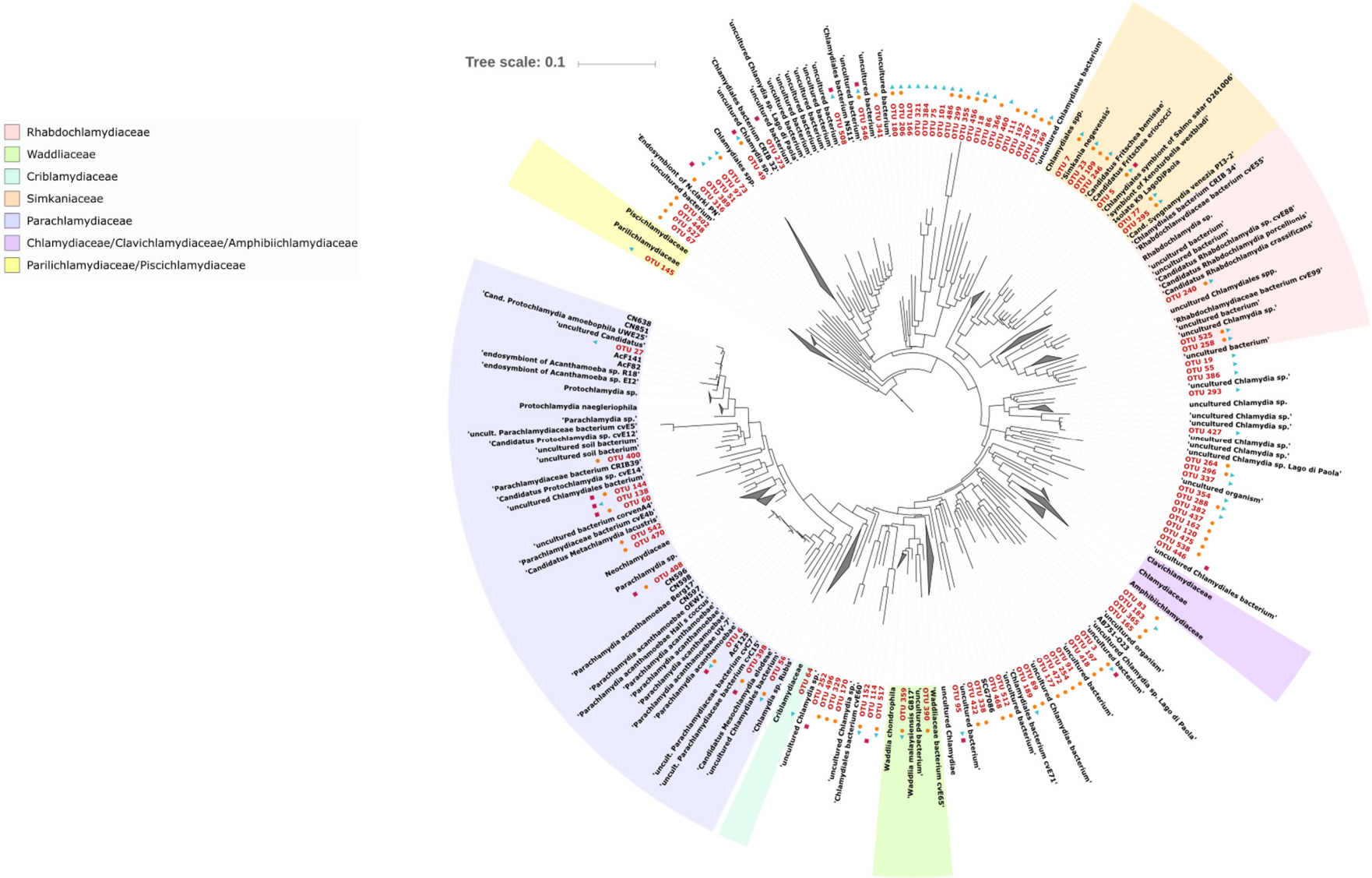



Figure 16: Bayesian inference tree of the phylum *Chlamydiae* including the OTUs obtained in this study (shown in red). The symbols attached to the OTUs indicate the type of samples they are associated with (circle – sediment, triangle – water, square – sludge).

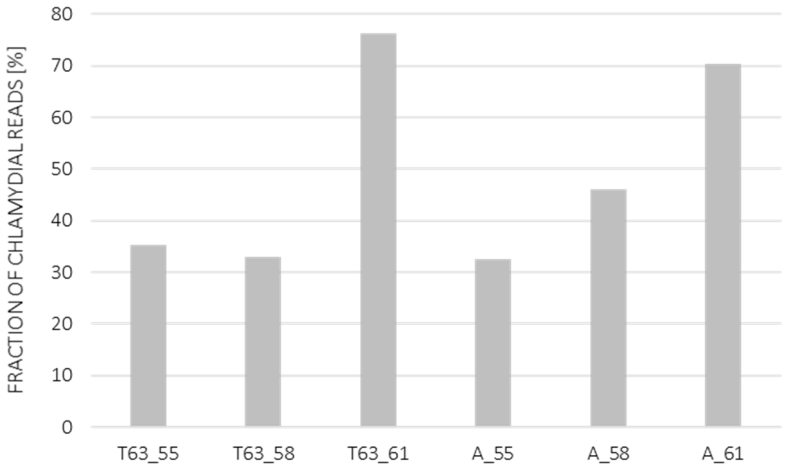



Figure 17: Fraction of chlamydial reads in percentage for different annealing temperatures.

### *3.1.6.2 Influence of increased cycle number*

In all samples, a higher cycle number led to higher PCR product concentrations of about 65 ng/µl on average. For sample $T6_2$ the highest concentrations were observed for an annealing temperature of 50 °C and 40 cycles (figure 18). The lowest concentrations were observed using an annealing temperature of 58 °C and 35 cycles. A similar pattern was observed for sample $T4_3$ (figure 18). However, here the concentrations were highest for an annealing temperature of 50 °C and 35 cycles and the difference between the two annealing temperatures was much less (figure 18). As shown for sample $T6_3$ and A, the specificity was increased from 50 °C to 58 °C (figure 19). The influence of the increased cycle number on the community composition of sample $T4_5$ is shown in table 8. As can be seen the overall pattern resembled those of the original MiSeq run (table 8).

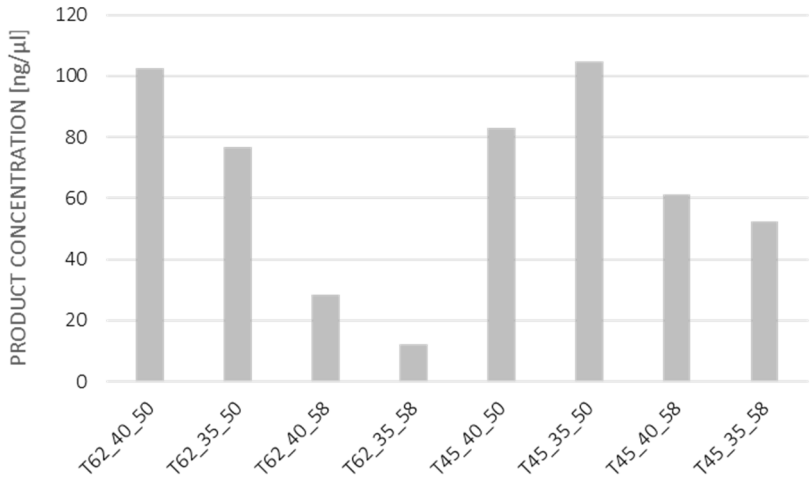



Figure 18: PCR product concentrations in ng/µl of sample $T6_2$ and $T4_5$.

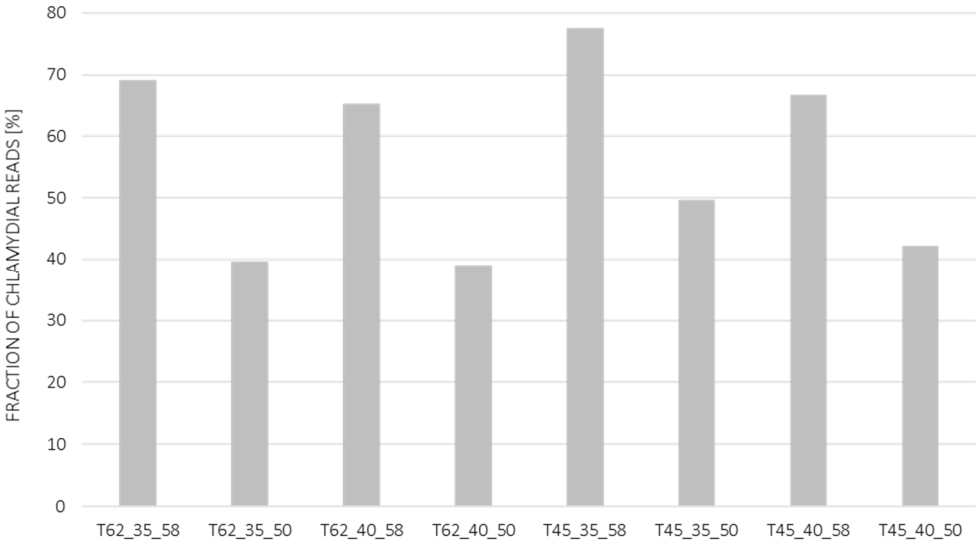



Figure 19: Fraction of chlamydial reads in % for sample $T6_2$ and $T4_5$.

Table 8: Community composition of sample $T4_5$ for an annealing temperature of 55 °C and 58 °C. The values represent the relative abundances of the genera in percentage. The original sample is shown in bold.

| | *Cand. Fritschea* | *Chlamydiales unc.* | *Parachlamydiaceae unc.* | *Criblamydia* | *Simkaniaceae unc.* | *Simkania* | *Cand. Metachlamydia* |
|---|---|---|---|---|---|---|---|
| T45_35_50 | 60.12 | 34.16 | 1.39 | 2.16 | 0.93 | 0.93 | 0.30 |
| T45_35_58 | 55.15 | 40.18 | 1.45 | 1.29 | 0.70 | 0.59 | 0.64 |
| T45_40_50 | 59.27 | 36.14 | 1.23 | 2.14 | 0.61 | 0.15 | 0.46 |
| T45_40_58 | 60.56 | 35.21 | 0.78 | 2.19 | 0.94 | 0.16 | 0.16 |
| **T45** | **60.61** | **30.42** | **5.45** | **1.33** | **0.60** | **0.12** | **1.46** |

## 3.2 Microdiversity of *Rhabdochlamydia porcellionis*

### 3.2.1 Screening of woodlice populations

Both the standard as well as the semi-nested PCR resulted in bands for the Slovenian population and one of the Bavarian populations (table 9). The dilution series revealed a detection limit of 1:100 for the standard PCR, i.e., populations with an infection status 100 times lower than the Slovenian population can be detected by the standard PCR. For the semi-nested PCR, the detection limit could be increased up to 1:10,000. The evaluation of potential PCR inhibitors in the hepatopancreas did not yield positive results. Thus, it can be excluded that the PCR was inhibited from host-derived substances. Some of the investigated animals are shown in figure 20.

Table 9: Sampling points and infection status of the investigated populations.

| Identifier | Sampling point | Infection status |
|---|---|---|
| LUB | Ljubljana (Slovenia) | Highly infected |
| OOE | Linz, Upper Austria (Austria) | Not infected |
| BA | Baden, Lower Austria (Austria) | Not infected |
| VI | Ottakring, Vienna (Austria) | Not infected |
| IT | South Tyrol (Italy) | Not infected |
| NW | Neuwaldegg, Vienna (Austria) | Not infected |
| OD | Kaisermuehlen - Old Danube, Vienna (Austria) | Not infected |
| LOB | Lobau, Lower Austria (Austria) | Not infected |
| DP | Kaisermuehlen - Donau Park, Vienna (Austria) | Not infected |
| PID | Piding, Bavaria (Germany) | Slightly infected |
| KB | Ottakring - Kongressbad, Vienna (Austria) | Not infected |
| AG | Ottakring - Arnethgasse, Vienna (Austria) | Not infected |
| DI | Brigittenau - Donauinsel, Vienna (Austria) | Not infected |
| WP | Floridsdorf - Wasserpark, Vienna (Austria) | Not infected |
| HS | Doebling - Hoehenstraße, Vienna (Austria) | Not infected |

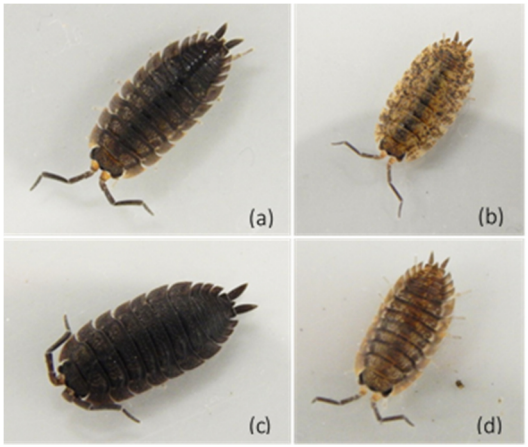



Figure 20: Individuals of investigated woodlice populations. *Porcellio scaber* from Slovenia (a), Lower Austria (b), Bavaria (Piding) (c) and Italy (d).

### 3.2.2 Cultivation of *Rhabdochlamydia porcellionis* from Slovenian population

The PCR and sequencing of the infected cell cultures revealed that the culture from the lowest dilution step ($10^{-6}$) was infected with *Rhabdochlamydia porcellionis*. In general, half of the original wells were contaminated. Just like PCR, FISH also turned out to be suitable as a screening method for *Rhabdochlamydia*. The respective FISH image can be seen in figure 21.

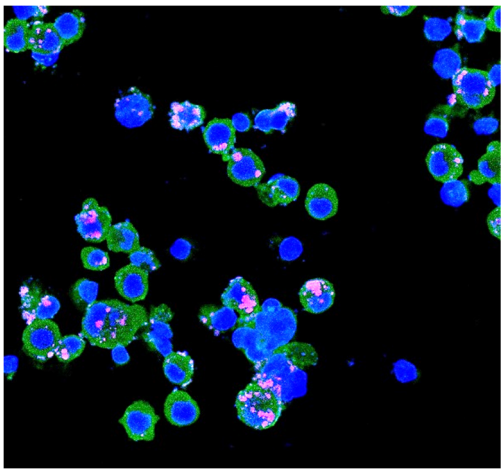


Figure 21: FISH image of the Sf9 cell cultures infected with *Rhabdochlamydia porcellionis*. DAPI signals are shown in blue, the eukaryotic probe EUK516 labelled with Cy5 is shown in green and the chlamydiae specific probe CHLS-0523 which is labelled with Cy3 is shown in pink.

### 3.2.3 High molecular weight DNA Isolation

The gel electrophoresis revealed a high quality of the isolated HMW-DNA with a size larger than 1,000 kB. The PicoGreen measurement showed a mean DNA concentration of 24 ng/µl for the Slovenian population and 10 ng/µl for the Bavarian population. In total, 19.5 µg DNA were isolated from the Slovenian population and 3.7 µg from the Bavarian population.

# 4 DISCUSSION

## 4.1 Biogeography of *Chlamydiae*

### 4.1.1 DNA Isolation and PCR

The choice of an adequate DNA extraction method is critical for a community analysis as it is the basis of all further steps and can introduce biases (Feinstein et al., 2009; Guo and Zhang, 2013; Martin-Laurent et al., 2001). Because *Chlamydiae* are rare in most environments (Pizzetti et al., 2015) it is of particular importance to use a very sensitive extraction method. As the water filter samples and the activated sludge samples resulted in a higher mean DNA concentration, P:C:I protocols seem to be more suitable than the DNA extraction kit used for the sediment samples. However, as the mean DNA concentration was still quite low for the water filter samples, it will be necessary to adapt the used DNA extraction protocol, e.g., pool more replicates per sample or filter a larger volume of water. The lower DNA concentration of the sediment samples can also be due to the nature of the samples as sediments have a complex structure (Roose-Amsaleg et al., 2001). For future studies, it would be favorable to evaluate different DNA extraction methods beforehand in order to find the best fitting method as the success largely depends on the type of samples. Furthermore, it is recommended to use the same extraction method for all samples to be able to reliably compare them. The low PCR product concentrations are probably due to the low abundance of *Chlamydiae* (Pizzetti et al., 2015) and the relatively low DNA concentrations. The low number of reads are mainly due to the unspecific binding of the primers but maybe also due to the low concentrations of PCR products obtained. As the number of reads obtained in the positive controls and the mock communities were in a normal range a general bias of the workflow or the sequencing can be excluded.

### 4.1.2 Evaluation of the mock communities

PCR as well as next generation sequencing techniques are associated with biases (Kunin et al., 2010; Lee et al., 2012) leading to shifts and biases in community composition analyses. Although some of these biases can be reduced, e.g., by pooling of replicates or strict computational processing most of them cannot be avoided (Lee et al., 2012). Thus, it is important to assess their influence in order to be able to draw correct conclusions. Mock communities are a favorable tool in this respect. In the current study, two different mock communities were tested in order to find the most sensitive one. Although both of them showed the same pattern regarding *Simkania negevensis,* the influence of the mismatches was more pronounced in the DNA mock community.

This could be due to the higher complexity of the DNA mock community with respect to accessibility of the 16S rRNA gene target. The higher abundance of UV7 can be explained by a binding of more primer version to this target due to the low annealing temperature. This scenario is also supported by the fact that the effect was much lower in the DNA mock community. Apart from the mentioned biases, the composition of the mock communities meets the expectations and there is no indication for a strong PCR or sequencing bias. Furthermore, the results indicate that the primers are highly sensitive for *Chlamydiae*. For future studies, it is recommended to include mock communities in each sequencing run to be able to detect possible PCR and sequencing biases. However, as the genomic DNA mock community represents a system more similar to an environmental sample and seems to be more sensitive it should be used preferentially.

### 4.1.3 Composition and dynamics of chlamydial communities

In general, due to the low read numbers, the bias against *Simkania negevensis* and the different DNA extraction methods, the statistical validity of the results is limited and the obtained results should be interpreted with caution.

The observations that members of the genus *Fritschea* were highly abundant in the Lago di Paola samples is in line with the expectations as close relatives were found to infect different fish species and marine worms (Nylund et al., 2015; Fehr, 2013; Israelsson, 2007). Furthermore, the high abundance of unclassified members was also expected and verified the assumption that there is a high, undiscovered diversity within the *Chlamydiae* (Lagkouvardos et al., 2014). As almost nothing is known about the diversity of *Chlamydiae* in marine sediments it is interesting that those samples showed an even more diverse and different pattern than the water samples. In general, the observed diversity supports previous studies, since there were already indications that *Chlamydiae* are highly diverse especially in marine environments (Lagkouvardos et al., 2014). The composition of the overall bacterial community resembles previous findings stating that *Proteobacteria* are highly abundant in marine sediments and seawater (Wang et al., 2012; Gonzalez and Moran, 1997). However, it is interesting that the distribution pattern was different in comparison to the *Chlamydiae* as the sediment sample showed a more equal distribution than the Lago di Paola sample. Further, it was surprising that *Chlamydiae* accounted for about 1 % of the total community in the River Po Delta sample H and the VetMed sample, as general bacterial primers are known to have mismatches for *Chlamydiae*. However, only three OTUs were amplified

with the general primers which is extremely low in comparison to the number of OTUs observed with the chlamydiae specific primers which amounts to 23 OTUs in the River Po Delta sample H and 13 OTUs in the VetMed sample. This difference in found OTUs demonstrates the power of the chlamydiae-specific approach used in the present study to investigate chlamydial diversity and distribution in the environment.

The results of the NMDS are in line with previous studies of the Lago di Paola where shifts in the chlamydial community were found (Pizzetti et al., 2012). However, the statistical analysis is probably biased due to the unequal sample sizes but as the results were highly significant the results are still regarded as valid. The observation that only few OTUs were significantly different between the time points is mainly due to the low read counts. The minor influence of time in the Po Delta samples is also in line with the expectations as deltas are highly dynamic ecosystems and the samples were taken from different sampling points. However, as shown by the SIMPER results some OTUs were still significantly different between the investigated time points supporting an influence of time on the community composition. In spite of that, it is recommended to use samples from the same sampling point for future time series in order to exclude influences of location specific parameters.

### 4.1.4 Phylogenetic analysis

A detailed phylogenetic analysis is advantageous in addition to the classification carried out by the RDP Classifier (Wang et al., 2007), which only provides reliable information about the affiliation to higher taxonomic levels such as genera or families. In comparison, phylogenetic trees comprise not only information about the closest relative on the species level but also about evolutionary distances. In general, the results of the phylogenetic analysis largely confirm the results of the RDP classification (Wang et al., 2007) and showed a high diversity of *Chlamydiae*. Furthermore, the observed distinct groups of OTUs were also previously reported for Lago di Paola (Lagkouvardos et al., 2014; Pizzetti et al., 2012). However, it is interesting that some of the OTUs (e.g. OTU_77, OTU_295 and OTU_296) of these clusters were also found in sediment samples of the River Po Delta. In combination with the observation that OTU_6, OTU_3, OTU_5, OTU_546 and OTU_49 were found in all three sample types, this indicates that there are some ubiquitous members with a potentially large host spectrum. The observation that it was not possible to resolve the phylogeny among families with Bayesian inference and maximum likelihood is likely due to the short sequence length of the MiSeq sequences. Furthermore, as the Bayesian inference tree partly

showed different patterns even within the families the amplified region is probably due to the high variability not ideal to resolve phylogeny of the *Chlamydiae*. Thus, it is recommended to calculate a reference tree with full-length sequences and to add the shorter MiSeq sequences afterwards for future studies.

### 4.1.5 Re-evaluation of the PCR primers

The results of the re-evaluation largely meet the expectations. The observation that an annealing temperature of 58 °C led to a decrease of specificity in sample $T6_3$ is likely due to a random PCR bias as for all other samples the specificity was increased at 58 °C. Also the influence of the increased cycle number was as expected. However, it was interesting that for sample $T4_5$ the concentrations were highest for an annealing temperature of 50 °C and 35 cycles, which could indicate a product inhibition at 50 °C and 40 cycles. The observation that the community composition did not change significantly although the number of cycles was increased confirms the observation of Sipos et al. (2007), who state that the influence of PCR cycles is minor in environmental samples. Based on the re-evaluation, it is recommended to use an annealing temperature of at least 58 °C but only in combination with an increased cycle number of at least 35 for future studies. Although the results of the re-evaluation allow an estimation about the optimal settings, further evaluations are necessary to find the optimum for all samples. Further, it will be necessary to add an additional primer version for *Simkania negevensis* to the reverse primer mix and to evaluate if the mismatch was causal for the shift in the mock communities and to what extent it affected the results of the environmental samples.

### 4.1.6 Concluding remarks and outlook

The workflow presented in this study can be used to screen a large number of samples efficiently for *Chlamydiae* and allow the detection of a broad spectrum of diversity. However, the settings have to be adapted a bit further to receive enough reads to allow strong statistical statements. In comparison, other studies carrying out similar statistical analyses used at least about 2,000 reads (Mosher et al., 2012). For future studies, it would be interesting to investigate more extreme environments, e.g., samples from ice lake or environments where nothing is known about the occurrence of *Chlamydiae,* e.g., arctic soils. But also a deeper investigation of the Lago di Paola and River Po Delta samples would be of interest especially with respect to co-occurring eukaryotes, as some of the observed OTUs were closely related to members infecting fish and invertebrates. In addition, it would be interesting to compare samples of the same type that differ

in location. In order to investigate if observed differences are due to a difference in host organisms or an adaption of the *Chlamydiae* to the environment. Moreover, to be able to make more reliable statements about the abundance of *Chlamydiae* it would be necessary to establish a qPCR protocol. Furthermore, it would be favorable to include a chlamydiae specific gene in order to get an even higher resolution of the phylogeny. In the course of this study, primers for the chlamydiae specific gene CT016, which encodes a hypothetical protein, were tested but the primers were too degenerated for PCR to work properly. Thus, it would be necessary to design less degenerated primers or to choose another gene. In addition, it would be interesting to develop a single cell genomics approach (Gawad et al., 2016) for *Chlamydiae* in environmental samples to get a deeper insight into the genomes of uncultivable members. To be able to verify that the *Chlamydiae* are really active in the samples, it would be further interesting to use the assay established here but to target RNA instead of DNA.

In conclusion, in the course of this study a chlamydial community composition was analyzed in environmental samples in detail for the first time. In conventional studies addressing the composition of microbial communities, *Chlamydiae* are underrepresented and underestimated as general bacterial primers have mismatches to *Chlamydiae*. Thus, the established workflow will in the future fundamentally contribute to complete microbial community analyses and to gain insights into the abundance, distribution, diversity, population structure and dynamics of these bacteria.

## 4.2 Microdiversity of *Rhabdochlamydia porcellionis*

### 4.2.1 Screening of woodlouse populations

As indicated by the results the presented workflow is suitable for a quick and sensitive screening of woodlouse populations. However, as the Slovenian population was known to be highly infected and the approach is only semi-quantitative, the detection limit can still be too low to detect very slightly infected animals. Thus, negative results should be treated with caution. Furthermore, to prevent overlooking infected populations, the screening must be repeated periodically. A quick and reliable screening is especially important as only two out of 15 populations were infected and sampling of the populations is time consuming. Moreover, a high number of infected population is needed for a microdiversity study. As the sensitivity could be increased substantially by the nested PCR it is recommended to use this approach for future studies. Although, the addition of

BSA did not remarkably influence the PCR results, its use is still recommended as the influence was only tested for the highly infected population.

### 4.2.2 Cultivation of *Rhabdochlamydia porcellionis*

Since it was already shown previously that *Rhabdochlamydia porcellionis* can be cultured in Sf9 insect cells (Sixt et al., 201), the purpose of this study was to optimize the established protocol and to assess its functionality for slightly infected populations. The high number of overgrown wells was expected, as the Grace's insect medium is very rich in nutrients and many bacteria are present in the hepatopancreas tissue. However, this could be a potential drawback when working with slightly infected populations. Thus, the initial surface sterilization of the animals should be increased up to 7 minutes. Further, as only the culture from the second dilution step ($10^{-6}$) was infected a relatively high amount of *Rhabdochlamydia* is needed to get them in culture. For slightly infected populations, it will be necessary to increase the number of used hepatopancreases and to reduce the number of passages. Although the screening of the cultures worked with both approaches, PCR and FISH, PCR should be used preferentially as the detection limit of FISH is relatively low ($10^3$ - $10^4$ cells per ml) and PCR turned out to be the more stable and reliable. However, FISH is an advantageous tool as it allows an estimation about the infection status and progress of the cells and thus can be used in addition.

### 4.2.3 High molecular weight DNA isolation

The analysis of the gel and the amount of isolated genomic DNA indicate that the workflow is suitable for isolation of high molecular weight DNA of *Rhabdochlamydia porcellionis*. The size of the genomic DNA resembles the expectations although the genomes of many environmental chlamydia have sizes of about 2 Mb (Collingro et al., 2011). However, as *Rhabdochlamydia porcellionis* infects multicellular organisms its genome size is expected to be more similar to that of *Chlamydiaceae* (~ 1 Mb) (Collingro et al., 2011). As the populations were highly infected and five hepatopancreases were used for each reaction the amount of yielded DNA is relatively low. Thus, for slightly infected populations it is recommendable either to get *Rhabdochlamydia* in culture and to isolate HMW-DNA from the cell cultures or to pool a higher number of hepatopancreases.

### 4.2.4 Concluding remarks and outlook

The workflow presented in this study provides a fast and relatively simple tool for screening large numbers of woodlice populations and isolating HMW-DNA for genome sequencing. Further, the

cultivation of *Rhabdochlamydia* allow the work with slightly infected populations and answering specific questions that may arise during genome analysis. Moreover, the workflow can not only be used for the model system used in this study but also be adapted for other insect host systems. However, as the success largely depends on the abundance of the bacteria the protocols may not be suitable for very slightly infected populations. The next step will be the sequencing of the isolated genomic DNA in order to evaluate the quality of the described workflow. For that purpose, the genomes will be sent to the Sanger Institute (Hinxton, UK) for PacBio sequencing. Afterwards, the genomes will be analyzed with respect to heterogeneity and strain-level diversity. The information from the genome analysis can then be used to perform infection assays in *Porcellio scaber* and insect cell cultures in order to gain insights into the biological importance of differently evolving strains e.g. with respect to infectivity.

In conclusion, in the course of this study a workflow was established that provides the possibility to gain deeper insights into the evolution and adaption mechanism of *Chlamydiae,* thereby contributing to a gain in knowledge about mechanisms that drive genome evolution in these rare and poorly characterized organisms.

# 5 LIST OF ABBERIVATIONS

| | |
|---|---|
| 16S rDNA | small subunit ribosomal RNA-encoding gene of prokaryotes |
| °C | degree Celsius |
| µ | micro ($10^{-6}$) |
| µm | micrometer |
| perMANOVA | analysis of variance |
| BSA | bovine serum albumin |
| bp | base pairs |
| Cy3 | indocarbocyanine |
| Cy5 | indodicarbocyanine |
| DAPI | 4',6-diamidino-2-phenylindole |
| dd$H_2O$ | double distilled and filtered water |
| DNA | deoxyribonucleic acid |
| EB | elementary body |
| EDTA | ethylenediaminetetraacetic acid |
| FISH | fluorescence in situ hybridization |
| FLUOS | fluorescein N-hydroxysuccinimidester |
| g | gram |
| HMW | high molecular weight |
| kb | kilobases |
| l | liter |
| m | milli ($10^{-3}$) |
| M | molar |
| Mb | megabases |
| MiSeq | illumina MiSeq high throughput sequencing |
| ng | nanogram ($10^{-9}$ g) |
| NMDS | non-metric multidimensional scaling |
| OTU | operational taxonomic unit |
| p | P value |
| PBS | phosphate buffered saline |
| P:C:I | phenol:chloroform:isoamyl alcohol |

| | |
|---|---|
| PCR | polymerase chain reaction |
| perMANOVA | permutational multivariate analysis of variance |
| PFA | paraformaldehyde |
| PVC | *Planctomycetes*, *Verrucomicrobia* and *Chlamydiae* |
| RB | reticulate body |
| rpm | rounds per minute |
| s | seconds |
| spp. | species |
| TBE | tris-borate-EDTA buffer |
| TE | tris-EDTA buffer |
| w/v | weight to volume |
| x | times |

# 6 REFERENCES

AbdelRahman, YM, Belland, RJ. The chlamydial developmental cycle. *FEMS Microbiol. Rev.* (2005).

Altschul, SF, Gish, W, Miller, W, Myers, EW, Lipman, DJ. "Basic local alignment search tool." *J. Mol. Biol.* (1990).

Amann, RI, et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* (1990).

Benson, DA, et al. GenBank. *Nucleic Acids Res.* (2013).

Cole, JR, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis *Nucl. Acids Res.* (2014).

Collingro, A, et al. Unity in Variety – The Pan-Genome of the *Chlamydiae*. *Mol. Biol. Evol.* (2011).

Corsaro, D, Venditti, D, Valassina, M. New parachlamydial 16S rDNA phylotypes detected in human clinical samples. *Research in Microbiology* (2002).

Corsaro, D, Venditti, D. Emerging chlamydial infections. *Crit. Rev. Microbiol.* (2004).

Daims, H, et al. Complete nitrification by *Nitrospira* bacteria. *Nature* (2015).

Fehr, A. Candidatus *Syngnamydia Venezia*, a Novel Member of the Phylum *Chlamydiae* from the Broad Nosed Pipefish, *Syngnathus typhle*. *PLoS One* (2013).

Feinstein, LM, Sul, WJ, Blackwood, CB. Assessment of Bias Associated with Incomplete Extraction of Microbial DNA from Soil. *Applied and environmental Microbiology* (2009).

Gawad, C, Koh, W, Quake, SR. Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics* (2016).

Gonzalez, JM, Moran, MA. Numerical Dominance of a Group of Marine Bacteria in the a-Subclass of the Class *Proteobacteria* in Coastal Seawater. *Applied and environmental Microbiology* (1997).

Gouy, M, Guindon, S, Gascuel, O. SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* (2010).

Gunnarsson, T. Selective feeding on a maple leaf by *Oniscus asellus* (Isopoda). *Pedobiologia* (1987).

Guo, F, Zhang, T. Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl. Microbiol. Biotechnol*. (2013).

Hassall, M, Jennings, JB. Adaptive features of gut structure and digestive physiology in terrestrial isopod *Philoscia muscorum* (Scopoli) 1763. *Biol. Bull.* (1975).

Herbold, C, et al. A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. *Front. Microbiol.* (2015).

Holger, D, Brühl, A, Amann, R, Schleifer, KH, Wagner, M. The Domain-specific Probe EUB338 is Insufficient for the Detection of all *Bacteria*: Development and Evaluation of a more Comprehensive Probe Set. *Systematic and Applied Microbiology* (1999).

Horn, M. *Chlamydiae* as Symbionts in *Eukaryotes*. *Annu. Rev. Microbiol.* (2008).

Israelsson, O. Chlamydial symbionts in the enigmatic *Xenoturbella* (Deuterostomia). *J. Invertebr. Pathol.* (2007).

Juck D, Charles T, Whyte LG, Greer CW. Polyphasic microbial community analysis of petroleum hydrocarbon-contaminated soils from two northern Canadian communities. *FEMS Microbiol. Ecol.* (2000).

Kim, YM, Ahn, CK, Woo, SH, Jung, GY, Park JM. Synergic degradation of phenanthrene by consortia of newly isolated bacterial strains. *J. Biotechnol.* (2009).

Kostanjsek, R, Marolt, TP. Pathogenesis, tissue distribution and host response to *Rhabdochlamydia porcellionis* infection in rough woodlouse *Porcellio scaber*. *Journal of Invertebrate Pathology* (2015).

Kostanjsek, R, Strus, J, Drobne, D, Avgustin, G. '*Candidatus Rhabdochlamydia porcellionis'*, an intracellular bacterium from the hepatopancreas of the terrestrial isopod *Porcellio scaber* (Crustacea: Isopoda). *International Journal of Systematic and Evolutionary Microbiology* (2004).

Kuo, C, Stephens, RS, Bavoil, PM, Kaltenboeck, B (2011). *Chlamydiaceae*. In Bergey 's Manual Systematic Bacteriology, 2nd edn, vol. 4. Springer Verlag, New York. 845-865.

Lagkouvardos, I, et al. Integrating metagenomic and amplicon databases to resolve the phylogenetic and ecological diversity of the *Chlamydiae*. *ISME* (2014).

Lau, R. www.chlamydiae.com. *Sexually Transmitted Infections* (2002).

Lee, CK, et al. Groundtruthing Next-Gen Sequencing for Microbial Ecology–Biases and Errors in Community Structure Estimates from PCR Amplicon Pyrosequencing. *PLoS ONE* (2012).

Ludwig, W, et al. ARB: a software environment for sequence data. *Nucleic Acids Research* (2004).

Lynch, MDJ, Neufeld, JD. Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology* (2015).

Martin-Laurent, F, et al. DNA Extraction from Soils: Old Bias for New Microbial Diversity Analysis Methods. *Appl. Environ. Microbiol.* (2001)

McAllister et al. Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing *Zetaproteobacteria*. *Appl. Environ. Microbiol.* (2011).

Methé, BA, et al. A framework for human microbiome research. *Nature* (2012).

Miller, MA, Pfeiffer, W, Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop* (2010).

Mosher, JJ, et al. Microbial Community Succession during Lactate Amendment and Electron Acceptor Limitation Reveals a Predominance of Metal-Reducing *Pelosinus spp*. *Appl. Environ. Microbiol.* (2012).

Nylund, S, et al. Characterization of '*Candidatus Syngnamydia salmonis*' (*Chlamydiales, Simkaniaceae*), a bacterium associated with epitheliocystis in Atlantic salmon (*Salmo salar L.*). *Arch. Microbiol.* (2015).

Oksanen, J, et al. vegan: Community Ecology Package. R package version 2.4-0. Available through: http://CRAN.R-project.org/package=vegan [26.10.2016].

Omsland, A, Sixt, BS, Horn, M, Hackstadt, T. Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities. *FEMS Microbiol. Rev.* (2014).

Pedrós-Alió, C. The rare bacterial biosphere. *Annu. Rev. Mar. Sci.* (2012).

Pizzetti, I, et al. Chlamydial seasonal dynamics and isolation of '*Candidatus Neptunochlamydia vexilliferae*' from a Tyrrhenian coastal lake. *Environ. Microbiol.* (2015).

Pizzetti, I, Fazi, S, Fuchs, BM, Amann, R. High abundance of novel environmental chlamydiae in a Tyrrhenian coastal lake (Lago di Paola, Italy). *Environmental Microbiology Reports* (2012).

Poppert, S, Essig, A, Marre, R, Wagner, M, Horn, M. Detection and Differentiation of *Chlamydiae* by Fluorescence In Situ Hybridization. *Appl. Environ. Microbiol.* (2002).

Quast, C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res*. (2013).

Roose-Amsaleg, CL, Garnier-Sillam, E, Harry, M. Extraction and purification of microbial DNA from soil and sediment samples. *Applied Soil Ecology* (2001).

Sipos, R, Szekely, AJ, Palatinszky, M, Revesz, S, Marialigeti, K, Nikolausz, M. Effect of primer mismatch, annealing temperature and PCR cycle number on16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology* (2007).

Sixt, BS, Kostanjsek, R, Mustedanagic, A, Toenshoff, ER, Horn, M. Developmental cycle and host interaction of *Rhabdochlamydia porcellionis*, an intracellular parasite of terrestrial isopods. *Environmental Microbiology* (2013).

Sogin, ML, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl Acad. Sci.* (2006).

Trifinopoulos, J, Nguyen, TL, von Haeseler, A, Minh, BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* (2016).

University of Michigan. *Animal Diversity Web*. Available through: http://animaldiversity.org/accounts/Porcellio_scaber/#80A678BC-BBCC-4CD9-9AE9-97A9F15326E3 [14.10.2016].

Vaughn, JL, Goodwin, RH, Tompkins, GJ, McCawley, P. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; *Noctuidae*). *In Vitro* (1977).

Viana, F, Buchrieser, C. Environmental treasures: co-isolation of the first marine *Chlamydiae* and its protozoan host. *Environmental Microbiology* (2016).

Wang, Q, Garrity, GM, Tiedje, JM, Cole, JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* (2007).

Wang, Y, et al. Comparison of the Levels of Bacterial Diversity in Freshwater, Intertidal Wetland, and Marine Sediments by Using Millions of Illumina Tags. *Appl. Environ. Microbiol*. (2012).

# 7 ABSTRACT

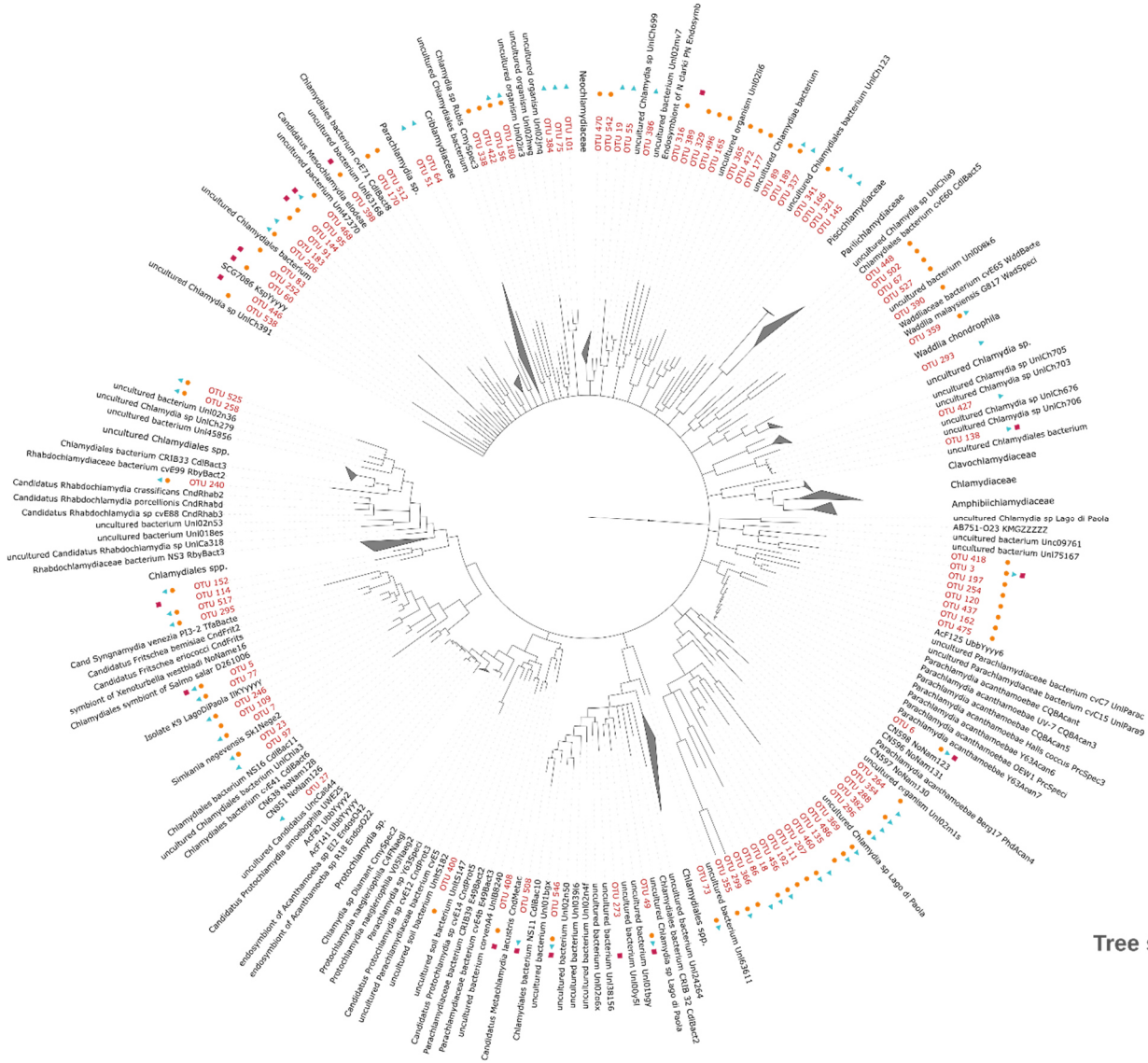
The rise of next generation sequencing techniques has enabled researches to extensively analyze microbes in different environments, thereby tremendously influencing our knowledge about the composition and dynamics of microbial communities. However, there are still some groups that are largely undiscovered due to either their low abundance or a lack of assays for their detection or suitable cultivation methods. *Chlamydiae* are one of those groups. Although some members are well studied due to their influence on human health, most chlamydiae are largely uncharacterized or even unknown. However, they are of particular interest not only due to their undiscovered diversity, wide distribution and ability to thrive in diverse animal and protist hosts but also as they are under suspicion of being emerging human pathogens. Thus, the current study focuses on extending the knowledge about these members. For that purpose, a 16S rDNA amplicon sequencing based workflow was established to reveal the distribution, diversity and population structure of *Chlamydiae*. Applying the workflow on a set of different environmental samples it could be verified that there is a huge diversity of undiscovered *Chlamydiae* and that chlamydial communities show a dynamic over time. Moreover, in order to gain a deeper insight into the evolution and adaptation mechanisms of *Chlamydiae*, another workflow was established. Here, a combination of HMW DNA isolation, cultivation, and whole genome sequencing was used to investigate genomic diversity in chlamydial populations. In the current study, the workflow was successfully applied to *Rhabdochlamydia porcellionis* that infects *Porcellio scaber* (Crustacean). However, the workflow can easily be adapted to other invertebrate host systems. Taken together the presented workflows will substantially increase our knowledge about *Chlamydiae* and contribute to complete microbial community analysis.

# 8 ZUSAMMENFASSUNG

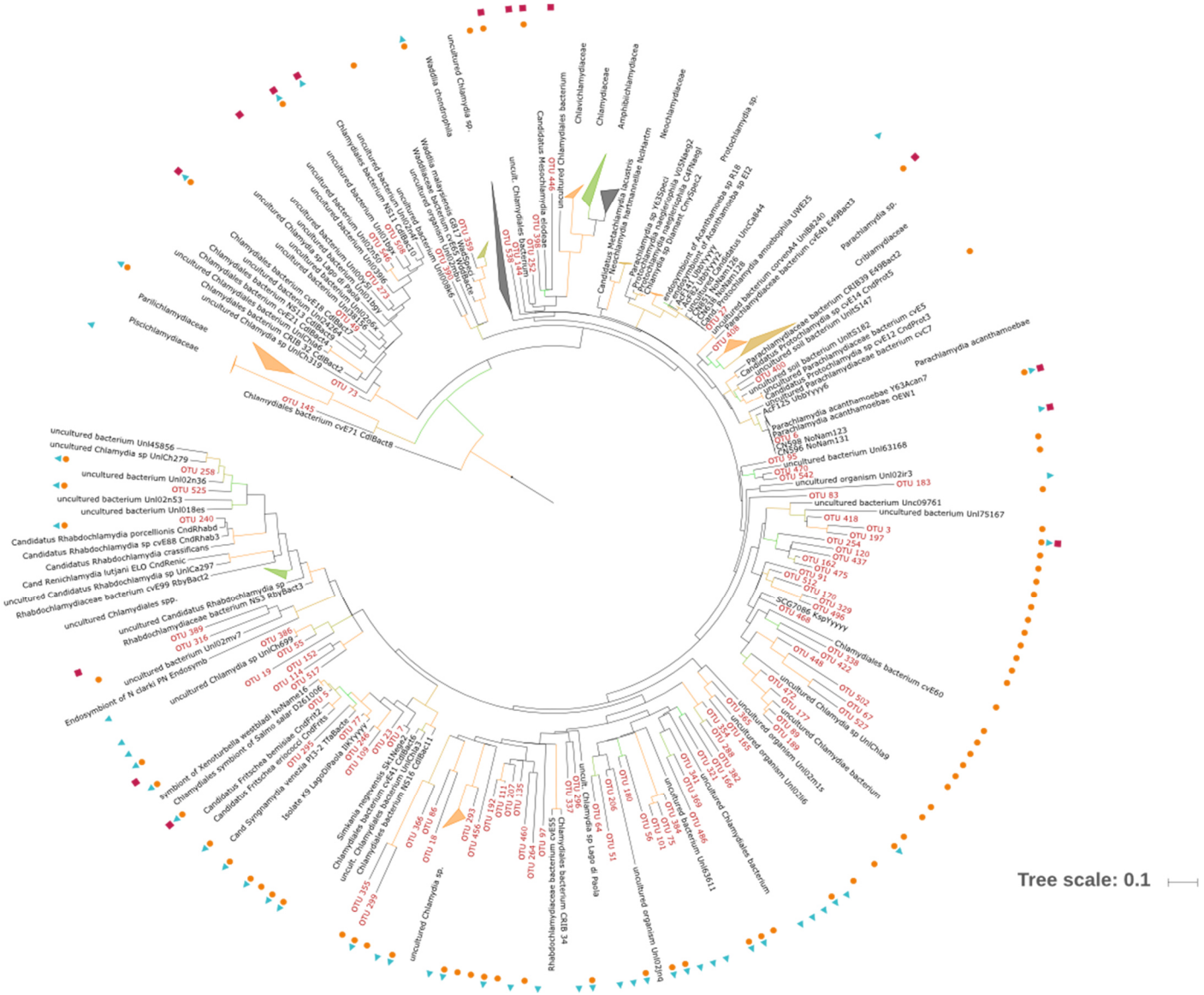
Die Fortschritte auf dem Gebiet der Hochdurchsatzsequenzierung, ermöglichten es Forschern, über die letzten Jahre Mikroorganismen in der Umwelt intensiv zu untersuchen. Dadurch wurde das Verständnis von der Zusammensetzung und Dynamik von mikrobiellen Lebensgemeinschaften substanziell verändert. Trotzdem gibt es nach wie vor Gruppen von Mikroorganismen, die sich nicht in Laborkulturen überführen lassen oder in extrem geringer Abundanz in der Umwelt vorkommen und daher weitgehend unerforscht sind. Chlamydien sind eine dieser Gruppen. Abgesehen von den humanpathogenen Arten, die schon sehr früh intensiv studiert wurden, ist ein Großteil der Chlamydien bis heute nur spärlich charakterisiert oder gar unbekannt. Doch sind gerade diese Arten von besonderer Bedeutung, nicht nur aufgrund ihrer potentiell hohen Diversität, weiten Verbreitung und Fähigkeit diverse Tiere und Protisten zu infizieren, sondern vor allem da einige Arten unter Verdacht stehen, zu Humanpathogenen zu evolvieren. Im Rahmen der aktuellen Studie wurde aus diesem Grund ein Workflow entwickelt, welcher es unter Verwendung von 16S rDNA Amplikon Sequenzierung ermöglicht, die Verbreitung, Diversität und Populationsstruktur dieser Arten zu erforschen. Durch die Anwendung des Workflows auf verschiedene Umweltproben konnte nachgewiesen werden, dass Chlamydien eine enorme Diversität umfassen und ihre Gemeinschaften eine zeitliche Dynamik aufweisen. Um einen tieferen Einblick in die Evolution und die Anpassungsmechanismen von Chlamydien zu erhalten, wurde ein zusätzlicher Workflow etabliert. Dieser ermöglicht, durch die Kombination von Kultivierung, HMW DNA Isolierung und Genomsequenzierung, Diversität in den Genomen von Chlamydien aufzudecken. In der Studie wurde der Workflow auf *Rhabdochlamydia porcellionis*, eine Art die *Porcellio scaber* (Crustacean) infiziert, angewandt. Dieser kann jedoch auch an andere Chlamydien-Wirtssysteme angepasst werden. Wie in der Studie gezeigt wurde, können die vorgestellten Workflows zukünftig substanziell zum besseren Verständnis der Verbreitung und Evolution von Chlamydien beitragen und somit zur gesamtheitlichen Erfassung und Analyse von mikrobiellen Lebensgemeinschafen.

# 9 SUPPLEMENT

Supplement 1: The figure shows the Bayesian inference tree of the phylum *Chlamydiae*. The sequences obtained in this study are shown in red. The symbols attached to the OTUs indicate the type of samples they are associated with (circle – sediment, triangle – water, square – sludge).



Supplement 2: The figure shows the maximum likelihood tree of the phylum *Chlamydiae*. The sequences obtained in this study are shown in red. The support values of the braches are indicated by their color (green – BS= 95, orange – BS= 100). Only braches with an UF-bootstrap value of ≥ 95 are shown. The symbols attached to the OTUs indicate the type of samples they are associated with (circle – sediment, triangle – water, square – sludge).



Supplement 3: The table shows all samples and contained OTUs, whereby the presence of an OTU is marked with +. Further, the total number of read assigned to an OTU are shown. The phylogenetic affiliation of the OTUs is shown in supplement 4.

| | A | B | C | D | F | G | H | L | M | $T4_1$ | $T4_2$ | $T4_3$ | $T4_4$ | $T4_5$ | $T4_6$ | $T4_7$ | $T4_8$ | $T4_9$ | $T5_2$ | $T5_4$ | $T5_7$ | $T5_9$ | $T6_1$ | $T6_3$ | VetMed | Total read count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU_6 | + | | | | | | + | | | | | | | | | | | | | | | | | | | 2 |
| OTU_3 | | | | | | | | | | + | + | + | + | + | + | + | + | + | | | | | + | | | 1442 |
| OTU_19 | | | | | | | | | | + | + | + | + | + | + | + | + | + | | | + | + | | + | | 260 |
| OTU_27 | | | | | | | | | | | | | | | | | | | + | | + | | | | | 2 |
| OTU_56 | | | | | | | | | | + | + | + | + | + | + | + | + | + | | | | + | | | | 49 |
| OTU_18 | | | | | | | | | | | | | | | | | | | + | + | + | + | + | + | | 423 |
| OTU_75 | | | | | | | | | | + | + | + | + | + | + | | + | + | | + | | | | + | | 52 |
| OTU_83 | | + | + | + | | + | | | | | | | | | | | | | | | | | | | | 16 |
| OTU_86 | | | | | | | | | | + | + | + | + | + | | + | + | + | | | + | + | + | + | | 54 |
| OTU_91 | + | + | + | + | + | + | | | | | | | | | | | | | | | | | | | | 31 |
| OTU_95 | | | | | | | | | | | | | | | | | | | | | | | | | + | 21 |
| OTU_55 | | | | | | | | | | + | + | + | + | + | | | + | + | + | + | + | + | | | | 31 |
| OTU_111 | | + | + | + | + | + | + | | | | | | | | | | | | | | | | | | | 33 |
| OTU_60 | + | + | | + | | | | | | | | | | | | | | | | | | | | | + | 88 |
| OTU_64 | | | | | | | | | | + | + | + | + | + | + | + | + | + | | + | | | + | + | | 47 |
| OTU_67 | | + | + | + | + | + | + | + | | | | | | | | | | | | | | | | | | 47 |
| OTU_120 | | + | | | + | + | + | | | | | | | | | | | | | | | | | | | 17 |
| OTU_144 | | + | | + | | | | | | | | | | | | | | | | | | | | | + | 22 |
| OTU_145 | | | | | | | | | | | | | | | | | | | | | + | + | | | | 18 |
| OTU_162 | | | | | + | + | + | | | | | | | | | | | | | | | | | | | 9 |
| OTU_170 | | | | | | | | | + | | | | | | | | | | | | | | | | | 17 |
| OTU_180 | | | | + | | | | | + | | + | + | | + | | | | + | | + | + | + | | + | | 19 |
| OTU_192 | + | + | | + | + | | + | | + | | | | | | | | | | | | | | | | | 25 |
| OTU_197 | | | | | | | | + | | | | | | | | | | | | | | | | | | 5 |
| OTU_252 | | | | | | | | | | | | | | | | | | | | | | | | | + | 8 |

| | A | B | C | D | F | G | H | L | M | $T4_1$ | $T4_2$ | $T4_3$ | $T4_4$ | $T4_5$ | $T4_6$ | $T4_7$ | $T4_8$ | $T4_9$ | $T5_2$ | $T5_4$ | $T5_7$ | $T5_9$ | $T6_1$ | $T6_3$ | VetMed | Total read count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU_296 | | + | + | + | + | + | + | + | | | | | | + | | | | | | | | | | | | 24 |
| OTU_337 | | | | | | | | | | | | + | | + | | | | | | | | | | | | 2 |
| OTU_384 | | | | | | | | | | | | + | | | | | | | | | | | + | + | | 5 |
| OTU_398 | + | + | + | | | + | | | | | | | | | | | | | | | | | | | + | 8 |
| OTU_400 | + | | | | | | | | | | | | | | | | | | | | | | | | | 2 |
| OTU_135 | + | | | | | | | | + | | | | | | | | | | | | | | | | | 13 |
| OTU_138 | | | | | | | | | | | | | | | | | | | | | | | | | + | 17 |
| OTU_408 | + | + | | | | | | | | | | | | | | | | | | | | | | | + | 12 |
| OTU_418 | | | | | | | | | + | | | | | | | | | | | | | | | | | 4 |
| OTU_437 | | + | + | | | | | | | | | | | | | | | | | | | | | | | 3 |
| OTU_456 | | + | + | | | + | + | | | | | | + | | | | | | | | | | | | | 7 |
| OTU_165 | | | | | + | + | + | + | | | | | | | | | | | | | | | | | | 14 |
| OTU_468 | | + | | + | | | | | | | | | | | | | | | | | | | | | | 4 |
| OTU_470 | | + | | | | + | | | + | | | | | | | | | | | | | | | | | 7 |
| OTU_177 | + | + | | | | | | | | | | | | | | | | | | | | | | | | 11 |
| OTU_475 | | | | | | + | + | | | | | | | | | | | | | | | | | | | 5 |
| OTU_183 | | | | | | | | | | + | + | + | + | + | + | + | + | + | | | | | | + | | 25 |
| OTU_486 | | | | | + | | | | | | + | | | | | | | | | | | | + | + | | 4 |
| OTU_512 | | | | | + | + | + | | | | | | | | | | | | | | | | | | | 3 |
| OTU_525 | + | | | | | | | | | | | | | | | | | | | | | | + | + | | 8 |
| OTU_206 | | | | | | | | | + | + | | + | + | + | | + | + | + | | | + | + | | + | | 20 |
| OTU_538 | + | + | | | | | | | + | | | | | | | | | | | | | | | | | 11 |
| OTU_542 | | + | | | + | | | | | | | | | | | | | | | | | | | | | 4 |
| OTU_546 | + | | | | | | | | | | | | | | | | | | | | | | + | | + | 6 |
| OTU_254 | | | + | | + | | + | + | | | | | | | | | | | | | | | | | | 4 |
| OTU_258 | | + | | + | | | | + | | | | | | | | | | | | | | | + | + | | 11 |
| OTU_264 | | | | | + | + | | + | | | | | | | | | | | | | | | | | | 9 |

| | A | B | C | D | F | G | H | L | M | $T4_1$ | $T4_2$ | $T4_3$ | $T4_4$ | $T4_5$ | $T4_6$ | $T4_7$ | $T4_8$ | $T4_9$ | $T5_2$ | $T5_4$ | $T5_7$ | $T5_9$ | $T6_1$ | $T6_3$ | VetMed | Total read count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU_5 | + | | | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | | 3162 |
| OTU_286 | | | | | | | | | | + | | | + | | | | | | | + | + | + | | | | 6 |
| OTU_288 | | | | | | + | + | | | + | | | | | | | | | | | | | | | | 5 |
| OTU_7 | | | + | + | | | | + | + | + | | | | | | | | | | + | | | | | | 7 |
| OTU_23 | | | | | | | | | + | + | + | + | + | + | + | + | + | + | | + | | + | | | | 178 |
| OTU_49 | + | + | | | | | | | | | | | | | | | | | | | | | + | | + | 37 |
| OTU_299 | + | | | | | | | + | | | | + | | | | | | | | | | | | | | 4 |
| OTU_316 | | | | | | + | + | + | | | | | | | | | | | | | | | | | | 9 |
| OTU_321 | | | | | | | | | | | | | | + | | | + | + | | | | + | | | | 5 |
| OTU_329 | + | | | | | | | | | | | | | | | | | | | | | | | | | 2 |
| OTU_51 | | | | | | | | | | + | + | + | + | + | | | + | + | | | + | + | | + | | 38 |
| OTU_73 | | | | + | + | + | + | | | | | | | | | | | | | | | | + | | | 30 |
| OTU_341 | | | | | | | | | + | | | | | | | | | | | | | | | | | 2 |
| OTU_354 | | | | | | | | | | | | + | + | + | | | | | | | | | | + | | 5 |
| OTU_77 | + | | + | | + | + | + | + | + | + | + | + | + | + | + | + | + | + | | + | + | + | | + | | 76 |
| OTU_359 | | | | | | + | | + | + | | | | | | | | | | | | | | | | | 3 |
| OTU_365 | | | | | | | | | + | | | | | | | | | | | | | | | | | 2 |
| OTU_89 | + | + | + | + | + | + | + | + | | | | | | | | | | | | | | | | | | 24 |
| OTU_97 | | | | | | | | | | | | + | | + | | | + | + | | + | + | + | | | | 27 |
| OTU_101 | | | | | | | | | | + | | + | + | + | + | | + | + | | | | + | + | + | | 27 |
| OTU_109 | | | | | + | + | + | + | | | | | | | | | | | | | | | | | | 39 |
| OTU_386 | | | | | | | | | | | | | | | | | | | | + | + | + | | + | | 5 |
| OTU_390 | | | | | | | | | + | | | | | | | | | | | | | | | | | 3 |
| OTU_114 | | | | | | | | | | | | | | | | | | | | | | | | | + | 13 |
| OTU_152 | | + | | | | | | | + | | | + | + | + | | | | | + | | | + | + | | | 15 |
| OTU_166 | | | | | | | | | | | | + | | | | | + | + | | | | + | | + | | 8 |
| OTU_189 | + | | + | + | | + | | | + | | | | | | | | | | | | + | + | | | | 10 |

| | A | B | C | D | F | G | H | L | M | $T4_1$ | $T4_2$ | $T4_3$ | $T4_4$ | $T4_5$ | $T4_6$ | $T4_7$ | $T4_8$ | $T4_9$ | $T5_2$ | $T5_4$ | $T5_7$ | $T5_9$ | $T6_1$ | $T6_3$ | VetMed | Total read count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| OTU_207 | + | + | + | + | | | | | | | | | | | + | | | | | | | | | | | 16 |
| OTU_427 | | | | | | | | | | | | | | | | | | | | + | + | | + | + | | 4 |
| OTU_240 | + | + | | + | | | | | | | | | | | | | | | | | | | | + | | 11 |
| OTU_246 | | | | | + | + | + | | + | | | | | | | | | | | + | | | | | | 6 |
| OTU_448 | | | | | + | | + | + | | | | | | | | | | | | | | | | | | 8 |
| OTU_273 | | | | | | | | | | | | | | | | | | | | | | | | | + | 4 |
| OTU_293 | | | | | | | | | | | | | | | | | | | | | | | + | + | | 8 |
| OTU_295 | | | | | | | | | + | | | | | | | | | | | + | + | | + | + | | 14 |
| OTU_338 | | | | | | | | + | | | | | | | | | | | | | | | | | | 1 |
| OTU_472 | | | + | | | + | | | | | | | | | | | | | | | | | | | | 2 |
| OTU_355 | | | + | | + | + | | | | | | | | | | | | | | | | | | | | 3 |
| OTU_496 | | | + | | + | + | | + | | | | | | | | | | | | | | | | | | 5 |
| OTU_502 | | | | | | | + | + | | | | | | | | | | | | | | | | | | 6 |
| OTU_369 | | | | | | + | | | | + | + | + | | + | + | | | + | | + | + | + | | + | | 16 |
| OTU_382 | | | | | | | | | | | | | | | | | | | + | | | + | | | | 2 |
| OTU_517 | + | | | | | | | | | + | | + | | + | + | | | + | | | | | | + | | 8 |
| OTU_422 | | | | + | | | | + | | | | | | | | | | | | | | | | | | 3 |
| OTU_527 | | | | | | | | + | | | | | | | | | | | | | | | | | | 2 |
| OTU_460 | | | | | | | | | | + | | + | | | | | | | | + | | | | | | 3 |
| OTU_508 | | | | | | | | | | | | | | | | | | | | | | | | + | + | 7 |

Supplement 4: The table shows the RDP classification on genus level of all observed OTUs.

| | Genus |
|---|---|
| OTU_6 | *Parachlamydia* |
| OTU_3 | unclassified *Chlamydiales* |
| OTU_19 | unclassified *Parachlamydiaceae* |
| OTU_27 | *Candidatus Protochlamydia* |
| OTU_56 | unclassified *Parachlamydiaceae* |
| OTU_18 | unclassified *Chlamydiales* |
| OTU_75 | *Candidatus Metachlamydia* |
| OTU_83 | unclassified *Parachlamydiaceae* |
| OTU_86 | *Candidatus Metachlamydia* |
| OTU_91 | unclassified *Parachlamydiaceae* |
| OTU_95 | *Candidatus Protochlamydia* |
| OTU_55 | *Criblamydia* |
| OTU_111 | unclassified *Parachlamydiaceae* |
| OTU_60 | unclassified *Chlamydiales* |
| OTU_64 | *Criblamydia* |
| OTU_67 | *Criblamydia* |
| OTU_120 | unclassified *Parachlamydiaceae* |
| OTU_144 | *Candidatus Protochlamydia* |
| OTU_145 | *Neochlamydia* |
| OTU_162 | unclassified *Parachlamydiaceae* |
| OTU_170 | *Neochlamydia* |
| OTU_180 | unclassified *Parachlamydiaceae* |
| OTU_192 | *Candidatus Metachlamydia* |
| OTU_197 | unclassified *Parachlamydiaceae* |
| OTU_252 | *Candidatus Protochlamydia* |
| OTU_296 | *Candidatus Metachlamydia* |
| OTU_337 | unclassified *Parachlamydiaceae* |
| OTU_384 | *Candidatus Metachlamydia* |
| OTU_398 | *Neochlamydia* |
| OTU_400 | *Candidatus Protochlamydia* |
| OTU_135 | unclassified *Chlamydiales* |
| OTU_138 | unclassified *Chlamydiales* |
| OTU_408 | *Candidatus Protochlamydia* |
| OTU_418 | unclassified *Parachlamydiaceae* |
| OTU_437 | unclassified *Parachlamydiaceae* |
| OTU_456 | *Neochlamydia* |
| OTU_165 | unclassified *Chlamydiales* |

| | |
|---|---|
| OTU_468 | unclassified *Parachlamydiaceae* |
| OTU_470 | *Neochlamydia* |
| OTU_177 | unclassified *Chlamydiales* |
| OTU_475 | *Neochlamydia* |
| OTU_183 | unclassified *Chlamydiales* |
| OTU_486 | *Neochlamydia* |
| OTU_512 | *Parachlamydia* |
| OTU_525 | *Neochlamydia* |
| OTU_206 | unclassified *Chlamydiales* |
| OTU_538 | *Neochlamydia* |
| OTU_542 | *Neochlamydia* |
| OTU_546 | unclassified *Parachlamydiaceae* |
| OTU_254 | unclassified *Chlamydiales* |
| OTU_258 | unclassified *Chlamydiales* |
| OTU_264 | unclassified *Chlamydiales* |
| OTU_5 | *Candidatus Fritschea* |
| OTU_286 | unclassified *Chlamydiales* |
| OTU_288 | unclassified *Chlamydiales* |
| OTU_7 | *Simkania* |
| OTU_23 | *Candidatus Fritschea* |
| OTU_49 | *Simkania* |
| OTU_299 | unclassified *Chlamydiales* |
| OTU_316 | unclassified *Chlamydiales* |
| OTU_321 | unclassified *Chlamydiales* |
| OTU_329 | unclassified *Chlamydiales* |
| OTU_51 | *Candidatus Fritschea* |
| OTU_73 | *Simkania* |
| OTU_341 | unclassified *Chlamydiales* |
| OTU_354 | unclassified *Chlamydiales* |
| OTU_77 | *Candidatus Fritschea* |
| OTU_359 | *Waddlia* |
| OTU_365 | *Criblamydia* |
| OTU_89 | unclassified *Simkaniaceae* |
| OTU_97 | *Candidatus Fritschea* |
| OTU_101 | *Simkania* |
| OTU_109 | *Candidatus Fritschea* |
| OTU_386 | unclassified *Chlamydiales* |
| OTU_390 | *Waddlia* |
| OTU_114 | *Simkania* |
| OTU_152 | unclassified *Simkaniaceae* |

| | |
|---|---|
| OTU_166 | unclassified *Simkaniaceae* |
| OTU_189 | unclassified *Simkaniaceae* |
| OTU_207 | unclassified *Simkaniaceae* |
| OTU_427 | unclassified *Chlamydiales* |
| OTU_240 | unclassified *Simkaniaceae* |
| OTU_246 | *Candidatus Fritschea* |
| OTU_448 | *Criblamydia* |
| OTU_273 | *Simkania* |
| OTU_293 | *Simkania* |
| OTU_295 | *Candidatus_Fritschea* |
| OTU_338 | unclassified *Simkaniaceae* |
| OTU_472 | unclassified *Chlamydiales* |
| OTU_355 | *Candidatus Fritschea* |
| OTU_366 | *Simkania* |
| OTU_496 | *Criblamydia* |
| OTU_502 | *Criblamydia* |
| OTU_369 | unclassified *Simkaniaceae* |
| OTU_517 | unclassified *Chlamydiales* |
| OTU_422 | *Candidatus Fritschea* |
| OTU_527 | *Criblamydia* |
| OTU_460 | *Simkania* |
| OTU_508 | unclassified *Simkaniaceae* |

# 10 CURRICULUM VITAE

## Education

| | |
|---|---|
| *Since 10/2014* | Master degree course in Molecular Microbiology, Microbial Ecology and Immunobiology, University of Vienna |
| *Since 03/2014* | Bachelor degree course in Bioinformatics, University of Vienna |
| *2011 – 2014* | Bachelor degree course in Biology, University of Vienna |
| *2006 – 2011* | Secondary college for occupation in economic professions, Pannoneum Neusiedl/See |

## Work Experience

| | |
|---|---|
| *08/2015* | Internship Baxalta, Vienna |
| *06/2009 – 09/2009* | Internship Donauturm Aussichtsturm- und Restaurantbetriebs GmbH, Vienna |

## Skills

| | |
|---|---|
| *Language*: | German (native speaker)<br>English (fluent)<br>French (basic) |
| *IT-Skills*: | Word, Power Point, Excel, Outlook<br>C++, SQL, HTML/CSS, Java<br>Windows, Linux |