universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

„Establishment of a Suitable Multilocus Sequence Analysis Procedure for Determining the Diversity of *Vibrio cholerae* Strains in Lake Neusiedler See“

verfasst von

Carina Pretzer

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 441
Studienrichtung lt. Studienblatt:	Diplomstudium Genetik - Mikrobiologie
Betreut von:	ao. Univ.Prof.Dr. Angela Witte

Die Diplomarbeit wurde durchgeführt am Institut für Hygiene und Angewandte Immunologie, Abteilung für Wasserhygiene, Medizinische Universität Wien.
Betreuer: Ass. Prof. Mag. Dr. Alexander Kirschner

Die phylogenetischen Analysen wurden unter der Betreuung von
Doz. Mag. Dr. Irina Druzhinina
vom Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften, Arbeitsgruppe Mikrobiologie, TU Wien durchgeführt.

ACKNOWLEDGEMENT

Ein ganz großes Dankeschön möchte ich meinen Eltern aussprechen…

Danke für Eure unermüdliche Bestärkung in allen Lebenslagen und Euer unerschütterliches Vertrauen in mich, meine Pläne und Vorhaben!

Ohne Euren beständigen Rückhalt hätte ich vielleicht nie das Durchhaltevermögen und die Kraft aufgebracht, die manchmal nötig waren, um mein Studium voranzubringen. Natürlich danke ich Euch auch für die finanzielle Unterstützung ohne die Studium und Leben in Wien nicht möglich gewesen wären.

Ich bedanke mich ganz herzlich bei Dr. Alexander Kirschner, in dessen Labor ich diese Diplomarbeit durchführen durfte, bei Dr. Andreas Farnleiter, der den Kontakt zu Dr. Irina Druzhinina hergestellt hat, die mir bei der Auswertung der Datenmengen zur Seite stand und mir in kurzer Zeit viel beigebracht hat - Thank you so much!

Ein weiterer Dank gilt Dr. Angela Witte für Ihren Einsatz.

Unerwähnt dürfen auch Mag. Sonja Schauer und Mag. Barbara Schrammel nicht bleiben, die in den 12 Monaten im Labor immer ein offenes Ohr für mich hatten, mir mit Ihrer Erfahrung zur Seite standen und maßgeblich an der lockeren Atmosphäre am Zimmermannplatz beteiligt waren.

Bedanken möchte ich mich auch bei meinen Freunden, die mich teilweise schon seit der Schulzeit begleiten oder die ich hier in Wien gefunden habe. Danke Annika, Edith, Kathi, Lisa und Ricki!

Ich danke dir mein Schatz, Richard, dass du immer mein Fels in der Brandung bist.

DANKE Euch allen für Eure Unterstützung, fürs Interesse, fürs Lernen, fürs Zuhören, fürs Zusammenarbeiten, fürs Beraten, für die schöne Zeit**!**

„Die Wissenschaft ist ewig in ihrem Quell,
unermesslich in ihrem Umfang,
endlos in ihrer Aufgabe,
unerreichbar in ihrem Ziel.“

Karl Ernst von Baer (1792 - 1876)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

°C	degree Celsius
µg	microgram
µl	microliter
µm	micrometer
ATP	adenosine triphosphate
bp	basepair
CARD-FISH	Catalyzed Reporter Deposition - Fluorescence *In Situ* Hybridization
CFU	colony forming units
cm	centimeter
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTP	desoxyribonukleosidtriphosphate
dNTPs	desoxynucleoside triphosphates
e.g.	for example (from *exemplī grātia*)
et al.	and others (from *et alii*)
EtBr	ethidium bromide
h	hours
$MgCl_2$	magnesium chloride
ml	milliliter
MLSA	multilocus sequence analysis
mm	milimeter
NCBI	National Center for Biotechnology Information
no.	number (from *numero*)
nt	nucleotide
Ø	diameter
omp	outer membrane protein
OTU	operational taxonomic unit
PCR	polymerase chain reaction
pmol	picomole
P_{tot}	total phosphorus
rcf	relative centrifugation force
rpm	revolutions per minute
SPC	Solid Phase Cytometry
TAE	tris acetat ethylendiamintetracid
TSS	total suspended solids
U	units
UDP	uridine diphosphate
UMP	uridine monophosphate
UV	ultraviolet
V	volt

1 INTRODUCTION

Vibrio cholerae belongs to the Gammaproteobacteria and to the family of Vibrionaceae, which also comprises such genera as *Allomonas*, *Catenococcus*, *Enterovibrio*, *Grimontia*, *Listonella*, *Photobacterium* and *Salinivibrio.* These bacteria are widely distributed in aquatic environments worldwide from brackish to deep-sea water; it is now accepted knowledge that the evolution of *Vibrio* spp. has occurred out of marine environments (83).

V. cholerae is a natural inhabitant of aquatic ecosystems (12, 15), preferring a saline environment (73). It is a rod shaped bacterium with a single polar flagellum allowing fast swimming (83). Two electron microscopy images of *Vibrio cholerae* by courtesy of Dr. Susanne Richter (Austrian Agency for Health and Food Safety, AGES) are shown in **Figure 1.**

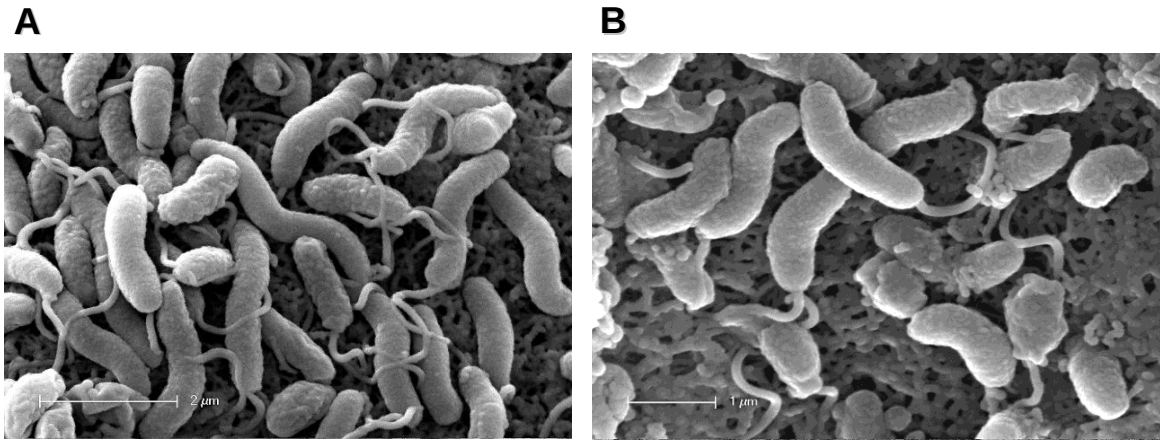


Figure 1. Electron-microscopy images of *Vibrio cholerae*

A: Scanning-electron-microscopy, scale = 2 µm, source: Dr. Susanne Richter, AGES Mödling;
B: Scanning-electron-microscopy, scale = 1 µm, source: Dr. Susanne Richter, AGES Mödling.

V. cholerae is a human pathogen like several species in the genus Vibrio are, for example *V. parahaemolyticus* and *V. vulnificus* (83).

For *V. cholerae* more than 200 serogroups are known, differentiated by the structure of their lipopolysaccharides, but only two of them, O1 and O139, are able to cause the disease cholera. However, non-O1/non-O139 were reported to cause diarrhea, wound infections or outbreaks of gastroenteritis (7, 33, 35). Serogroup O1 can be subdivided into the serotypes, Ogawa and Inaba and the rarely reported Hikojim, which are further subdivided in biotypes with different biochemical properties, Classical and ElTor. The two cholera associated serogroups O1 and O139 can be further differentiated according to genetic variations like antibiotic or phage resistance properties (25, 31, 88).

Since 1817, cholera has swept the world in seven major pandemics (54). It is an acute life-threatening diarrhoeal disease caused by *V. cholerae* contaminated food or water, which now causes 100,000 – 120,000 deaths every year of whom only a small proportion are reported to WHO (91). The reported number of cholera deaths caused by *V. cholerae* in the years 2009 to 2011 counts a total number of 20,000 (92-94).

Developing countries, particularly in such regions as South-East Asia, Africa and Latin America have been highly affected by cholera outbreaks or cholera epidemics due to the lacking hygiene standards, and because the access to safe drinking water and adequate sanitation are not always guaranteed (22, 39, 91).

PATHOGENESIS AND *VIBRIO CHOLERAE* GENOME

After ingestion of contaminated food or water the acid barrier of the stomach will be passed and the bacteria reach the small intestine. Generally an ingestion of 10^8 *Vibrio cholerae* is needed to cause disease in healthy persons but a much lower dose of 10^5 *V. cholerae* is sufficient if given with antacids to neutralise the stomach acid (34, 69, 70) but if ingested with food even reduced numbers can lead to infection because food might protect them from the stomach activity (54). Mediated by the toxin co-regulated pili (encoded by the *tcp* genes) *V. cholerae* is able to colonize the epithelium. Then cholera toxin (encoded by *ctxA* and *ctxB of the ctxAB* operon) is produced in two steps. First, the B-subunits of the enterotoxin bind the gangliosid GM1, allowing the release of the A subunit across the cell membrane. Subunit A activates epithelial adenyl cyclase, which leads to disruption of sodium influx of the epithelial cells - resulting in loss of water and electrolytes. This watery diarrhoea can lead to dehydration and death if not treated (39, 54).

The pathogenesis of *V. cholerae* is encoded in virulence genes like *ctx* and *tcp* genes. Whereas *ctx* is part of a lysogenic filamentous bacteriophage, named CTX Φ (86), the *tcp* genes comprise a large cluster of genes which is encoded on chromosome I (24). In the year 2000, the complete genome of *V. cholerae* O1 biovar El Tor strain N16961 was sequenced (32). The genome consists of two chromosomes, Chr. I with 2,961,149 bp, and Chr. II with 1,072,315 bp circular DNA, together encoding more than 3,800 proteins (60).

Serogroups other than O1 and O139 are not linked to epidemic and pandemic disease and are therefore collectively named non-O1/non-139 serogroups (25). Non-O1/non-O139 strains were reported to cause mild to moderate severe watery diarrhoea, or wound, ear and blood infections (7, 59). Nevertheless horizontal gene transfer provides

the possibility transforming a non-O1/non-O139 into a pathogen, as there is evidence that also the O139 strain emerging 1992 in India arose from a O1 strain (2, 3, 9, 13, 38).

OCCURRENCE OF *VIBRIO CHOLERAE* IN AUSTRIA AND PREVIOUS INVESTIGATIONS

13 infections with non-O1/non-O139 *V. cholerae* were reported in Austria in the years 2000 to 2005. For eight of these patients which suffered from otitis media (inflammation of the middle ear), or otitis externa (inflammation of the outer ear or ear canal), and one lethal sepsis, a stay abroad was not associated, and for five of them a visit at Lake Neusiedler See was documented (35). For this reason, a surveillance program was started in 2001 to observe the ecology, genetic diversity and potential pathogenicity of *V. cholerae* by Kirschner and colleagues (FWF-project P21625-B20), in which *V. cholerae* has been detected not only in lake Neusiedler See but also in its adjacent shallow saline alkaline lakes (46, 71).

The Lake Neusiedler See is a large steppe lake spanning the Austrian-Hungarian border with a surface area of 321 km^2. 80 % of the total area is located in the easternmost Austrian state Burgenland and 20 % is on Hungarian state territory (36). A 178 km^2 big reed belt, composed of *Phragmites australis*, covers the lake, which represents 55 % of the total area. The maximum depth is 1.8 m and mean depth is about 1.1 m. The pH amounts 8.5 to 9.1 and salinity is 1-2 g/L which are ideal conditions for *V. cholerae,* which is known to be a moderately halophilic bacterium (45, 57).

In the period from 2001-2004 a monitoring program was investigating the presence of *V. cholerae* in Lake Neusiedler See. In the course of the study 300 culture based isolates were identified as non-O1/non-O139 by agglutination tests and about 133 isolates were observed by PCR analysis for the presence of *ctx* or *tcp*, with the result that no isolate contained one of those pathogenicity-related genes (46).

In a previous pulsed-field gel electrophoresis (PFGE) study (29) 84 isolates of two different lake habitats, open water and reed belt, were analysed to determine *V. cholerae* genotypes, in relationship to their seasonal occurrence, differences in sampling sites and their association with zooplankton (45). Out of 84 isolates 64 different genotypes could be identified and four of them were found at both sampling sites. Three genotypes originating from zooplankton isolates were also found in open water samples, but a seasonal trend wasn't confirmed (29). To assess the genetic diversity of *Vibrio cholerae* in the lake, further investigations were needed and therefore multilocus sequence analysis (MLSA) was chosen.

DIFFERENTIATION OF *V. CHOLERAE* BY MULTILOCUS SEQUENCE ANALYSIS (MLSA)

For identifying bacteria at strain level several molecular methods have been developed, in the past moving more and more towards computer-assisted design/analysis and increased automation. Pulsed-field gel electrophoresis, restriction fragment length polymorphism, arbitrarily primed PCR/random amplification of polymorphic DNA, repetitive sequencing-based PCR (like Multilocus variable number tandem repeat analysis), denaturing gradient gel electrophoresis, high-resolution melting analysis, multilocus sequence typing, sequencing of 16S-23S rRNA gene internal transcribed spacer, whole genome sequencing and DNA hybridization-based methods are useful tools, each with advantages and disadvantages also in reference to resolution, throughput and reproducibility (52).

MLSA basically is sequence comparison of selected housekeeping genes and is part of multilocus sequence typing (MLST), an approach first described in 1998 (55). In 2003, Kotetishvili et al (49), showed that MLSA discriminates better different *V. cholerae* types than PFGE does.

In MLSA genetic loci serve as phylogenetic markers to identify and subtype species and to perform phylogenetic studies (72, 81, 84). Housekeeping genes have been usually used for this approach (40, 72, 84). Housekeeping genes are constitutively expressed genes, required for basic cellular functions, like genes involved in transcription or translation or genes encoding for ribosomal proteins, and therefore also proposed as “maintenance” genes (6). These genes are highly expressed under all conditions (21). As housekeeping genes encode essential biochemical functions nucleotide changes are restricted and therefore they are evolving slowly and selectively neutral (56, 74). The ensuing minimal allelic variations within a population provide high levels of discrimination if many loci are analysed (55) and the population structure can be defined (74).

Virulence genes, or virulence-associated genes, are usually more polymorphic and diversify more rapidly (78) and are therefore applicable to discriminate better between specific clones (14) and have been used in multiple studies (27, 49, 51, 78).

Since there is no widely accepted species concept for prokaryotes, the assignment of isolates to species is an integration of phenotypic and genotypic features and population genetics, ecology and genomics are linked to it and should be considered (28). MLST/MLSA were approved in several studies for analyzing population structure and evolutionary relationship (56) and also for taxonomic usage, and presents a ‘baseline’ for species assignment (28).

WHY NOT 16S RRNA?

The 16S rRNA gene is a highly conserved region present in the genome of all prokaryotes. It encodes the 16S rRNA, which functions as a part of the small subunit of the ribosome (54).

16S rRNA sequence-based methods were reported as methods of choice for phylogenetic analysis in microorganisms in the late 1980s (10, 90). Although the applicability of 16S rRNA gene analysis for distinguishing highly related species is limited, especially for those with 98% 16S rRNA sequence similarity (1); such highly conserved genes usually can't unveil intra-species of bacteria (56).
Vibrio mimicus is *V. cholerae's* nearest neighbour and was seen as a biotype of *V. cholerae* in the past, but now it is known as a different species (87). Kita-Tsukamoto and colleagues determined in 1993 (47) the phylogenetic relationships of different marine bacteria, containing *Vibrio cholerae* ElTor and Classical and *Vibrio mimicus*, on the basis of 16S rRNA sequences. In this study five strains of *V. cholerae* formed one cluster along with *V. mimicus* with a large distance from them to other Vibrios. Similarity levels ranging from 99.06 to 100.00 % and K values (evolutionary distance per site) from 0.00 to 0.78 indicated that these are very closely related organisms. Later on these two species were reported to have nearly identical 16S rRNA, differing only in 6 out of 1,456 nucleotides (9, 68).
A study in 2003 tested the applicability of the DNA sequence of 32 widely distributed bacterial genes to assign strains to species, with the outcome that 16S rRNA, is not able to differentiate closely related taxa because of its high conservation (96).

SELECTED GENES AND ENCODING PROTEINS

The genes chosen for this multilocus sequence approach are: *gyrB*, *recA*, *toxR*, *pyrH* and *rpoA*; their functions are described briefly in the following.

DNA gyrase subunit B, encoded by the gene ***gyrB***, is a Type II topoisomerase. In bacterial cells there are two topoisomerases: gyrase and topoisomerase IV, which are both needed in all phases of DNA replication. DNA gyrase introduces supercoils in the DNA strands, important for unwinding them in initiation, minimizing stress in front of the replication fork and packing in termination (43).

The ***recA*** gene encoded protein, recombinase-A, is involved in bacterial homologous recombination reactions, where sequence exchange between up to four strands can occur, as well as in the repair of stalled replication forks and double-strand break repair. Another task of recombinase-A is SOS response and SOS mutagenesis which are global

responses to DNA damage in which cell cycle is arrested and DNA repair and mutagenesis is induced; with its protease activity it can cleave various repressors and more than 15 SOS repair genes were activated(16, 53).

The transmembrane protein **TOXR** is the cholera toxin transcriptional activator, a membrane protein, important for the regulation of virulence genes. The expression of the *tcp* and *ctx* operons, which genes encode for TCP and CT, the key proteins in *V. cholerae* pathogenicity, is regulated by TOXR and TOXS. TOXR binds directly to the promotors of the *ctx* operon, the genes encoding for outer membrane proteins OMPU and OMPT and the *toxT* gene, and affects their expression directly. TOXT is another downstream regulator in the *toxR* regulon, which can then independently activate *ctx* and *tcp* promotors (4).

PyrH encodes a protein called uridylate kinase, a key enzyme in pyrimidine biosynthesis. Dependent on ATP uridine monophosphate is phosphorylated to uridine diphosphate which is a precursor for the synthesis of uridine triphosphate and cytosine triphosphate. Uridylate kinases are present in bacteria, archaea and eukarya (85).

The gene ***rpoA*** encodes the DNA-directed RNA polymerase subunit alpha, which is the smallest subunit of the RNA polymerase holoenzyme (RNAP). It is involved in RNAP assembly, in promotor recognition by sequence specific protein-DNA interaction and in transcriptional regulation via a large set of transcriptional activator proteins (18).

Figure 2 shows the position of the five chosen loci for this MLSA on chromosome I of *Vibrio cholerae* O1 biovar ElTor str. N16961, which is important because there is a need of unlinked loci.

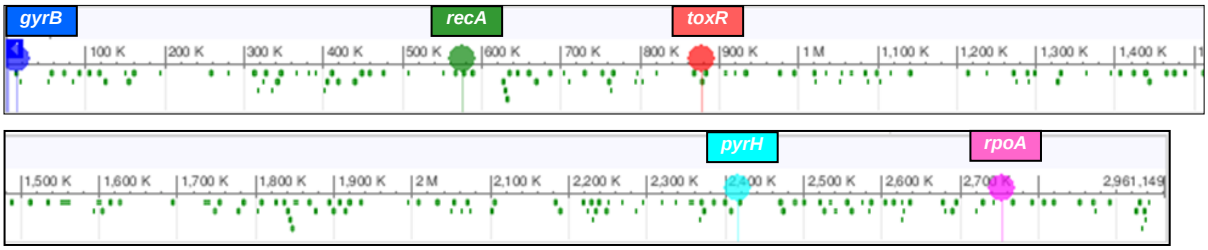


Figure 2. Position of the five chosen loci (*gyrB, recA, toxR, pyrH* & *toxR*) on Chromosome I of *Vibrio cholerae* O1 biovar El Tor str. N16961.

AIMS OF THE STUDY

The aim of this study was to establish a suitable multilocus sequence analysis procedure to assess the diversity and genetic variability of *Vibrio cholerae* at Lake Neusiedler See. For this purpose the following objectives have been formulated:

- Collection of 300 isolates out of three different sampling sites in the lake, representing three different habitats.
- Screening for *ompW* (encoding a specific outer membrane protein of *V. cholerae*) and identify *Vibrio cholerae.*
- Selection of loci to be used for this MLSA.
- Amplification of genes by PCR.
- DNA Sequencing.
- Do phylogeny.
- Estimate the diversity of *V. cholerae* in the lake.

2 MATERIAL AND METHODS

2.1 SAMPLING SITES

The investigated water samples originated from three different locations in Lake Neusiedler See/Austria and were collected in July 2012. The three sampling sites are named by their position within the lake respectively by the representing habitat: open water, reed belt and intermediate habitat. The geographical position of the lake within Central Europe is shown in **Figure 3**, and a detailed location of the three sampling points in **Figure 4.**

Figure 3. Map of central Europe.
Blue arrow indicates location of Lake Neusiedler See. Scale 200 km. Location data © 2013 GeoBasis-DE/BKG (©2009), Google.

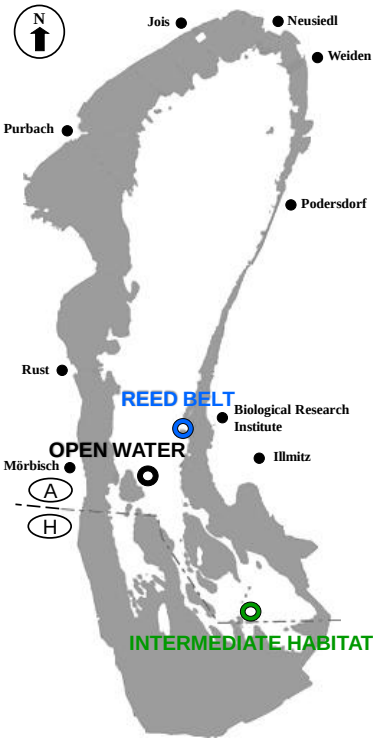


Figure 4. View of Lake Neusiedler See.
View of Lake Neusiedler See with the surrounding municipalities and the three sampling sites, indicated by coloured circles. The dark shaded area indicates the reed belt and the light shaded area indicates open water.

2.2 ISOLATION AND IDENTIFICATION OF PRESUMPTIVE *VIBRIO CHOLERAE*

2.2.1 COLLECTION AND PREPARATION OF ENVIRONMENTAL ISOLATES

Isolates and strains used in this study were listed in **Table 1**.
Water samples were collected in July 2012, from three sampling sites of Lake Neusiedler See, tagged in **Figure 4.** 500 ml water sample were collected using sterile 500 ml glass-bottles in duplicates out of a depth of 30 cm under the surface of the lake and transported to the laboratory in a box at ambient temperature.

Diverse parameters like water temperature, Total Phosphorus (P_{tot}), Total Suspended Solids (TSS), Chlorophyll-a, Dissolved Organic Carbon and Secchi depth (a measure for turbidity) were measured for each sample (46). Total bacterial cell numbers were determined by SybrGold staining according to a modified protocol of Riepl et al, 2011 (64). Additionally, cultivation-based quantification of *Vibrio cholerae* (71) and quantification with Catalyzed Reporter Deposition Fluorescence In Situ Hybridization combined with Solid-Phase Cytometry (71) was performed.
For cultivation, different volumes of the water samples were filtered through 0.45 µm pore size cellulose nitrate filters (Ø 50 mm, Sartorius Stedim Biotech, Göttingen, Germany). From sampling site intermediate habitat 10 × 1 ml water were used, from sampling site open water 10 × 5 ml and from sampling site reed belt, 5 × 1 ml and 5 × 0.5 ml water sample were filtered. The filters were placed on Thiosulfate Citrate Bile Sucrose agar plates (TCBS, Merck, Darmstadt, Germany) and incubated for 18 h at 37 °C (Inucell, MMM Group, Munich, Germany). Yellow, round, 1 to 2 mm in diameter colonies were picked and streaked onto nutrient agar plates without sodium chloride (see **7.1.1 Media**) and then incubated (Inucell) for 24 h at 37 °C. If the culture was grown, half of each bacterial smear was then scraped off and used to inoculate a cryo-vial (Roti®-Store Cryo-vials, Roth GmbH, Karlsruhe, Germany). After mixing and removing the supernatant, the cryo-vial was frozen with liquid nitrogen and stored at -80 °C as a sample bank. The remaining bacterial smear was also scraped off and used to inoculate 500 µl of 1 × phosphate buffered saline (PBS), and stored at -20 °C till usage for subsequent DNA-extraction.

This procedure was extended till 100 presumptive *V. cholerae* isolates of each sampling site were frozen in duplicates, in a cryo-vial and in 500 µl PBS.

2.2.2 PREPARATION OF CLINICAL *VIBRIO CHOLERAE* ISOLATES

Additionally, eight *V. cholerae* samples isolated from patients in the years from 2008 to 2012 were provided by the Institute of Medical Microbiology and Hygiene, AGES. Seven of the isolates have been identified as *V. cholerae* non-O1/non-O139 and one (920001/12) was diagnosed according to examinations of the AGES laboratory as suspicious for O1 subtype Inaba. The cultures were streaked onto nutrient agar plates without sodium chloride and subsequently handled like the isolates of Lake Neusiedler See, as previously described in **2.2.1.**

Table 1. Isolates used in this study.

Isolate no.	Date of Isolation	Place of Isolation	Source
		Environmental	
01 - 100	2012-09-07	Austria, Lake Neusiedler See, Intermediate habitat	Water
101 - 200	2012-09-07	Austria, Lake Neusiedler See, Open water	Water
201 - 300	2012-09-07	Austria, Lake Neusiedler See, Reed belt	Water
		Clinical	
CHT 81/08	2008	Austria	Human
CHT 83/08	2008	Austria	Human
CHT 84/08	2008	Austria	Human
CHT 85/09	2009	Austria	Human
CHT 86/09	2009	Austria	Human
920003/10	2010	Austria	Human
920001/11	2011	Austria	Human
920001/12	2012	Austria	Human
		Reference strains*	
***Vibrio cholerae* M66-2**	1937	Indonesia	Human
***Vibrio cholerae* O395**	1965	India	Human
***Vibrio cholerae* O1 biovar ElTor str. N16961**	1971	Bangladesh	Human
***Vibrio cholerae* MJ-1236**	1994	Bangladesh	Human

* Sequences obtained from GenBank – NCBI

2.2.3 REFERENCE STRAINS

Four different *Vibrio cholerae* strains, which whole genome sequences are known and available on the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) were included in the phylogenetic analysis for comparison of these cholera causing strains - all of clinical origin - with the ones from Lake Neusiedler See and the clinical strains from Austria. ***Vibrio cholerae* M66-2:** Outbreaks between 1937 and 1957 in Indonesia were called as pre 7th pandemic outbreaks. It belongs to serogroup O1 and was isolated in 1937 in Makassar, Indonesia (26). ***Vibrio cholerae* O395:** This strain belongs to serogroup O1, serotype Ogawa, classical. O395 was isolated in 1965 in India and was sequenced by random shotgun sequencing, as M66-2 was too, published 2008 (26). ***Vibrio cholerae* O1 biovar ElTor str. N16961:** Belongs to serogroup O1 and was isolated in Bangladesh in 1971. The genome was sequenced by the whole genome random sequencing method and published in 2000 (32). ***Vibrio cholerae* MJ-1236:** MJ-1236 belongs to serogroup O1, serotype Inaba, Classical. It was isolated in Matlab, Bangladesh in 1994 (8).

2.2.4 DNA-EXTRACTION

DNA was extracted by boiling the isolates at 99 °C for 10 min at 400 rpm in a Thermomixer comfort (Eppendorf, Hamburg, Deutschland). After subsequent centrifugation (Hermle Z233 MK-2, Rotor 220.59 VO4) at 12.000 rcf for 10 min at 4 °C the supernatant was diluted 1:10 (further noted as 10^{-1} dilution) with water (DEPC treated, sterile filtered, Sigma Life Sciences, Swiss) and used as DNA template for PCR.

2.2.5 SCREENING FOR *VIBRIO CHOLERAE* BY *OMPW* PCR

The *V. cholerae* specific outer membrane protein encoding gene *ompW* was amplified by PCR for each isolate of Lake Neusiedler See. Used primer sequences for *ompW*-PCR are: forward, `5′CAC CAA GAA GGT GAC TTT ATT GTG 3′;` reverse, `5′ GGT TTG TCG AAT TAG CTT CAC C 3′` (30).

PCR mixtures, based on a multiplex PCR (5), composed of 12.2 µl water (DEPC treated, sterile filtered, Sigma Life Sciences), 2.5 µl reaction buffer B [10×] (Solis Biodyne, Tartu, Estonia), 2 µl magnesium chloride [0.025mol/l] (Promega, Madison, USA), 1.2 µl forward primer *ompW* [10 pmol/µl], 1.2 µl reverse primer *ompW* [10 pmol/µl] (Thermo Fisher Scientific GmbH, Ulm, Germany), 0.5 µl desoxynucleoside triphosphates [20mM each] (Solis BioDyne), 0.4 µl HOT FIREPol® DNA polymerase [5 U/µl] (Solis BioDyne) and 5.0 µl template DNA, to a final volume of 25 µl per reaction. The mix was then vortexed (lab dancer, IKA, Germany), and span down with a microcentrifuge (Qualitron, Pakistan).

PCR was subsequently performed with a Biometra T Gradient Thermoblock (Biometra, Göttingen, Germany). The thermal program was set as 15 min at 94 °C (initial denaturation), 30 cycles of 30 sec at 94 °C (denaturation), 30 sec at 59 °C (annealing), 2 min at 72 °C (elongation) and 10 min at 72 °C (final elongation).
A negative control, with water (Sigma Life Sciences) and a positive control, a *V. cholerae* O139 strain as DNA template, were additionally carried along in each PCR approach.

2.2.5.1 Amplification control via gel electrophoresis

Amplified PCR products were checked by gel electrophoresis (Electrophoresis System Perfect Blue Gel System, PeqLab, Erlangen, Germany). 3 µl marker and 5 µl sample (5 µl sample mixed with 1 µl 6×loading dye) were loaded onto a 2 % agarose gel and the DNA fragments were separated with 100-120 V (Power Supply, Biorad) for 70-90 minutes in 1× TAE buffer (40 mM Tris, 1 mM EDTA, pH 8) and additionally stained in EtBr-solution (1 µg/ml) for 20 minutes, subsequently washed in 1x TAE for 15 minutes and visualized under UV light (GelDoc, PeqLab) with the imaging system Quantum ST4v16.03 (PeqLab).

2.3 DEVELOPMENT OF MULTILOCUS SEQUENCE ANALYSIS PROCEDURE

The genes to be used for MLSA were chosen based on literature research by the use of PubMed (http://www.ncbi.nlm.nih.gov/pubmed), resulting in multiple papers which formed the basis of the research. The publications were studied and the used genes, primer sequences, software and tools for sequence and phylogeny analysis compared.

2.3.1 SELECTION OF HOUSEKEEPING GENES AND PRIMER PAIRS

Four housekeeping genes, *recA*, *gyrB*, *pyrH* and *rpoA* and additionally a virulence gene, *toxR*, were selected for this analysis by comparing former publications dealing with similar approaches (27, 41, 49, 58, 65, 78, 79, 81, 84) on criteria like observed serogroup, application frequency, results and recommendations. The respective primer pairs were also selected by their application frequency and the resulting amplicon length, which should be less than 1000 bp because of restrictions given by custom sequencing services. Selected primer pairs were ordered at Thermo Scientific and prepared according to their recommendations. The degenerate primer-sequences were listed in **Table 2**. Additionally forward and reverse primer-sequence, start position, length of amplicon, reference and annealing temperature for PCR are summarized. (Length and position based on *Vibrio cholerae* O1 biovar ElTor str. N16961.)

Table 2. Primers used for this multilocus sequence analysis.

Primer	Primer sequence (5'-3')	Position	Fragment Length [nt]	Reference	PCR Annealing Temp. [°C]
gyrB F	GAAGGBGGTATTCAAGC	655	629	Garg, 2003 (27), Kotetishvili, 2003 (49)	50
gyrB R	GAGTCACCCTCCACWATGTA	1283			
recA-01-F	TGARAARCARTTYGGTAAAGG	222	837	Thompson et al., 2011 (79)	50
recA-02-R	TCRCCNTTRTAGCTRTACC	1058			
toxR-F	CCTTCGATCCCCTAAGCAATAC	80	779	Rivera et al., 2001 (65)	55
toxR-R	AGGGTTAGCAACGATGCGTAAG	858			
pyrH-04-F	ATGASNACBAAYCCWAAACC	1	617	Thompson et al., 2011, 2005, 2008 (79, 81, 84)	55
pyrH-02-R	GTRAABGCNGMYARRTCCA	617			
rpoA-01-F	ATGCAGGGTTCTGTDACAG	1	970	Thompson et al., 2011, 2005, 2008 (79, 81, 84)	53
rpoA-03-R	GHGGCCARTTTTCHARRCGC	970			

2.3.2 ESTABLISHMENT OF SPECIFIC PCR CONDITIONS

The optimum PCR conditions, consisting of thermal program, concentration and volumes of the components had to be established. Starting from the recommended PCR reaction mix and the recommended PCR cycles according to the manufacturer´s instructions (Solis BioDyne) and the preexisting PCR reaction mix and respective thermal program for *ompW* amplification (5), different modifications were tested for each of the five chosen *V. cholerae* specific genes. Because no exact melting temperature of the degenerate primer sequence could be given, an approximate melting temperature was determined with PrimerBlast (95) based on the corresponding gene sequence provided by NCBI.

Attending the recommendations from BioDyne the annealing temperatures were set between 2 and 6 °C lower than the calculated primer melting temperatures.

2.3.2.1 Primer-test approaches

The five different primer pairs were tested in single-plex PCR approaches with DNA from a *V. cholerae* O1 strain and a *V. cholerae* O139 strain from the strain collection, using three different concentrations of template DNA (**2.2.4 DNA-Extraction**): supernatant pure, 1:10 and 1:100 diluted with water (Sigma Life Sciences), in the following named: 10^0, 10^{-1} and 10^{-2}. Different PCR compositions, summarized in **Table 3**, were tested to determine the most effective primer conditions.

PCR's were performed with a Biometra T Gradient Thermoblock (Biometra). The thermal program consisted of 15 min at 95 °C (initial denaturation), 30 cycles of 1 min at 95 °C for (denaturation), 1 min at primer specific annealing temperature (**Table 3**), 2 min at 72 °C (elongation) and 10 min at 72 °C (final elongation). A negative control, with water (Sigma Life Sciences) was additionally carried along in each PCR. Correct amplification was checked via gel electrophoresis as described in chapter **2.2.5.1**, resulting in similar pictures like **Figure 5,** with more or less visible bands depending on the efficiency of the chosen conditions.

Table 3. Different PCR test approaches and corresponding annealing temperatures.
Given are final concentrations per PCR mix with a volume of 25 µl.

	gyrB	***recA***						***toxR***	***pyrH***							***rpoA***
Annealing temperature	**50 °C**	**50 °C / 53 °C / 55 °C**						**55 °C**	**53 °C**	**55 °C**						**53°C**
Forward primer Thermo Fisher Scientific GmbH **[pmol/µl]**	0.48	0.48	1.2	2.0	3.0	4.0	4.8	0.48	0.48	0.48	1.2	2.0	3.0	4.0	4.8	0.48
Reverse primer Thermo Fisher Scientific GmbH **[pmol/µl]**	0.48	0.48	1.2	2.0	3.0	4.0	4.8	0.48	0.48	0.48	1.2	2.0	3.0	4.0	4.8	0.48
Reaction buffer BD (10×) Solis BioDyne	1×															
$MgCl_2$ Promega	2000 pmol/µl															
dNTPs Solis BioDyne	400 pmol/µl															
hot FIREPol polymerase Solis BioDyne	0.08 U/µl															
Water Sigma Life Sciences	Up to a final volume of 25 µl															

Figure 5, a gel image, visualizes the efficiency of different approaches for *recA* amplification.

On top, a *recA*-PCR with an annealing temperature of 50 °C, in the middle one with 53 °C and beneath one performed with 55 °C is shown. The clearly visible dark bands illustrate that amplification was more efficient at an annealing temperature of 50 °C than at 53 °C or 55 °C. The primer concentration-series ascertained that 2.0 pmol/µl gives sufficient amplification to obtain a strong signal without wasting too much material.

According to this test, for amplification of *recA*, the adjustable PCR conditions - annealing temperature and final primer concentration - were set at 50 °C and 2 pmol/µl.

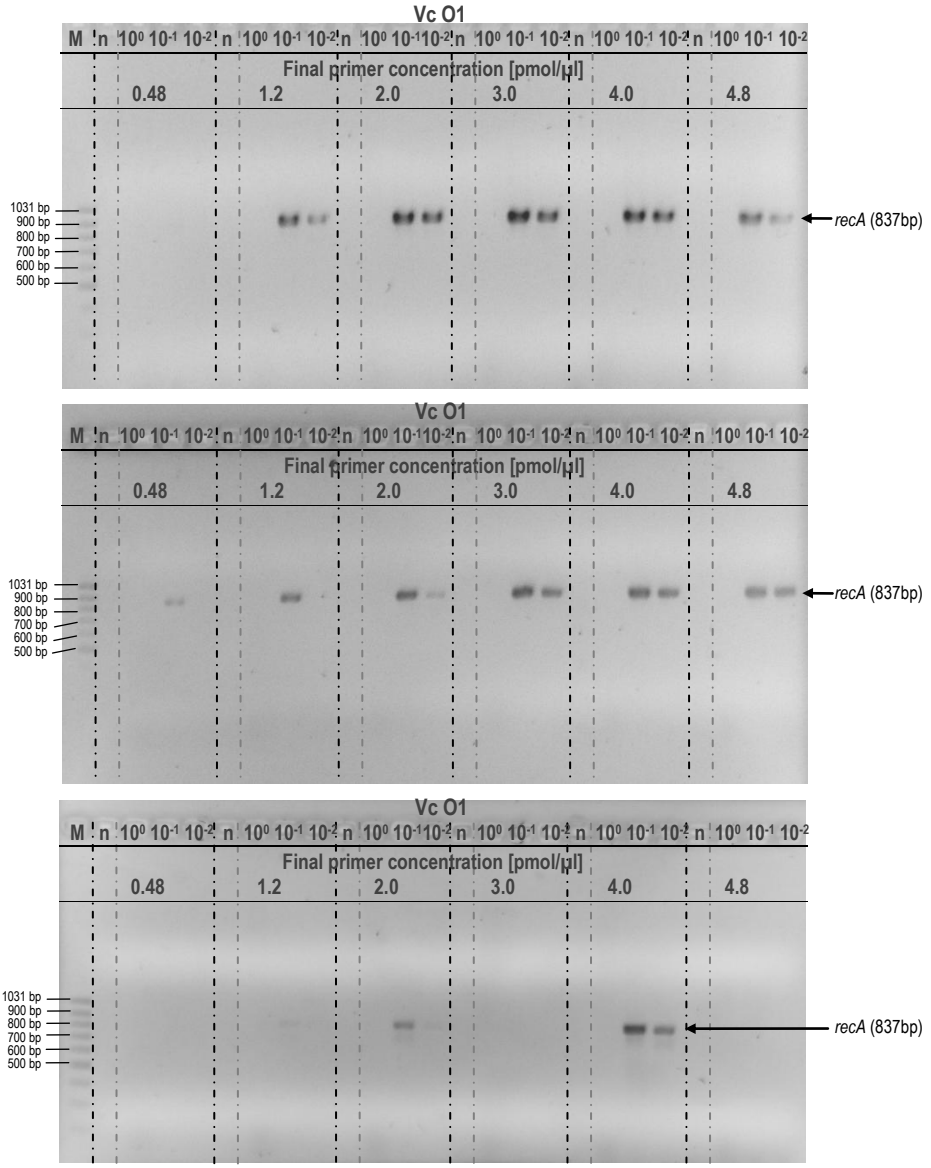


Figure 5. Primer test for primer pair *recA*

Template *V. cholerae* O1 (Vc O1) with different DNA extract dilutions, pure (10^0), 1:10 (10^{-1}), 1:100 (10^{-2}), performed with different final primer concentrations at different annealing temperatures, top 50 °C middle 53 °C and bottom 55 °C. PCR conditions according to **Table 3**. Negative control with water (n). MassRuler Low Range DNA Ladder (M). 2 % agarose gel stained with EtBr.

For each primer pair the PCR compositions and annealing temperatures were examined by performing individual approaches resulting in final compositions to be used for further gene amplification of the *V. cholerae* isolates from Lake Neusiedler See for MLSA.
The final established PCR conditions and annealing temperatures for each gene are listed in **Table 4**. The listed concentrations are the final concentrations in one PCR approach with a volume of 25 µl.

Table 4. Final PCR conditions; concentrations given per reaction volume of 25 µl.

Reagent	*gyrB*	*recA*	*toxR*	*pyrH*	*rpoA*
Annealing temperature	50 °C	50 °C	55 °C	55 °C	53 °C
Forward primer Thermo Fisher Scientific GmbH	0.48 pmol/µl	2 pmol/µl	0.48 pmol/µl	3 pmol/µl	0.48 pmol/µl
Reverse primer Thermo Fisher Scientific GmbH	0.48 pmol/µl	2 pmol/µl	0.48 pmol/µl	3 pmol/µl	0.48 pmol/µl
Reaction buffer BD (10×) Solis BioDyne	1×				
$MgCl_2$ Promega	2000 pmol/µl				
dNTPs Solis BioDyne	400 pmol/µl				
hot FIREPol polymerase Solis BioDyne	0.08 U/µl				
Template DNA	5 µl				
Water Sigma Life Sciences	Up to a final volume of 25 µl				

2.4 SEQUENCE AMPLIFICATION

For MLSA the five genes, *gyrB, recA, toxR*, *pyrH* and *rpoA* of each isolate were amplified via PCR under the established conditions (**Table 4**). As the 1:10 diluted DNA extract affords throughout positive amplification results (**Figure 14**) the 1:10 diluted DNA extract was chosen as template for the PCRs (according to **2.2.4**).

2.4.1 SEQUENCE AMPLIFICATION OF ENVIRONMENTAL AND CLINICAL ISOLATES VIA PCR

PCR mixtures per sample composed of 12.6 µl water (DEPC treated, sterile filtered, Sigma Life Sciences), 2.5 µl reaction buffer B [10x] (Solis BioDyne), 2 µl magnesium chloride [0.025 mol/l] (Promega), forward and reverse primer [concentration according to **Table 4**] (Thermo Fisher Scientific GmbH), 0.5 µl desoxynucleoside triphosphates [20mM each] (Solis BioDyne), 0.4 µl HOT FIREPol® DNA polymerase [5 U/µl] (Solis BioDyne) and 5.0 µl template DNA, to a final volume of 25 µl per reaction.

The PCR's were usually done at once for 46 samples, one positive control, with *V. cholerae* O139 as template, and one negative control with water (Sigma), in sum 48 samples, prepared with a mastermix for 52 samples. **Table 5** gives an example for volume calculation: e.g. *recA*-amplification.

The mastermix was prepared in a 1.5 ml reaction tube (Eppendorf), vortexed with a lab dancer (IKA) and portioned with a multipette plus (Eppendorf) to 20 µl in each PCR tube (PCR SingleCap 8er-SoftStrips 0.2 ml, Biozym Scientific GmbH, Germany). Then 5 µl of the DNA extract of each isolate (diluted 1:10) were added to one PCR tube (5 µl water for negative control, or 5 µl DNA extract of *V. cholerae* O139 for positive control) to a final volume of 25 µl. The mix was then finally vortexed, span down with a microcentrifuge (Qualitron) and amplification started in a Biometra T Gradient Thermoblock (Biometra). The thermal program was set as 15 min at 95 °C, 30 cycles of 1 min at 95 °C, 1 min at primer specific annealing temperature (**Table 4**), 2 min at 72 °C and 10 min at 72 °C. Successful amplification was checked via gel electrophoresis (**2.2.5.1**).

Table 5. *recA*-PCR calculator for 46 samples (+ 2 controls → Mastermix 52)

Components	Stock solution		Working solution		Final conc.	Per sample	Mastermix
	[mmol/l]	[pmol/µl]	[mmol/l]	[pmol/µl]	[pmol/µl]	[µl]	[µl]
Water Sigma Life Science						12.6	**655.2**
10× Reactionbuffer BD Solis BioDyne	**10**	10000	**10**	10000	1000	2.5	**130.0**
$MgCl_2$ Promega	**25**	25000	25	25000	2000	2.0	**104.0**
Primer recA F Thermo Fisher Scientific GmbH	1×10^{-7}	**100**	1×10^{-8}	**50**	2	1.0	**52.0**
Primer recA R Thermo Fisher Scientific GmbH	1×10^{-7}	**100**	1×10^{-8}	**50**	2	1.0	**52.0**
dNTP Mix Solis BioDyne	**20**	20000	20	20000	400	0.5	**26.0**
hotFirePol DNA polymerase Solis BioDyne	**5 U/µl**		5 U/µl		0,08 U/µl	0.4	**20.8**
Volume						**20.0**	**1040.0**
Template *V.c.-Isolate*						**5.0**	

After visualization of correct amplification the single PCR's were stored at -20 °C till usage for sequencing.

2.4.1.1 Concentration determination

The concentration of the PCR amplificates was determined approximately via comparison with a marker (MassRuler™ Low Range DNA Ladder).

For concentration determination three *recA*-amplification products of the isolates 40, 50 and 70 were diluted 1:2, 1:5 and 1:10 with water (Sigma) and 5 µl of each loaded onto a 2 % agarose gel. 5 µl or 3 µl marker were loaded in between the samples for comparison. The topmost band of the MassRuler™, 1031 bp, has a concentration, according to manufacturer´s instructions (**Figure 6-A)** of 50ng/5µl, and therefore 30ng/3µl.

By closer examination of **Figure 6-B** the visualized band of the 1:10 dilution (of all three samples) is comparable with the topmost band of the marker with 3 µl, which means that the 1:10 dilutions contains approximately 30 ng PCR product in 5 µl. The pure PCR product contains therefore 300ng/5µl, which is an approximate concentration of about 60ng/µl at one PCR reaction.

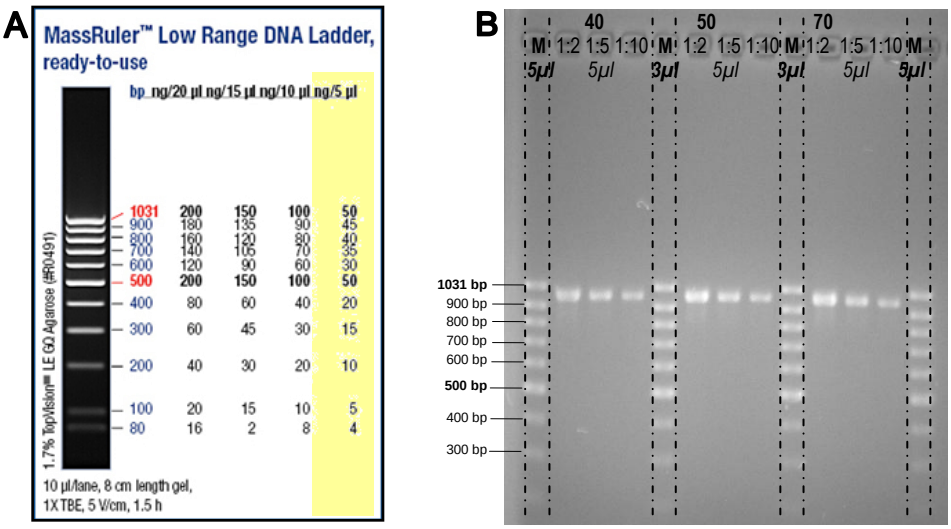


Figure 6. Concentration determination via Marker.
A: MassRuler™ Low Range DNA ladder (Fermentas). B: PCR-amplificates *(recA)* for comparison with marker on a 2 % agarose gel, stained with EtBr.

2.4.2 DNA SEQUENCING

The sequencing facility Microsynth (Balgach, Switzerland) was chosen for sequencing. The whole-plate sequencing of PCR products by Barcode Economy Run Service includes PCR purification and is applicable for read length up to 1,100 bases. They require the PCR products in liquid form at room temperature in pure water with correct sample amount/concentrations.

2.4.2.1 Default sample amounts and concentrations

The sequencing facility requires for unpurified PCR products an amount of 20 µl with a concentration of 3 ng/µl (per 100 bases) and for separate sent Primer aliquots of 200 µl [10 pmol/µl]. To calculate the amount of sample and water to achieve 20 µl with the required concentration of 3 ng/µl per 100 bases, the estimated concentration of 60 ng/µl and the individual length of the PCR product was included. **Table 6** lists the results.

Table 6. Amount & concentration calculations for sequencing service.

Sample		Concentration	Volume	
Required for unpurified PCR product		**3 ng/100 bases**	**20 µl**	
MLSA PCR product		**60 ng/µl**	**20 µl**	
amplificate	**bases**	**required [ng]**	**PCR product [µl]**	**water [µl]**
recA	837	25	8.4	11.6
gyrB	629	19	6.3	13.7
pyrH	617	19	6.2	13.8
rpoA	970	29	8.7	10.3
toxR	779	23	7.8	12.2

Both direction sequencing, forward and reverse, was performed for all isolates of the intermediate habitat (1-100), for half of the isolates of sampling site reed belt (201-248) and for the eight clinical isolates. As the resulting belonging forward and reverse sequences of these samples were in nearly all cases completely conform, the remaining PCR products were sequenced in one direction.

2.4.2.2 Plate preparation

For example the *recA* PCR products of isolates 1-48 were prepared for forward and reverse sequencing. Therefore 11.6 µl water (Sigma) plus 8.4 µl of the PCR product *recA* (**Table 6**) were pipetted in two of the 96 wells, according to the pipetting scheme (**Figure 7**) generated during the online ordering process where the according primer for each well is noted. This was carried on for all 48 PCR products; then the plate was sealed and sent to the facility. The PCR primers were applicable for the sequencing reactions and sent separately, (in proper concentration, **2.4.2.1**) in 1.5 ml Safe-Lock PCR tubes (Eppendorf) with the plates. The scheme ensures correct sample-primer combination.

	1	2	3	4	5	6	7	8	9	10	11	12
A	1 *recA_F*	9 *recA_F*	17 *recA_F*	25 *recA_F*	33 *recA_F*	41 *recA_F*	1 *recA_R*	9 *recA_R*	17 *recA_R*	25 *recA_R*	33 *recA_R*	41 *recA_R*
B	2 *recA_F*	10 *recA_F*	18 *recA_F*	26 *recA_F*	34 *recA_F*	42 *recA_F*	2 *recA_R*	10 *recA_R*	18 *recA_R*	26 *recA_R*	34 *recA_R*	42 *recA_R*
C	3 *recA_F*	11 *recA_F*	19 *recA_F*	27 *recA_F*	35 *recA_F*	43 *recA_F*	3 *recA_R*	11 *recA_R*	19 *recA_R*	27 *recA_R*	35 *recA_R*	43 *recA_R*
D	4 *recA_F*	12 *recA_F*	20 *recA_F*	28 *recA_F*	36 *recA_F*	44 *recA_F*	4 *recA_R*	12 *recA_R*	20 *recA_R*	28 *recA_R*	36 *recA_R*	44 *recA_R*
E	5 *recA_F*	13 *recA_F*	21 *recA_F*	29 *recA_F*	37 *recA_F*	45 *recA_F*	5 *recA_R*	13 *recA_R*	21 *recA_R*	29 *recA_R*	37 *recA_R*	45 *recA_R*
F	6 *recA_F*	14 *recA_F*	22 *recA_F*	30 *recA_F*	38 *recA_F*	46 *recA_F*	6 *recA_R*	14 *recA_R*	22 *recA_R*	30 *recA_R*	38 *recA_R*	46 *recA_R*
G	7 *recA_F*	15 *recA_F*	23 *recA_F*	31 *recA_F*	39 *recA_F*	47 *recA_F*	7 *recA_R*	15 *recA_R*	23 *recA_R*	31 *recA_R*	39 *recA_R*	47 *recA_R*
H	8 *recA_F*	16 *recA_F*	24 *recA_F*	32 *recA_F*	40 *recA_F*	48 *recA_F*	8 *recA_R*	16 *recA_R*	24 *recA_R*	32 *recA_R*	40 *recA_R*	48 *recA_R*

Figure 7. Pipetting scheme of 96-well plate for sequencing.
Scheme of plate preparation for sequencing *recA* of *V. cholerae* environmental isolates 1-48. Number of isolate and according primer noted within each cell.

2.5 MULTILOCUS SEQUENCE ANALYSIS

2.5.1 SEQUENCE ASSEMBLY

The sequence data is provided as Secure HTTPS download in two file formats .ab1 and .fasta by the sequencing facility. For correct identification the files were named with:

sample name__no.of plate__primer name.file-format

e.g.: 01 3635 recA-F.ab1
01 3635 recA-F.fasta

01 3635 recA-R.ab1
01 3635 recA-F.fasta

Raw .ab1 files were imported to CLC Main Workbench (Version 6.8.1-3, CLC bio). Contigs were assembled from forward and reverse sequences with included trace data of each locus for each isolate, **Figure 8.** Then positions with conflicting data were located and corrected manually. Sequences from one-direction sequencing were equally imported and the belonging trace data examined manually for contingently appearing errors. The resulting consensus sequence was then saved separately, named with no. of isolate and coding gene.

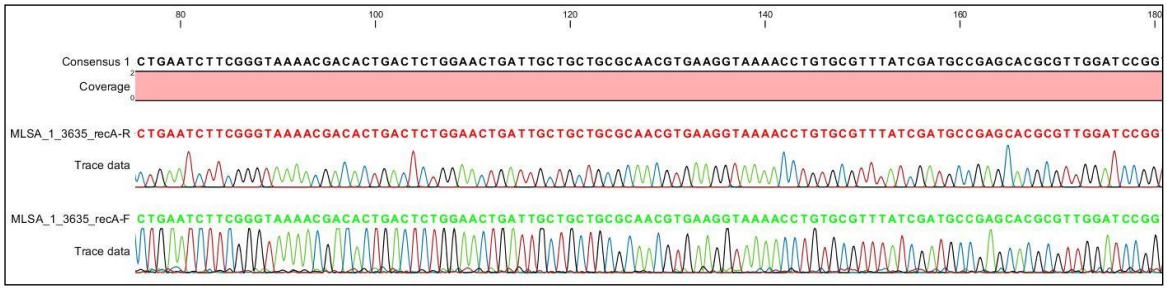


Figure 8. Contig of isolate no. 01, forward and reverse sequence of *recA* assembled.
Position 76 to 180 shown, Screenshot of CLC Main Workbench.

2.5.2 SEQUENCE TRIMMING

The read lengths of the sequences differed from reaction to reaction so the consensus sequences were all different in length. Because afterwards the phylogenetic tree was computed based on the DNA sequence differences all the sequences of one locus had to start and end at equal positions otherwise the differences in lengths would be weighted. Therefore all consensus sequences of the same locus were aligned and the overlapping region determined. Subsequently all sequences were trimmed manually to equal length.

All actions were performed for each locus separate with CLC Main Workbench.
Table 7 summarizes the final lengths of the fragments used for this MLSA approach, compared with the size of the corresponding gene and the size of the PCR product.

Table 7. Comparison of length of the gene, the PCR product and the fragment used for the analysis.

Gene	Gene size [nt]	PCR product [nt]	Fragment for analysis [nt]
gyrB	2,418	629	487
recA	1,065	837	648
toxR	885	779	658
pyrH	732	617	477
rpoA	993	970	(+712)
		Concatenated sequence	**2,270 (2,982)**

2.5.3 SEQUENCE CONCATENATION

The five loci of each environmental/clinical isolate were then concatenated. The order was set according to the distribution on Chr.I of *Vibrio cholerae* O1 biovar ElTor str. N16961. If sequences were missing, due to sequencing failures, they were skipped. The concatenated sequences comprise a minimum of three loci.

Reference strains: The five chosen genes of each reference strain were downloaded separate from NCBI GenBank database and imported into CLC Main workbench. Each locus was trimmed to the previously determined individual start and end positions and the five belonging sequences were then concatenated.

2.5.4 PHYLOGENETIC ANALYSIS

2.5.4.1 Nucleotide properties and details of phylogenetic analysis

The nucleotide properties of the used loci were computed by MrBayes v. 3.0B4 (66) and the details of phylogenetic analysis were computed with the software DnaSP 5.10.01 (67) and the implemented tools “Analysis of Polymorphic Sites” and “Analysis of DNA Polymorphism”. Therefore the alignment file (NEXUS) was used but the four isolates with missing data (no. 46, 151, 207, 280) were excluded. The analysis was performed for each locus, by defining the “Region to Analyze” individually and also for the concatenated dataset.

2.5.4.2 Phylogenetic analysis

The concatenated sequences were exported as NEXUS files and subsequently imported to MEGA, Vers. 5.2 (77) and finally aligned using the muscle algorithm (19, 20), settings **Figure 9**. As alignment (see part of it in **Figure 10**) did not contain any ambiguous areas no special treatments were applied. Alignments were saved for further steps as FASTA files which is the compatible format for different programs.

The steps for this analysis were performed according to Druzhinina et al., 2010, (17).
The possibility of intragenic recombination, which would prohibit the use of the respective loci for phylogenetic analysis, was tested by linkage disequilibrium based statistics as implemented in DnaSP 4.50.3 (67). The neutral evolution of the used fragments (*gyrB, recA, toxR, pyrH* and *rpoA*) was tested by Tajima’s D test implemented in the same software (**Table 10**). The interleaved NEXUS files were formatted using PAUP*4.0b10 (75). The unconstrained GTR + I + G substitution model was applied to all sequence fragments.
Metropolis-coupled Markov chain Monte Carlo (MCMC) sampling was performed using MrBayes v. 3.0B4 with two simultaneous runs of four incrementally heated chains that performed 1 or 10 million generations, depending on the data (**Table 10**). The length of run (number of generations) for each dataset was determined using AWTY graphical system (89) to check the convergence of MCMC. Bayesian posterior probabilities (PP) were obtained from the 50 % majority rule consensus of trees sampled every 100 generations after removing the first trees using the "burnin" command. Number of discarded generations was determined for every run based on visual analysis of the plot showing generation versus the log probability of observing the data. According to the protocol of Leache and Reeder (50) PP values lower than 0.95 were not considered significant while values below 0.9 are not shown on the resulting phylograms.

Maximum-parsimony method of phylogenetic reconstructions (observed p-distance, no evolutionary modeling required) was done using PAUP*4.0b10 applying a heuristic search (n = 1,000) with the random addition of sequences and the TBR tree-swapping algorithm.

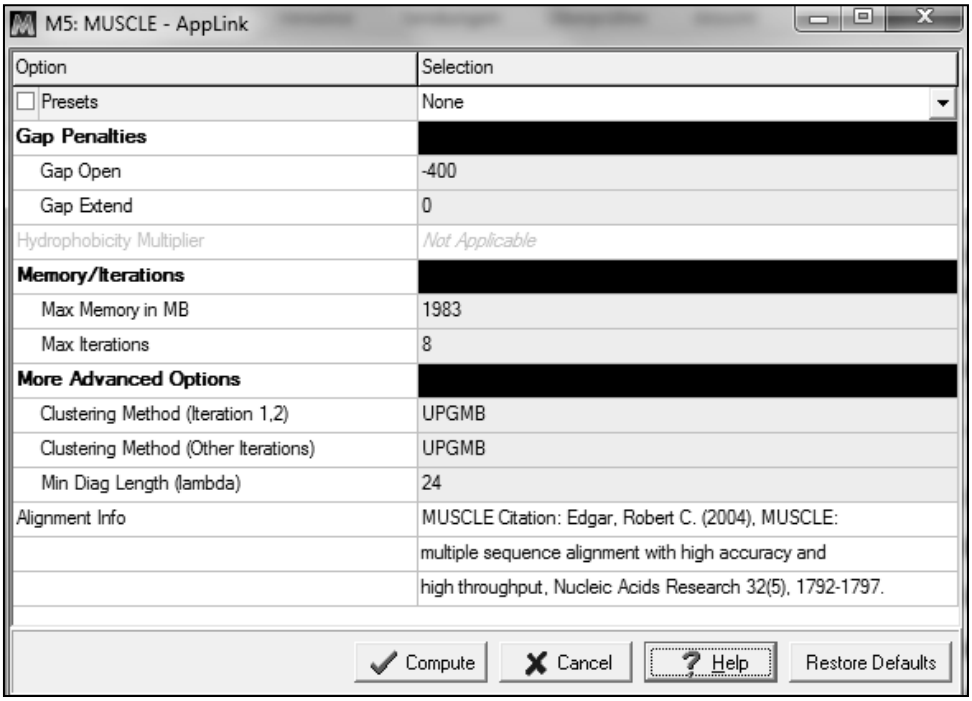


Figure 9. Default settings used for alignment with the muscle algorithm.
Snapshot MEGA 5.2.2, Win7.

Figure 10. Part of Muscle Alignment, site 1326 to 1492- locus *toxR*.
First column lists the isolate number followed by corresponding DNA-Sequence. Each nucleotide marked with a color: guanine in purple, adenine in green, cytosine in blue and thymine in red. Positions 1,411-1,413 is a three nucleotide spanning indel. Snapshot Alignment Explorer, MEGA 5.2, Win7.

The *Vibrio cholerae* isolates were assigned to phylotypes and haplotypes based on their arrangement on the phylograms. These phylotypes generated by single trees were then numbered for convenience consecutively from 1-98 and the haplotypes generated by the concatenated sequences were numbered from 99-146; in the following named as haplotype ID, listed in **Table 11**.

2.5.5 COMPARISON OF *VIBRIO CHOLERAE* DIVERSITY BETWEEN DIFFERENT HABITATS

Another way to describe the relation between the *Vibrio cholerae* isolates of the ecosystems is the Venn diagram. Therefore the tool VENNY was used (63). This Venn diagram is based on the parsimony informative sites of each sampling site, as created with the software DnaSP 5.10.01 (67) and the implemented tool “Analysis of Polymorphic Sites”. The alignment file (NEXUS) is the source.
“A site is parsimony-informative if it contains at least two types of nucleotides, and at least two of them occur with a minimum frequency of two” (76, 77).

The analysis was performed for each sampling site and also for the references/clinical isolates. First four “Sequence Sets” (references/clinical; intermediate habitat; open water; reed belt) were defined, comprising of the isolates belonging to the chosen sampling site (46, 151, 207, 280 were again excluded). The sequences of the references and clinical isolates had to be pooled due to the fact that only four sets can be depicted with this tool. Subsequently the “Analysis of Polymorphic Sites” is selected, the “Data Set” (equal to “Sequence Set”) desired to analyze chosen and computed for the whole concatenated sequence: “Region to Analyze” 1-2,270. The Parsimony Informative Sites were then listed one beneath the other to be ready for import in the tool.
To estimate the species richness of *Vibrio cholerae* in the lake rarefaction curves were generated. Therefore matrices for each of the three sampling sites were created individually based on the concatenated Bayesian tree, and imported into EstimateS (11).

Table 8 illustrates the matrix generated for the intermediate habitat. It comprises of number of clades (rows) present for the selected loci and divisions of 10 with the following scheme: isolate numbers from 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100 (columns). Then the number of isolates with names between those divisions counted, noted in a table, and saved as TXT file.

Table 8. Matrix to estimate species richness, format compatible for import into EstimateS.

Intermediate habitat									
18	10								
0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	0	0
0	1	2	0	0	0	0	0	0	1
0	1	1	2	0	3	1	0	1	0
1	0	1	0	1	2	0	0	1	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	0	1
1	1	0	0	1	0	0	1	0	0
0	0	0	1	0	0	2	0	0	1
0	1	0	0	0	0	1	0	0	0
2	0	0	0	0	1	0	1	0	1
0	0	0	0	0	0	0	1	0	0
4	5	5	6	5	4	4	6	7	5
0	0	0	1	0	0	0	0	0	0
0	1	0	0	1	1	0	0	0	0

18 clades (rows) exist where isolates of this habitat were present. In the first clade is one isolate with a number between 61-70 and one isolate with a number between 81-90. Continued for all 18 clades.

3 RESULTS

3.1 SAMPLING SITES

Different environmental parameters were measured at each sampling site. The differences between the sampling sites are summarized in **Table 9** and graphically depicted in **Figure 11** and **Figure 12**. The whole dataset is listed in **Table S 1**.

Table 9. Lake parameters, median, min. and max. values in brackets, n=8 (05.06.12-10.09.12).

Sample site	Water temp. [°C]	CFU *V.c.* [CFU/ml]	SPC *V.c.* [cells/ml]	Total bacterial cell nr. [cells/ml]	pH	Conductivity µS/cm
Intermediate habitat	23.6 (19.4 - 27.0)	46.0 (6.1 - 90.0)	40.3 (5.5 - 86.7)	9.29×10^6 (6.10×10^6- 1.20×10^7)	8.9 (8.8 - 8.9)	2200 (1800 - 2600)
Open water	22.7 (18.1 - 27.5)	21.9 (0.5 - 54.0)	35.8 (2.8 - 54.0)	8.59×10^6 (5.94×10^6- 1.58×10^7)	8.9 (8.8 - 8.9)	2000 (1700 - 2500)
Reed belt	23.6 (17.6 - 28.3)	52.5 (27.3 - 84.0)	164.6 (57.5 - 241.6)	1.31×10^7 (8.42×10^6- 2.08×10^7)	8.7 (8.6 - 8.8)	2450 (1900 - 2700)

Sample site	P_{tot} [µg/l]	Chl-a [µg/l]	DOC [mg/l]	Humics Ratio	TSS [mg/l]	Secchi [cm]
Intermediate habitat	31.0 (16 - 132)	6.1 (4.7 - 20.8)	17.6 (15.8 - 19.2)	15.0 (6.5 - 21.7)	40.0 (30.0 - 105.3)	23.0 (10.0 - 38.0)
Open water	52.5 (22 - 122)	11.5 (6.0 - 29.5)	14.6 (13.1 - 15.5)	15.9 (8.3 - 19.5)	50.2 (27.3 - 326.9)	16.5 (3.0 - 42.0)
Reed belt	45.0 (35 - 56)	12.9 (7.1 - 17.5)	26.3 (23.1 - 35.8)	9.5 (5.8 - 11.0)	21.6 (14.9 - 35.3)	47.5 (21.0 - 80.0)

Water samples were collected biweekly during summer; the sampling for this MLSA approach was in between two standard samplings of the survey, mentioned in the introduction, and is indicated in each diagram by a grey shaded area.

The water temperature behaved similar at the three sampling sites during the summer, the median temperature was about 23.6 °C, with a minimum temperature at 17.6 °C and a maximum at 28.3 °C measured in the reed belt and similar values at the other two sampling sites. *Vibrio cholerae* colony forming units and *V. cholerae* cells/ml, measured with Solid Phase Cytometry, show different values because of the higher sensitivity of the solid phase cytometry approach. At the intermediate habitat 46 and 40 *Vibrio cholerae* cells were detected in median per ml, in the open water 21 and 35 ml^{-1}, and in

the reed belt 52 and 164 ml^{-1} *V. cholerae* were counted with cultivation and SPC, respectively. Behind this median values there was a high range of concentrations, ranging from 6 to 90 CFU ml^{-1} at the intermediate habitat, from 0.5 to 54 CFU ml^{-1} at the open water area and from 27-84 CFU ml^{-1} in the reed belt. A similar variability for cells detected with Solid Phase Cytometry was observed. Compared with the total cell number, ranging from a median of 8.6×10^6 cells per ml in the open water, to 9.3×10^6 cells per ml in the intermediate habitat, and up to 1.3×10^7 cells per ml in the reed belt, the number of *Vibrio cholerae* individuals is very low.

The pH values of the sampling sites intermediate habitat and open water were similar with a median value of 8.9. In the reed belt pH was slightly lower with a median value of 8.7 and with less fluctuations. Conductivity was very similar at all three stations, ranging from 2000 µS/cm in the open water to 2450 µS/cm in the reed belt. The median total phosphorus values ranged from 31 µg/l in the intermediate habitat, to 45 µg/l in the reed belt and to 52.5 µg/l in the open water with big fluctuations between minimum and maximum values. Lowest chlorophyll a values were measured at the intermediate habitat with about 6 µg/l, ranging from 4.7 µg/l to 20.8 µg/l Chl-a during summer. At the open water and the reed belt the median Chl-a values were similar with 11.5 µg/l and 12.9 µg/l, respectively, ranging from 6.0 µg/l to 29.5 µg/l and 7.1 µg/l to 17.5 µg/l.

Dissolved organic carbon concentrations were lowest in the open water with a median value of 14.6 mg/l and little fluctuation. Slightly more DOC was present in the intermediate habitat with 17.6 mg/l. In the reed belt between 23.1 mg/l and 35.8 mg/l with a median of 26.3 mg/l DOC was observed. The reed belt had the lowest humics ratio of 9.5 with little fluctuation during the whole sampling period. Intermediate and open water showed higher humic ratio's of 15.0 and 15.9, with a minimum and maximum of 6.5 and 21.7 and 8.3 and 19.5, respectively. A low humic ratio indicates a high amount of humic substances among the dissolved organic material in the water. Total Suspended Solids median values ranged from 21.6 mg/l in the reed belt to 40 mg/l at the intermediate habitat up to 50.2 mg/l in the open water. Whereas the reed belt values varied slightly between 14.9 mg/l to 35.3 mg/l a very high value of 326.9 mg/l was detected in the open water in the middle of July, due to the influence of strong winds.

Secchi depths of the intermediate habitat and the open water differed little from each other but fluctuated from 3 cm to 47.5 cm during the sample period. The Seechi depth measured at the reed belt rose till the middle of summer from 30 cm up to 80 cm, however at the last sampling date it was again at a low level like in the open water.

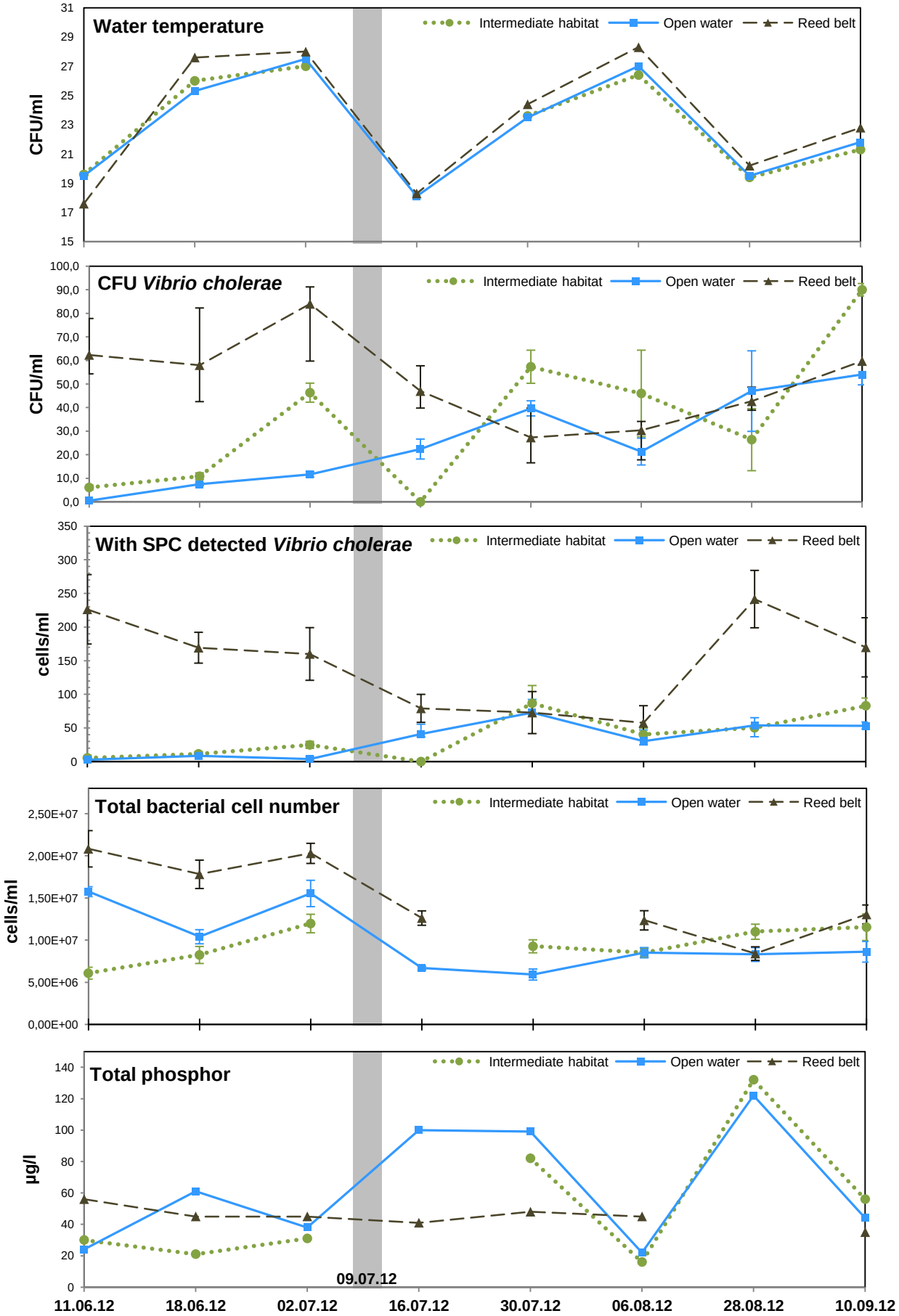


Figure 11. Development of certain water parameters at the three habitats in the period from 11.06.12 to 10.09.12. Green dotted line: intermediate habitat; blue line: open water; black dashed line: reed belt. Vertical bars correspond to standard deviation. Grey shaded area indicates time of sampling for MLSA. Date of sampling plotted on the x-axis, y-axis indicates parameter specific unit; data according to **Table S 1**.

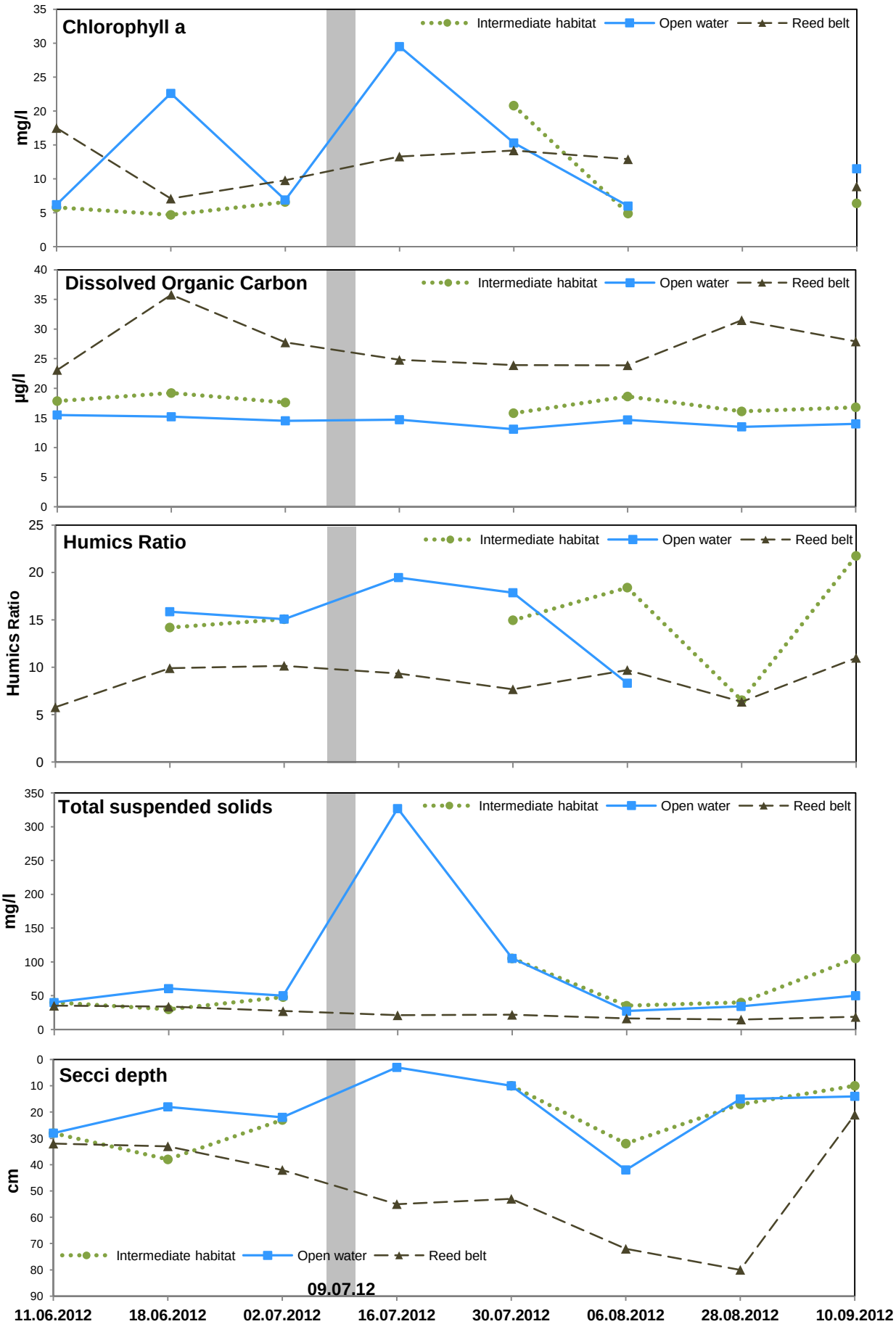


Figure 12. Development of certain water parameters at the three habitats in the period from 11.06.12 to 10.09.12. Green dotted line: intermediate habitat; blue line: open water; black dashed line: reed belt. Grey shaded area indicates time of sampling for this MLSA. Date of sampling plotted on the x-axis, y-axis indicates parameter specific unit; data according to **Table S 1**.

3.2 ISOLATION AND IDENTIFICATION OF PRESUMPTIVE *VIBRIO CHOLERAE*

3.2.1 COLLECTION AND PREPARATION OF ENVIRONMENTAL ISOLATES

100 presumptive *V. cholerae* colonies of each of the three sampling sites were collected by cultivation, resulting in 300 isolates, each one stored in 500 µl PBS at -20 °C and additionally in a cryo-vial in the sample bank at -80 °C. The isolates were numbered from 1 to 300, see **Table 1.**

3.2.2 SCREENING FOR *VIBRIO CHOLERAE* BY *OMPW* PCR

Out of 300 July isolates analysed by *ompW* PCR, 299 were detected as *ompW* positive. Two isolates from open water, no. 145 and no. 191, showed up as *ompW* negative, therefore PCR was repeated twice, which resulted in two positive signals for no. 145 and two another negative for no. 191. As isolate no. 191 was certainly *ompW* negative, so not a *Vibrio cholerae* isolate, it was excluded from further experiments.

Figure 13 shows *ompW* PCRs for isolates 178 to 200, detected on a 2 % agarose gel stained with EtBr. All isolates except of no. 191 showed an *ompW* positive signal at the length of 300 bp.

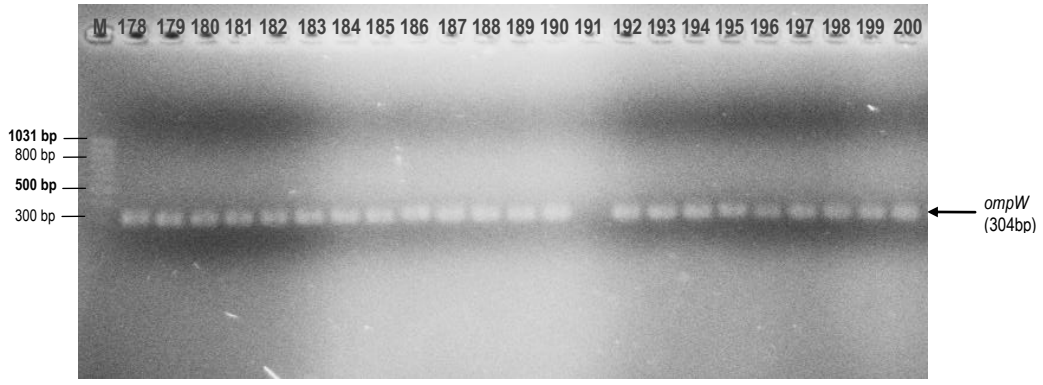


Figure 13. *ompW*-PCR for isolates 178 to 200, originating from open water.
PCR conditions according to **2.2.5.** MassRuler Low Range DNA Ladder (M). 2 % agarose gel stained with EtBr.

3.3 DEVELOPMENT OF MULTILOCUS SEQUENCE ANALYSIS PROCEDURE

3.3.1 ESTABLISHMENT OF SPECIFIC PCR CONDITIONS

3.3.1.1 Primer-test approaches

Final DNA amplification of the five genes and additionally *ompW* of *V. cholerae* O1 and *V. cholerae* O139 strains with three different DNA extract dilutions for comparison, were summarized on a 2 % agarose gel, as shown in **Figure 14**. (PCR conditions according to **Table 4**.)

By comparing the amount of the amplified fragments, the DNA extract dilution 1:10 afforded the best results and was used as template for further PCR's. As strain *V. cholerae* O139 gave a better amplification result than *V. cholerae* O1 for the gene *pyrH* it served as positive control in all further sequence amplifications for all PCR's.

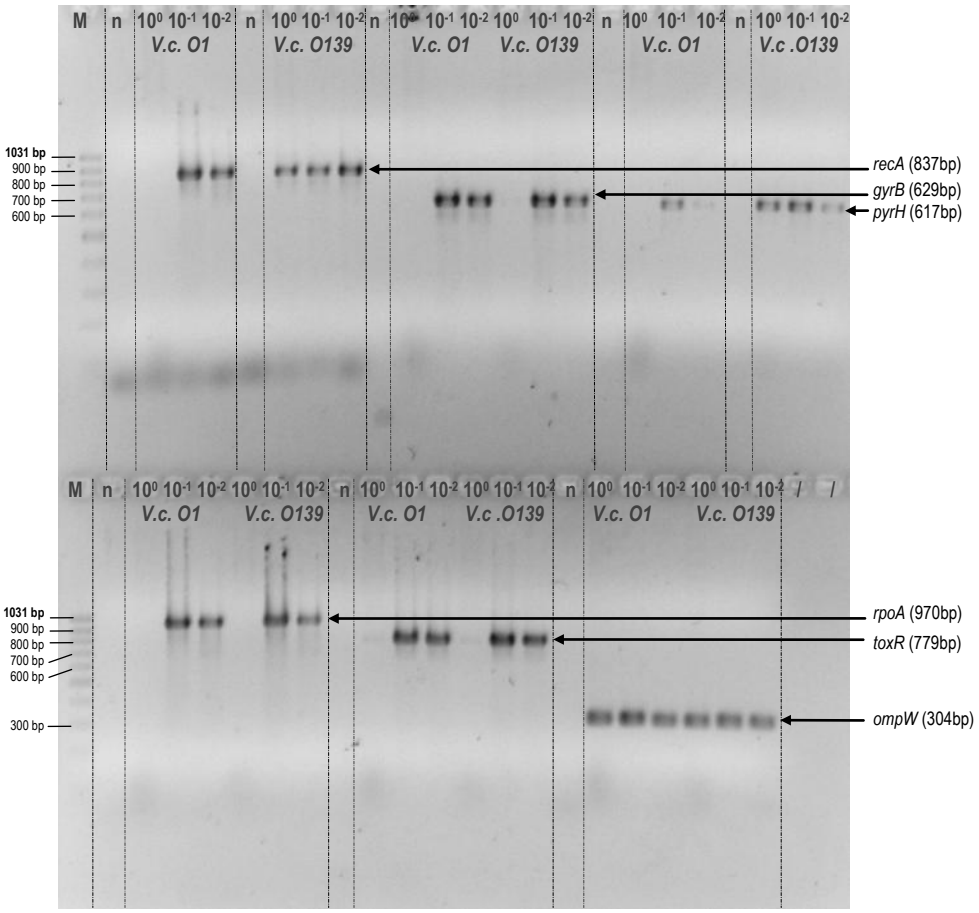


Figure 14. Result of the primer-test approaches.
As DNA templates serve *V. cholerae* O1 and *V. cholerae* O139 with different DNA extract dilutions: pure (100), 1:10 (10-1), 1:100 (10-2). Negative control with water (n). PCR conditions according to **Table 4.** MassRuler Low Range DNA Ladder (M). 2 % agarose gel stained with EtBr.

3.4 SEQUENCE AMPLIFICATION

3.4.1 SEQUENCE AMPLIFICATION OF ENVIRONMENTAL AND CLINICAL ISOLATES VIA PCR

Figure 15 shows the successful sequence amplification of the gene *recA* (837bp) for the isolates 1 to 46, originating from the intermediate habitat, visualized on an agarose gel stained with EtBr. PCR's prepared according to **2.4.1**.

This picture is an example out of many similar pictures; it shows only 46 out of 1,500 successful single PCR's, arising from the amplification of five genes from 300 different isolates.

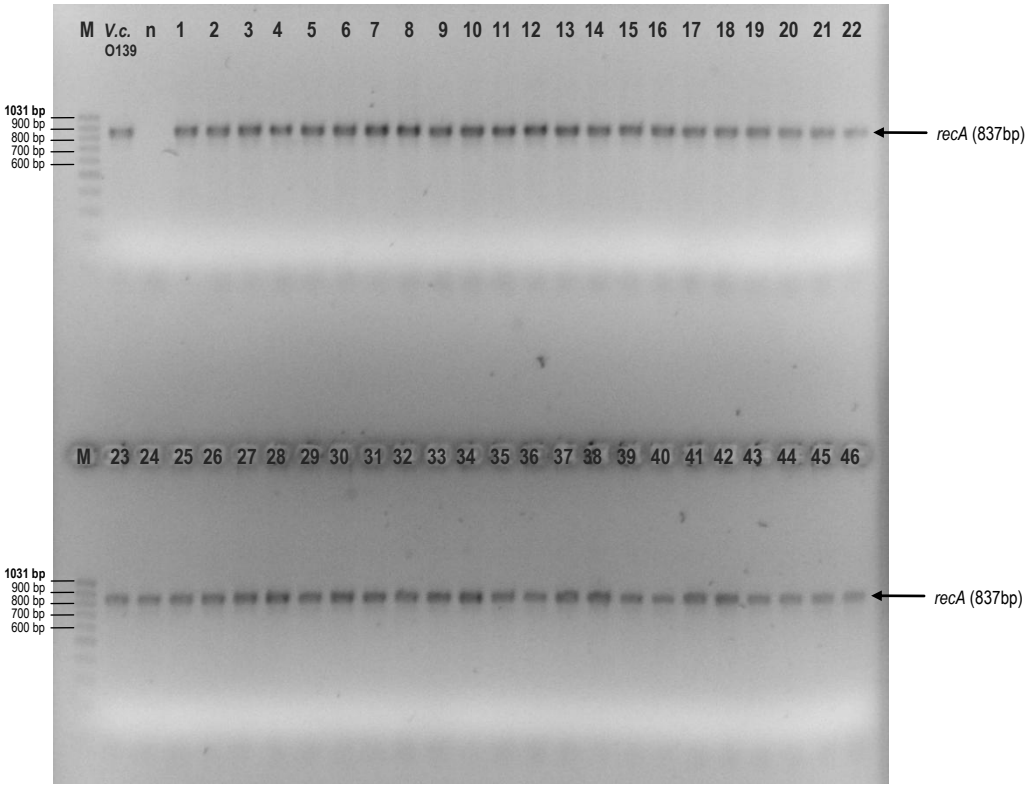


Figure 15. *recA*-amplification of isolates 1-46.
Vibrio cholerae O139 as positive control; negative control with water (n). PCR conditions according to **Table 4.** MassRuler Low Range DNA Ladder (M). Number of Isolates (1-46). 2 % agarose gel stained with EtBr.

3.5 MULTILOCUS SEQUENCE ANALYSIS

After analysis of 148 isolates the preliminary results showed that *rpoA* sequences contained only 8 polymorphic sites (total number of mutations: 8) out of 712 sites, and only 5 parsimony informative sites. A phylogenetic tree based on this locus (data not shown) illustrated clearly that *rpoA* failed to differentiate the *Vibrio cholerae* strains of Lake Neusiedler See and was therefore excluded from further sequencing reactions. Additionally five more single sequences could not be obtained completely and where therefore also excluded from concatenation. This means that four concatenated sequences (isolates 46, 151, 207, 280) comprise only three genes.

For the intermediate habitat, 495 sequences (four *rpoA* and one *recA* sequence [isolate 46] excluded) were obtained, 446 resulting from forward and reverse sequencing, 49 from one direction sequencing. From this 100 concatenated sequences were generated, where 95 comprised the entire set of five genes and five were composed of four genes.

Open water yielded 395 sequences (minus isolate 191 and all *rpoA sequences*), whereof 377 were obtained from one direction and 18 from both direction sequencing, resulting in 99 concatenated sequences, 98 comprising of 4 genes and 1 of 3 genes [isolate 151].

From the reed belt, 445 sequences were acquired (*rpoA* was sequenced for one half of the sample set), 256 from both and 189 from one direction sequencing, which resulted in 100 concatenated sequences whereof 47 comprised 5 genes, 51 comprised 4 genes and two comprised 3 genes [isolates 207, 280].

In sum 1,335 sequences from environmental samples were acquired, resulting in 299 concatenated sequences which were used for diversity calculations.
Additionally eight concatenated sequences resulting from eight clinical *V. cholerae* isolates were included.

3.5.1 PHYLOGENETIC ANALYSIS

The analysis was based on three housekeeping genes (*gyrB*, *recA*, *pyrH*) and one virulence-associated gene (*toxR*).

3.5.1.1 Nucleotide properties and details of phylogenetic analysis

The detailed nucleotide properties of all four loci used and of the phylogenetic analysis made for all of them individually and for the concatenated data set, are provided in **Table 10.**

The fragments of the three housekeeping genes *gyrB*, *recA* and *pyrH a*nd the virulence associated factor *toxR* used for this analysis have a comparable size, all between 477 and 658 nucleotides. The numbers of parsimony informative sites of the three housekeeping genes were between 36 and 56. The concatenated sequence had 269 parsimony informative sites out of 2,270 total nucleotides, whereof almost 50 % were based on the *toxR* fragment, which contributed 126 parsimony informative sites. However this higher number was not associated with length. Hence the number of constant sites (monomorphic sites) for the housekeeping genes was around 90 %, for *toxR* only 80 % of all 658 nucleotides were constant sites.

The parameters of the MCMC analysis included relatively balanced mean nucleotide values and substitution rates. As expected the transitions were generated at a higher frequency than transversions. For example, at locus *gyrB* the substitution rates for A-G and C-T were 0.29 and 0.40, respectively, in contrast to the substitution rates for A-C, A-T, C-G, G-T that were 0.05, 0.08, 0.08 and 0.10, respectively.

Indels were only found in the *toxR* fragment, with a size of three nucleotides. The number of segregation sites (or variable sites, or polymorphic sites) was comparable for the housekeeping genes with values between 40 to 60 but for *toxR* it is more than twice as high with a number of 133. This, of course, affects the number of segregation sites of the concatenated dataset. The number of phylotypes was almost equal for *gyrB*, *toxR* and *pyrH* with 28, 29 and 30, only *recA* had an increased value of 39. The concatenated dataset resulted in only 59 haplotypes.

The phylotype diversity values of the four loci were with approximately at 0.823 close to each other, while for the concatenated sequences the haplotype diversity value was 0.851. The nucleotide diversity was low between 0.008 and 0.03 whereof the *pyrH* sequence fragment gave lowest value.

Table 10. Nucleotide properties of used loci and details of phylogenetic analyses.

Parameters	phylogenetic marker				
	gyrB	*recA*	*toxR*	*pyrH*	concatenated dataset
Fragment characterization	DNA gyrase subunit B	recombinase A	cholera toxin transcriptional activator	uridylate kinase	not applicable
Number of sequences	307	307	307	307	307
Number of characters: total/pars. inform./constant	487 / 51/ 430	648 / 56 / 588	658 / 126 / 522	477 / 36 / 437	2270 / 269 / 1977
	Parameters of MCMC analysis				
Mean nt frequencies A / C / G / T	0.27 / 0.22 / 0.26 / 0.25	0.26 / 0.22 / 0.27 / 0.25	0.31/ 0.24 / 0.21/ 0.24	0.21/ 0.20 / 0.30 / 0.29	not applicable
Substitution rates A-C/A-G/A-T/C-G/C-T/G-T	0.05 / 0.29 / 0.08 / 0.08 / 0.40 / 0.10	0.05 / 0.39 / 0.02 / 0.14 / 0.30 / 0.10	0.10 / 0.37 / 0.06 / 0.07 / 0.29 / 0.11	0.02 / 0.06 / 0.05 / 0.02 / 0.82 / 0.03	not applicable
Number of generations / discarded first generations	1000 000 / 700	1000 000 / 1000	1000 000 / 500	1000 000 / 500	10 000 000 / 700
Total tree length	63.23	62.25	62.01	60.45	62.23
	DNA polymorphism analysis				
Number of sites excluding gaps and missing data	487	648	655	477	2267
Segregation sites / Number of phylotypes or haplotypes	57 / 28	60 / 39	133 / 29	40 / 30	290 / 59
Haplotype diversity / Nucleotide diversity, Pi	0.816 / 0.02077	0.828 / 0.01336	0.829 / 0.03083	0.820 / 0.00792	0.851/ 0.01886
	Neutrality analysis				
Tajima's D test	n.s. (0.34307, P>0.1)	n.s. (-0.26336, P>0.1)	n.s. (-0.13020, P>0.1)	n.s. (-1.13887, P>0.1)	not applicable

3.5.1.2 Phylogenetic analysis of *V. cholerae* isolates

Bayesian trees and Maximum Parsimony trees were calculated resulting from the analysis of four single loci and of the concatenated gene sequences of 299 *Vibrio cholerae* isolates of Lake Neusiedler See. Concatenated sequences comprised the gene fragments of *gyrB*, *recA*, *toxR* and *pyrH*, giving a total length of 2,270 nucleotides. Additionally to the environmental samples eight clinical samples from Austria were included. Moreover, the sequences of four epidemic *Vibrio cholerae* O1 strains retrieved from NCBI GenBank, isolated between 1937 and 1994 in Indonesia, India or Bangladesh, served as references in the depicted trees.

The four Bayesian single loci trees are shown in **Figure 16** to **Figure 18** and the Bayesian tree of the concatenated sequences in **Figure 20**. The topology of the Maximum Parsimony tree of the concatenated dataset and single loci trees were not supported in a course of this study and are therefore shown in **Figure S 1 - Figure S 5. Table 11** and **Table S 2** provide the according data to the Bayesian and Maximum Parsimony trees, respectively.

Composition of *Vibrio cholerae* at the three habitats of the lake is different.

The analysis of the *V. cholerae* isolates by this MLSA revealed 41 different haplotypes within the lake, whereof 20 are only present in the reed belt, one is unique to the intermediate habitat and the remaining 20 different haplotypes are distributed in all three habitats .

In the Bayesian phylogram (**Figure 20**) based on the aligned concatenated sequences, 116 *V. cholerae* isolates from Lake Neusiedler See form one big clade, ID 142, which accounts for around 40 % of all collected isolates. 51 % of these isolates originate from the sampling site open water, 44 % from the intermediate habitat and only the remaining 5 % from the reed belt. Out of 100 *V. cholerae* from the intermediate habitat and out of 99 *V. cholerae* from open water - 51 or 59, respectively, were the same haplotype. A very similar pattern of this big clade is formed by all four loci (→ ID 21, ID 50, ID 68, ID 71). Another conspicuous clade is formed out of 20 isolates, ID 125. This clade is very distant from the others. Isolate-sequences from the intermediate habitat belonging to this group have a three nucleotide gap in the *toxR* sequence and a look at the single gene tree of *toxR* shows clearly that this locus is responsible for this distance (→ ID 83). These isolates seem to have beyond a three nucleotide gap (see also in **Figure 10**) an altered *toxR* sequence. Three further clades like ID 109 (→ *toxR ID* 85), ID 130 (→ *toxR* ID 96) and ID 139 (→ *toxR* ID 74) exist which are only based on differences in the *toxR* sequence.

The very distant groups, ID 144 and ID 145, obviously contain *V. cholerae* isolates which are more different. Their distant arrangement is mainly based on the *gyrB* locus (→ ID 22, 23). Interestingly only locus *recA* differentiates the two isolates of haplotype ID 144 also as a single phylotype (→ ID 47), whereas for *recA* the clade accordant to ID 145 (→ ID 53) includes two other isolates and the *toxR* based tree pools the corresponding isolates of ID 144 and 145 to one phylotype (→ ID 98). However, in all the three mentioned single trees these isolates included in these groups were similarly classified as distant.
Small individual clades comprising one up to 10 isolates (ID 99-102, 106, 111, 113, 114, 118-121, 127-130, 137-138, 144) distributed about the whole tree are formed by more than half by the isolates from the reed belt. Some of these clades are concordant with one or two single gene trees. Haplotype ID 102 and 104 are grouped together to one phylotype in the *toxR* (→ ID 82) and *pyrH* (→ 70) single trees, haplotype ID's 114, 118, 119, 120, 121, 127 are present alike in *recA* and *toxR* single trees and two of them (ID 120, ID 128) are concordant with the *gyrB* tree (→ ID 6, 8). The two distant haplotypes 99 and 100 were influenced by the locus *pyrH*, where they (→ ID 62, 63) were assigned very distant. Some clades including ID's 105, 110, 111, 133 and 137, and are directly concordant to the locus *recA* (→ ID 45, ID 30, ID 42, ID 24, ID 33), and others are equal for *recA* and *toxR:* ID 114 (→ ID 44, ID 76), ID 118 (→ ID 37, ID 88) and ID 121 (→ ID 38, ID 91).

There were also some clades which are equally present in three to four single gene trees and therefore also in the concatenated tree: ID's 115 (→ ID 13, ID 46, ID 67, ID 78), 116 (→ ID 20, ID 29, ID 61, ID 77), 127 (→ ID 14, ID 31-32, ID 95), 135 (→ ID 10, ID 27, ID 57, ID 73).

For comparability environmental isolates were interrelated with clinical references. Four of the clinical strains, which caused infections and sepsis, isolated in the years from 2009 to 2012 were grouped in the concatenated tree near to each other and also next to the clade of three reed belt isolates, ID 121 (→ ID 91-94). Moreover the probability of a common ancestor of these 7 strains is supported (corresponding to posterior probabilities > 0.94). One of these strains, no. 920001/12 was diagnosed according to examinations of the AGES laboratory as suspicious for O1 ElTor subtype Inaba, but did not cluster with the three ElTor reference strains (VcM662, VcMJ1236 and VcN16961). It has to be noted that this particular strain is not member of this clade in the single locus tree of *toxR* (→ ID 86) but its *recA* sequence accounted this grouping near to the three reed belt isolates (→ ID 38, ID 39).

With the differentiation method of the present study three of the four reference strains from GenBank, isolated from patients suffering from cholera between 1937 to 1994, form one clade together with a clinical isolate from Austria isolated in 2008 (CHT81/08). This

formation is based on the *recA* and *toxR sequences*, as can be seen in those single trees (→ ID 25, ID 80). The *recA* sequence of these strains is evidently quite different to the others.

The fourth reference strain *V. cholerae* O395 (a classical cholera strain), haplotype ID 140, based on the concatenated data set is fairly distant from the other strains. This arrangement is due to the *toxR* locus (→ ID 98), since at the single tree of *gyrB* and *pyrH* it is grouped to the other references, and *recA* isn't able to differentiate this one from the majority.

Interestingly two other clinical isolates form one group with five environmental isolates, haplotype ID 105, also with supported probability of a common ancestor. This classification is again caused by an altered *recA* sequence since in the other single trees this group is pooled to other isolates. The eighth clinical isolate CHT83/08 is segregated in all four single loci trees and consequently also in the concatenated tree (ID 103) and has a supported probability to have a common ancestor with 14 isolates of the reed belt (ID's 99-102).

Table 11. Distribution of OTU's in three habitats of Lake Neusiedler See, July 2012.

Phylotype ID	Locus	Habitat			Total this study	References	Clinical	Total
		Intermediate habitat	Open water	Reed belt				
1	*gyrB*	15	11	6	32			32
2	*gyrB*	10	7	10	27			27
3	*gyrB*	2	1	18	21		2	23
4	*gyrB*						1	1
5	*gyrB*	1	3	13	17			17
6	*gyrB*			1	1			1
7	*gyrB*						1	1
8	*gyrB*			2	2			2
9	*gyrB*	1			1			1
10	*gyrB*	5	6	6	17			17
11	*gyrB*			10	10			10
12	*gyrB*			7	7			7
13	*gyrB*	1	1		2			2
14	*gyrB*			7	7			7
15	*gyrB*	5			5		1	6
16	*gyrB*					4	1	5
17	*gyrB*	1	1		2			2
18	*gyrB*			3	3			3
19	*gyrB*		1	9	10		2	12
20	*gyrB*	4	4		8			8
21	*gyrB*	54	61	6	121			121
22	*gyrB*			2	2			2
23	*gyrB*	1	3		4			4
24	*recA*	3	3	4	10			10
25	*recA*					3	1	4
26	*recA*	4	2		6			6
27	*recA*	6	6	2	14			14
28	*recA*	15	11	5	31			31
29	*recA*	4	4		8			8
30	*recA*	1	2	4	7			7
31	*recA*			3	3			3
32	*recA*			7	7			7
33	*recA*			1	1		1	2
34	*recA*			1	1			1
35	*recA*			13	13			13
36	*recA*			1	1		1	2
37	*recA*			7	7			7
38	*recA*			3	3			3
39	*recA*						1	1
40	*recA*			5	5			5
41	*recA*			2	2			2
42	*recA*			2	2			2
43	*recA*	1	2	7	10			10
44	*recA*			10	10			10
45	*recA*	2	1	2	5		2	7
46	*recA*	1	1		2			2
47	*recA*			2	2			2
48	*recA*	1		1	2			2
49	*recA*			2	2			2
50	*recA*	59	62	13	134	1	2	137
51	*recA*	1	1		2			2
52	*recA*		1	2	3			3
53	*recA*	2	3	1	6			6
54	*pyrH*	21	17	15	53		2	55
55	*pyrH*	1	2	7	10			10
56	*pyrH*	4	2		6			6
57	*pyrH*	6	5	2	13			13
58	*pyrH*	1	1	11	13			13
59	*pyrH*	5	1	3	9			9
60	*pyrH*						1	1
61	*pyrH*	4	4		8			8
62	*pyrH*			6	6			6
63	*pyrH*			1	1			1
64	*pyrH*					4	1	5
65	*pyrH*	1		1	2			2
66	*pyrH*	2	3	1	6			6
67	*pyrH*	1	1		2			2
68	*pyrH*	54	63	44	161		3	164
69	*pyrH*						1	1
70	*pyrH*			9	9			9
71	*toxR*	54	61	6	121			121
72	*toxR*	7	5	4	16			16
73	*toxR*	4	4	2	10			10
74	*toxR*	5			5			5
75	*toxR*	2	2	4	8			8
76	*toxR*			10	10			10
77	*toxR*	4	4		8			8
78	*toxR*	1	1		2			2
79	*toxR*						1	1
80	*toxR*					3	1	4
81	*toxR*	2	1	10	13		2	15
82	*toxR*			9	9			9
83	*toxR*	9	8	3	20			20
84	*toxR*	1	1	6	8			8
85	*toxR*		2	5	7			7
86	*toxR*	1	3	10	14		1	15
87	*toxR*			1	1			1
88	*toxR*			7	7			7
89	*toxR*			1	1			1
90	*toxR*			1	1			1
91	*toxR*			3	3			3
92	*toxR*						1	1
93	*toxR*						1	1
94	*toxR*						1	1
95	*toxR*			11	11			11
96	*toxR*	1	1	3	5			5
97	*toxR*	7	3	1	11			11
98	*toxR*	2	3	3	8	1		9

Haplotype ID	Locus	Habitat			Total this study	References	Clinical	Total
		Intermediate habitat	Open water	Reed belt				
99	concatenated			6	6			6
100	concatenated			1	1			1
101	concatenated			1	1			1
102	concatenated			6	6			6
103	concatenated						1	1
104	concatenated			3	3			3
105	concatenated	2	1	2	5		2	7
106	concatenated			1	1			1
107	concatenated		1	4	5			5
108	concatenated	1		2	3			3
109	concatenated		2	5	7			7
110	concatenated	1	2	4	7			7
111	concatenated			2	2			2
112	concatenated		1	2	3			3
113	concatenated			2	2			2
114	concatenated			10	10			10
115	concatenated	1	1		2			2
116	concatenated	4	4		8			8
117	concatenated					3	1	4
118	concatenated			7	7			7
119	concatenated			1	1			1
120	concatenated			1	1			1
121	concatenated			3	3			3
122	concatenated						1	1
123	concatenated						2	2
124	concatenated						1	1
125	concatenated	9	8	3	20			20
126	concatenated	6	3		9			9
127	concatenated			7	7			7
128	concatenated			2	2			2
129	concatenated			2	2			2
130	concatenated			3	3			3
131	concatenated	1	1		2			2
132	concatenated	1		1	2			2
133	concatenated	3	3	4	10			10
134	concatenated	4	2		6			6
135	concatenated	4	4	2	10			10
136	concatenated	2	2		4			4
137	concatenated			1	1			1
138	concatenated			3	3			3
139	concatenated	5			5			5
140	concatenated					1		1
141	concatenated	1		1	2			2
142	concatenated	51	59	6	116			116
143	concatenated	3	2		5			5
144	concatenated			2	2			2
145	concatenated	1	3		4			4

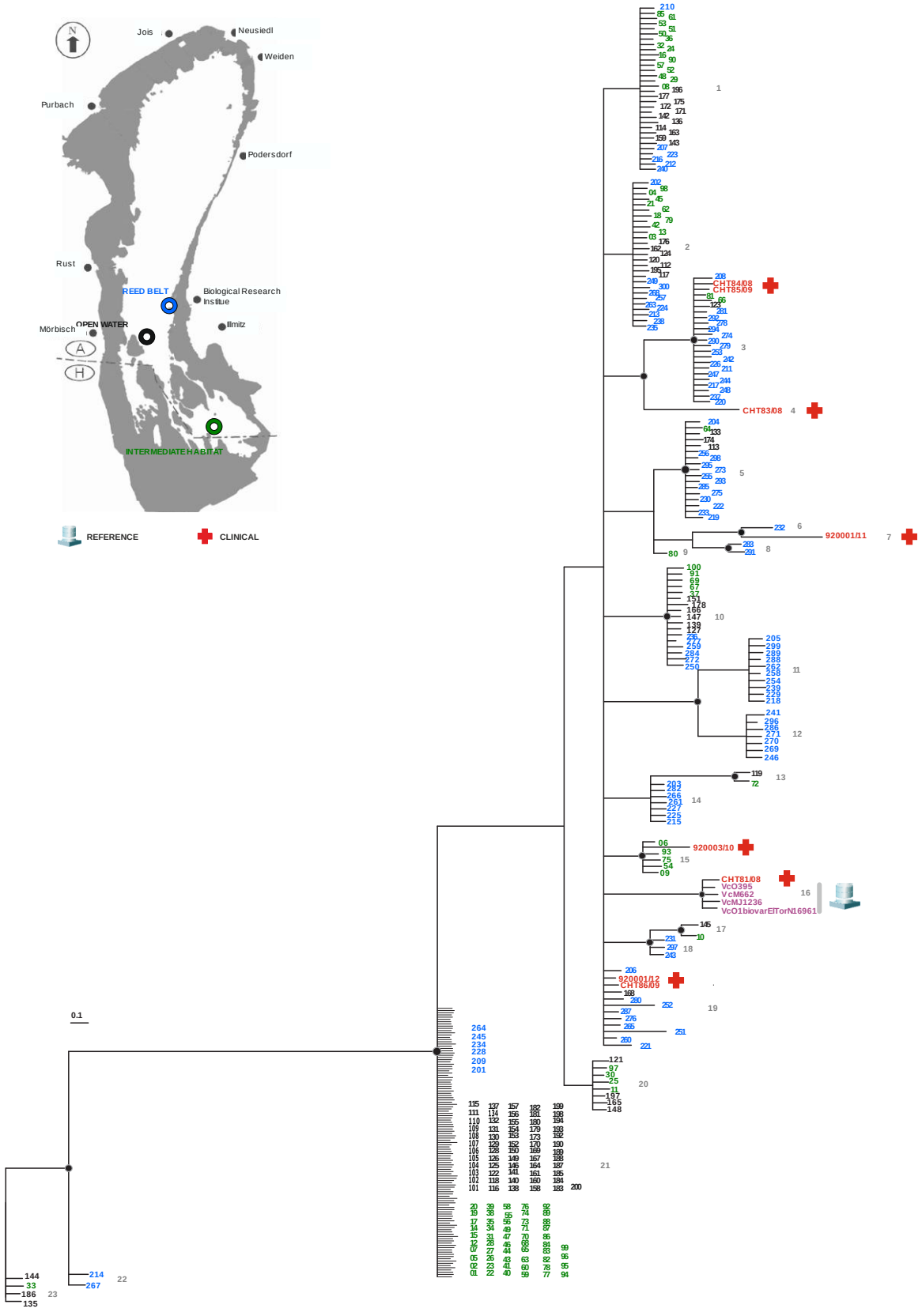


Figure 16. Single loci tree *gyrB*.

Bayesian tree resulting from analysis of the *gyrB* fragment of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Branching points indicated by black circles correspond to posterior probabilities of > 0.94 (="supported"). Haplotype ID's are given in grey colored numbers, according to **Table 11.** Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

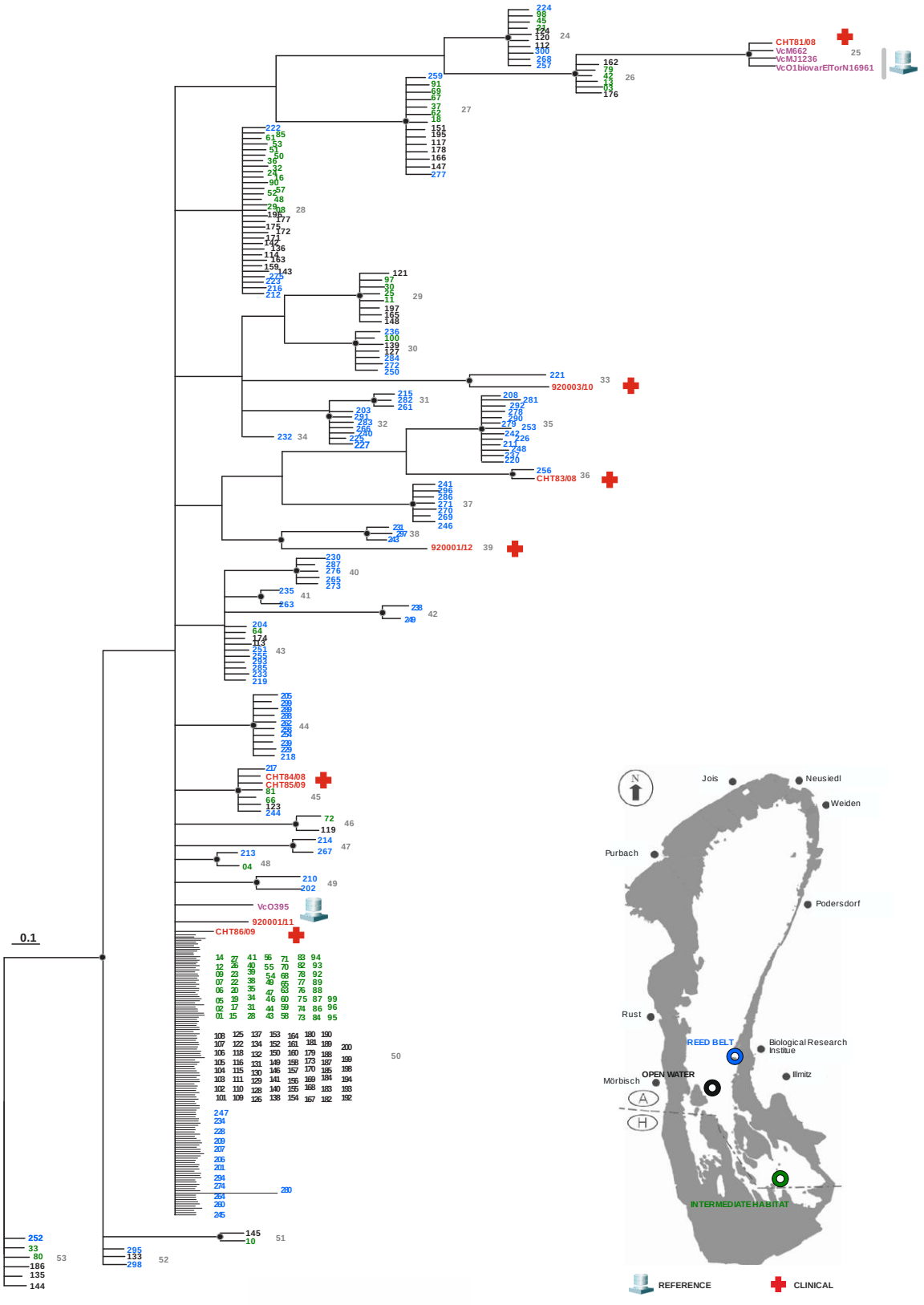


Figure 17. Single loci tree *recA*.
Bayesian tree resulting from analysis of the *recA* fragment of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Branching points indicated by black circles correspond to posterior probabilities of > 0.94 (="supported"). Haplotype ID's are given in grey colored numbers, according to **Table 11.** Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

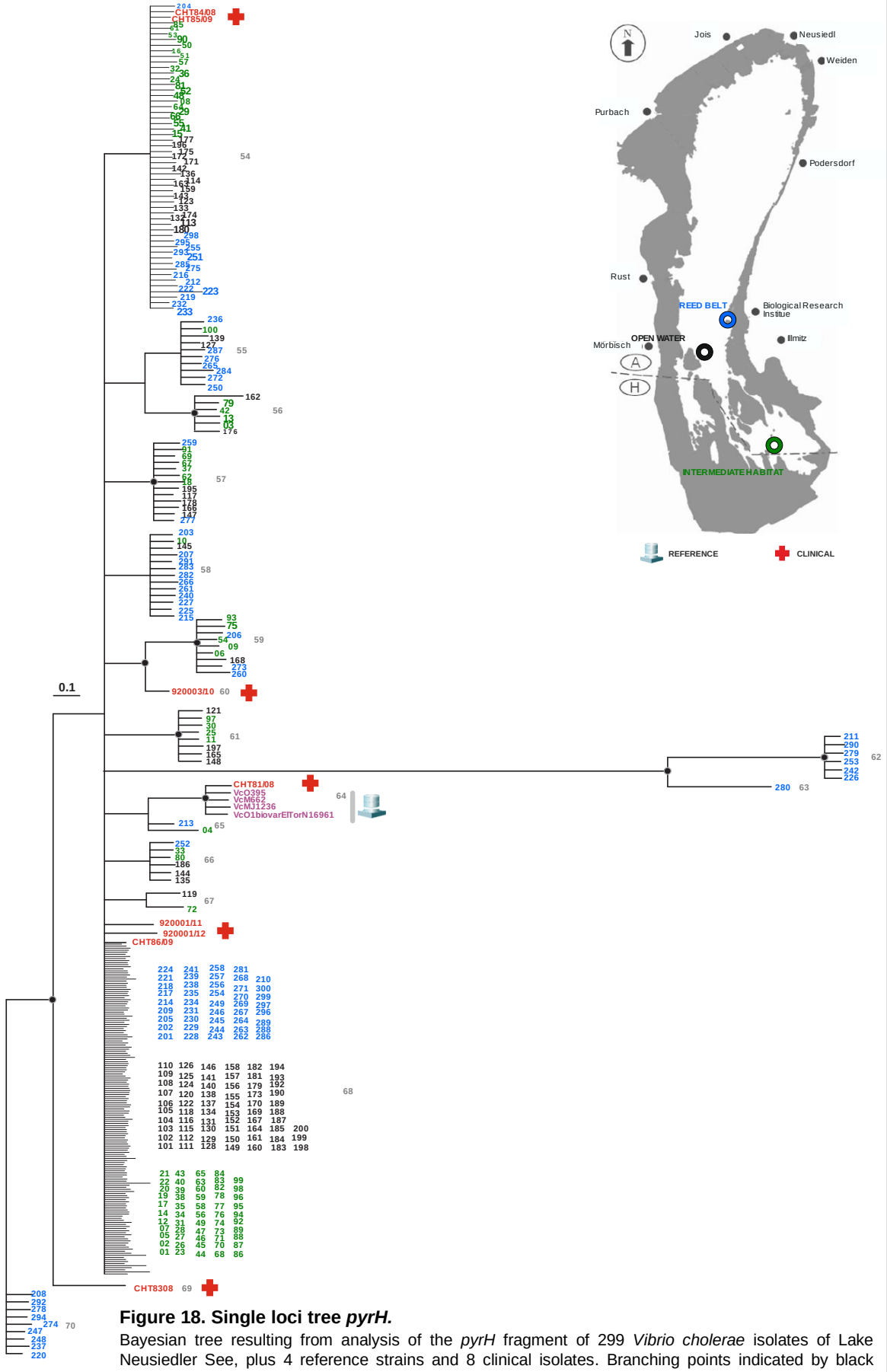


Figure 18. Single loci tree *pyrH.*

Bayesian tree resulting from analysis of the *pyrH* fragment of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Branching points indicated by black circles correspond to posterior probabilities of > 0.94 (=”supported”). Haplotype ID’s are given in grey colored numbers, according to **Table 11.** Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

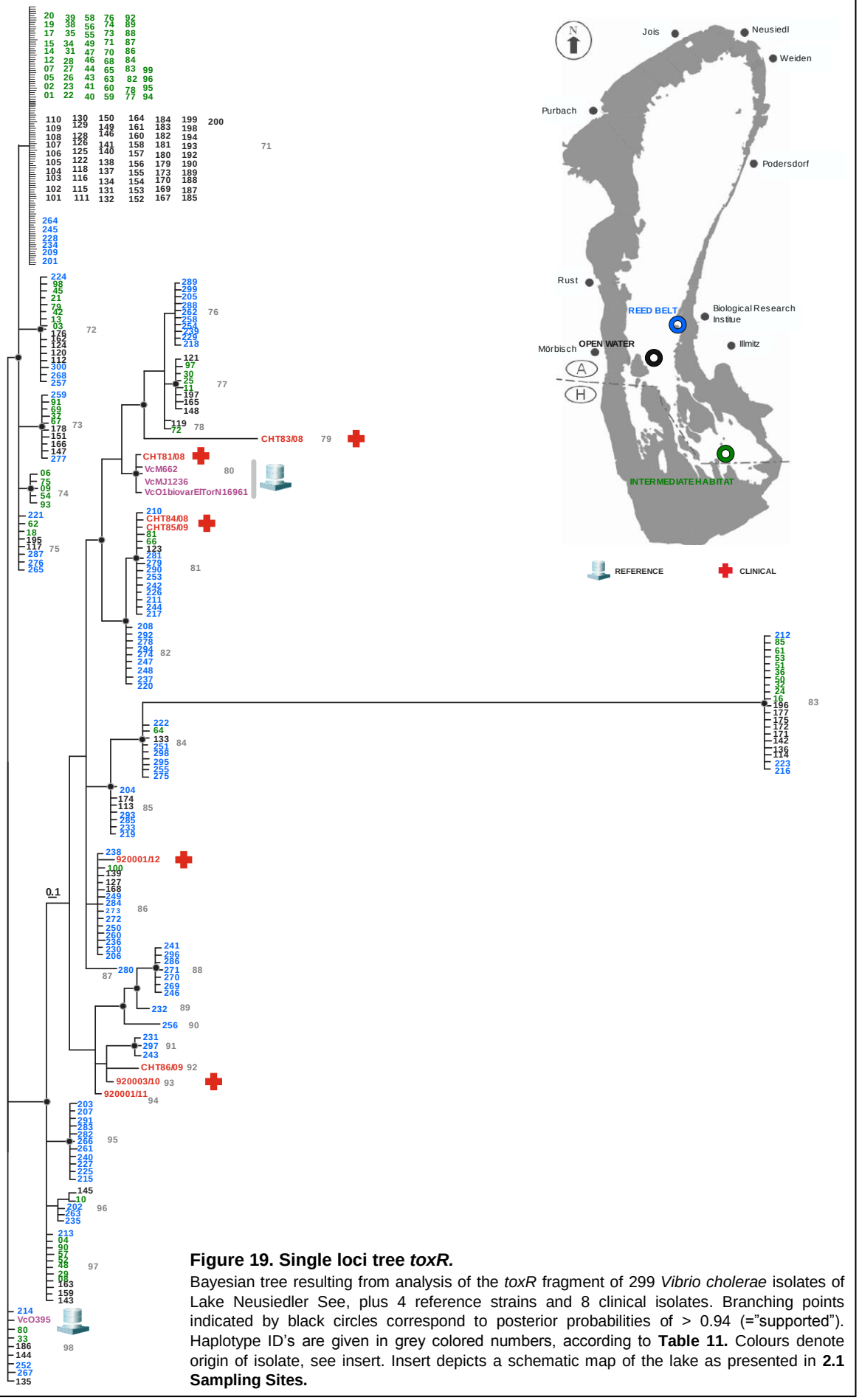


Figure 19. Single loci tree *toxR*.

Bayesian tree resulting from analysis of the *toxR* fragment of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Branching points indicated by black circles correspond to posterior probabilities of > 0.94 (="supported"). Haplotype ID's are given in grey colored numbers, according to **Table 11.** Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

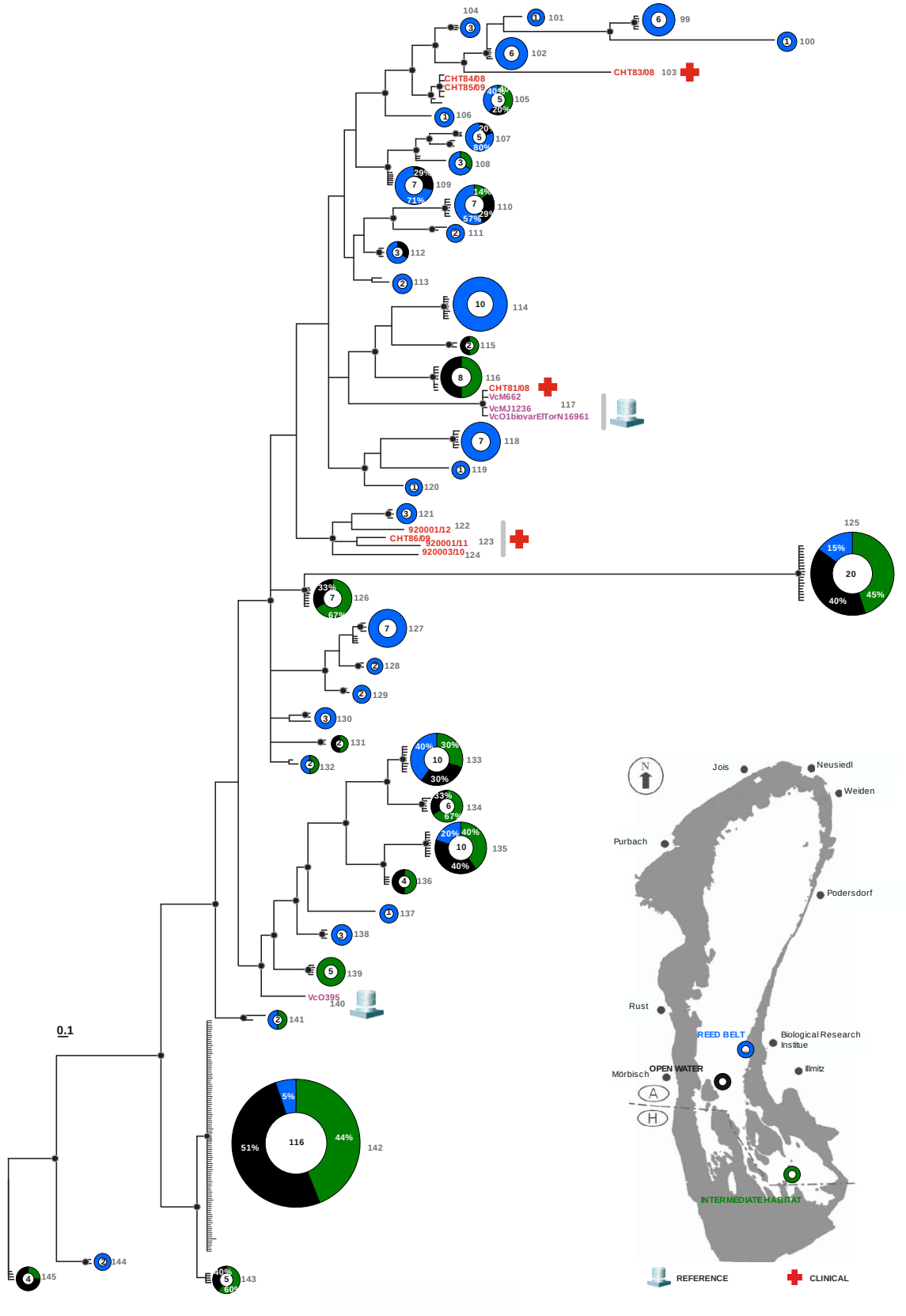


Figure 20. Multilocus tree of *V. cholerae* isolates of Lake Neusiedler See.
Bayesian tree resulting from analysis of the concatenated sequences comprising of *gyrB, recA, pyrH* and *toxR* [2,270 bp] of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Branching points indicated by black circles correspond to posterior probabilities of > 0.94 (="supported"). Haplotype ID's are given in grey colored numbers, according to **Table 11**. Piecharts indicate the distribution of each clade among sampling sites, numbers of isolates per sampling site declared in percent and the total number of isolates is centered. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

3.5.2 COMPARISON OF *VIBRIO CHOLERAE* DIVERSITY BETWEEN DIFFERENT HABITATS

Each position (site) in the alignment of strains is either constant or variable. Constant sites (or monomorphic sites) are identical in the whole alignment; these positions don't say anything about phylogeny (48). Variable (or segregation or polymorphic sites) but parsimony uninformative sites are those which are different in maybe just one of the analysed strains. But if the site contains at least two types of nucleotides and they occur with a minimum frequency of two this site is parsimony informative (48).

The parsimony informative sites of the concatenated sequences comprising the four gene fragments of *V. cholerae* isolates in Lake Neusiedler See of the three different habitats were analysed for the total number of parsimony informative sites and number of shared parsimony informative sites and plotted in two Venn diagrams (**Figure 21 A, B**). The number of unique parsimony informative sites of each sequence set is depicted on the external sides of the diagram and between them are the numbers of shared sites among each habitat. The highest number of parsimony informative sites had the sequence set "reed belt" with a total of 225, whereof 43 were unique to this data set. The *V. cholerae* sequences originating from "open water" provided 190 parsimony informative sites, whereof only two were unique to this ecosystem. A similar situation was observed for the sequences of the "intermediate habitat" where 175 parsimony informative sites were found of which only two were unique to this site. The sampling set references/clinical provided 6 unique parsimony informative sites out of 57 in total.

The number of parsimony informative sites of the concatenated sequences shared by all of the three analysed habitats of the lake was 158. If the sequences of the references and clinical isolates are added, only 122 remained unique for the habitats of the lake, while 36 of them are present in the environmental and reference/clinical isolates.

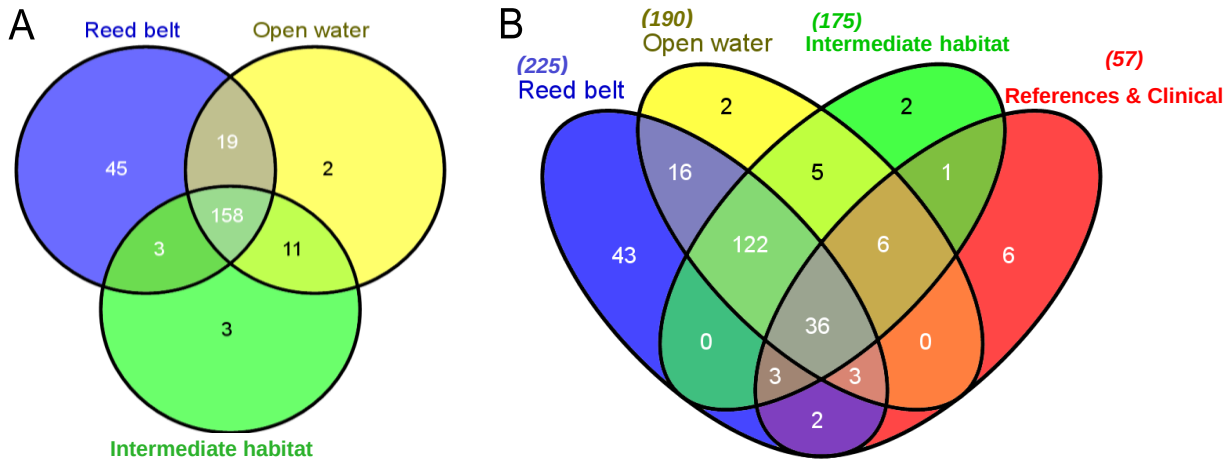


Figure 21. Venn diagrams of *V. cholerae* diversity in three ecosystems of Lake Neusiedler See based on parsimony informative sites. Diagrams created by the tool VENNY (63) based on the parsimony informative sites of the concatenated sequences of *V. cholerae* isolates from three different ecosystems in Lake Neusiedler See (A) and after adding the sequences of references and clinical isolates, which were pooled to a fourth sample set (B). Quantity of parsimony informative sites of each sample set noted in italics and brackets in B.

In order to estimate whether the sample size of this study was sufficient to characterize the entire diversity of *V. cholerae* in each habitat of the lake Neusiedler See rarefaction curves were calculated. **Figure 22** depicts the rarefaction curves to estimate the species richness of *Vibrio cholerae*.

For open water and the intermediate habitat similar values were obtained. The *V. cholerae* isolates based on the concatenated sequences were grouped to 17 and 18 haplotypes, respectively and the curves of both sites were saturated at the number of analysed isolates. Based on these curves, one more haplotype would be revealed if 20 additional isolates were included. The reed belt habitat behaved in a comparable manner although 32 different haplotypes were identified. The rarefaction curve shows that the analysis of 100 isolates is nearly leading to saturation; additional 10 isolates would reveal one additional haplotype.

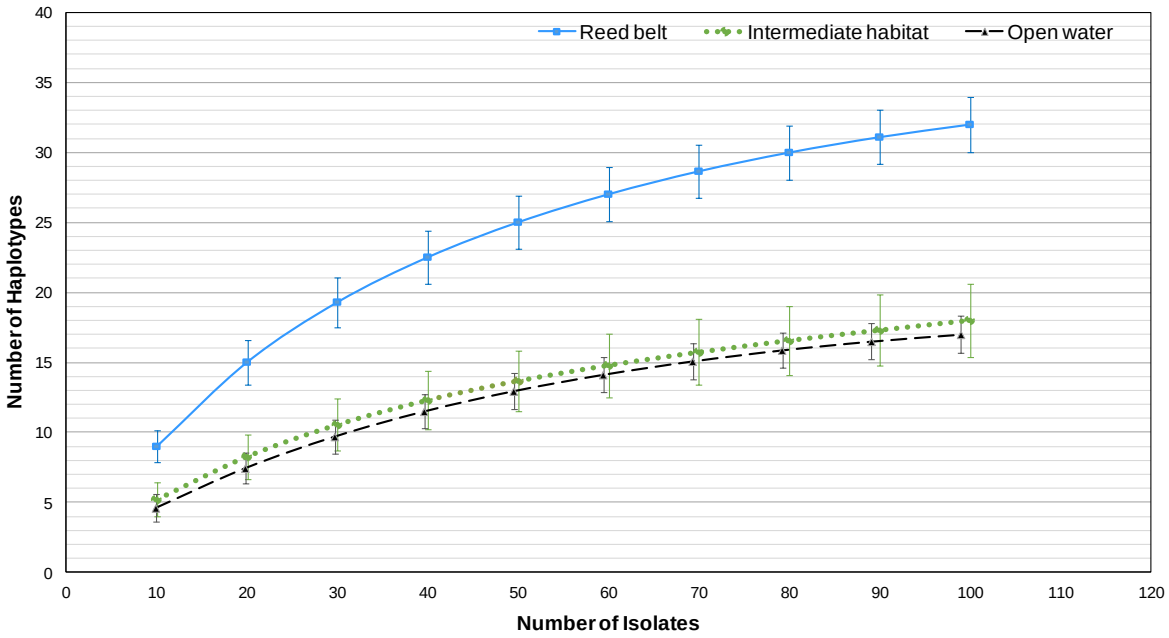


Figure 22. Rarefaction curves of the three habitats.
Green dotted line: intermediate habitat; blue line: open water; black dashed line: reed belt. Vertical bars correspond to standard deviation. X-axis: Number of isolates; y-axis: Number of haplotypes.

4 DISCUSSION AND CONCLUSION

As part of a survey program started in 2001, this thesis targeted at the establishment of a suitable multilocus sequence analysis procedure to determine the diversity of *Vibrio cholerae* strains in three habitats of Lake Neusiedler See. Therefore, five genes (four housekeeping genes and one virulence-associated gene) were selected from recent literature. Single PCR was established for each of these genes and PCR products were sequenced by an external sequencing service. For MLSA, 100 *Vibrio cholerae* isolates per each habitat were analysed and interrelated with references and clinical strains of *V. cholerae*. One housekeeping gene, that delivered too few information concerning phylogenetic diversity, was excluded from analysis.

Different ecology of the three sampling sites.

The measured environmental parameters of the three sampling sites revealed significant ecological differences within the lake corresponding to a different inherent biodiversity. The quantification of *V. cholerae* with a culture based and a culture-independent method (CARD-FISH combined with SPC) showed similar trends, despite higher concentrations were found with the latter method. *V. cholerae* numbers were highest in the reed belt, where also highest total bacterial numbers were observed. Lowest cell numbers and *V. cholerae* concentrations were observed in the open water habitat. Bacterial concentrations in the intermediate habitat were in between. The water in the reed belt is mainly characterized by a much lower turbidity (lower concentrations of total suspended solids and higher Secchi depth) as well as by a much higher concentration of dissolved organic carbon (DOC) and – concomitantly – by higher concentrations of humic substances (lower humics ratio). These elevated DOC and humics concentrations are caused by the presence of decomposing reed. Another study (61) on environmental parameters in Lake Neusiedler See had found out that the wind-speed correlates significantly with fluctuations in turbidity, but also in total phosphorus (P_{tot}) and chlorophyll a. The measured total phosphorus concentrations of the open water and the intermediate habitat showed a similar pattern throughout the summer, while P_{tot} of the reed belt was rather constant, because this habitat is less exposed to the wind. A similar observation was made for chlorophyll a values, with much higher fluctuations in the open water habitat than in the reed belt.

MLSA revealed 41 different haplotypes in the lake.

For the present study four housekeeping genes and one virulence-associated gene of the *Vibrio cholerae* genome were chosen based on their application frequency in former publications dealing with *V. cholerae* (mentioned in **2.3.1**). Fragments of *gyrB* (487 nt), *recA* (648 nt), *toxR* (658 nt), *pyrH* (477 nt) and the later on excluded locus *rpoA* (712nt) of 299 *V. cholerae* isolates plus references were sequenced. This combination of genes has not been applied before for the differentiation of environmental *V. cholerae*. However, most of those studies have considered the whole Vibrio genus (80, 81, 84) or cholera-toxigenic *V. cholerae* O1 or O139 strains (27, 51, 79).

The analysis of the first 148 concatenated sequences revealed that locus *rpoA* wasn't able to differentiate the isolates (data not shown). The DNA-directed RNA polymerase subunit alpha is - according to the results - highly conserved among the *V. cholerae* subtypes in the lake Neusiedler See and not applicable for usage as a phylogenetic marker in MLSA. It was therefore excluded from the phylogenetic analysis. With the remaining dataset comprising four genes, the *V. cholerae* population in three different habitats of Lake Neusiedler See could be resolved in detail. The numbers of identified haplotypes are illustrated in **Figure 23**. 41 different *Vibrio cholerae* haplotypes were identified in total, out of 299 analysed isolates. When references and clinical isolates are included, 45 haplotypes were identified. Almost 40 % of all the collected environmental isolates were accounted to one type, which occurred rather equally in the open water and the intermediate habitat but less often in the reed belt. The rest of about 60 % of the environmental isolates was separated into 40 haplotypes.

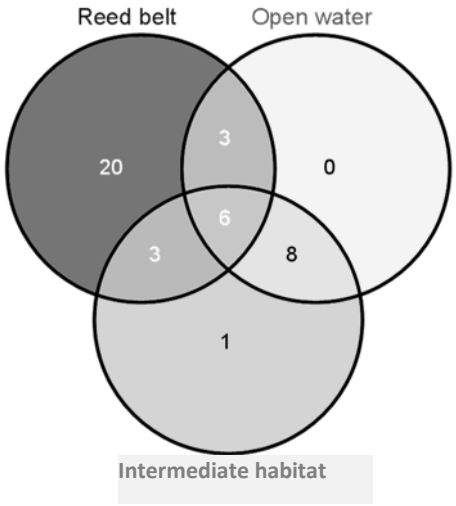


Figure 23. Venn diagram of haplotypes. Diagram created by the tool VENNY (63) based on the distribution of different *V. cholerae* haplotypes in three different ecosystems in Lake Neusiedler See. The three circles indicate the ecosystems and the numbers the quantity of unique or shared haplotypes.

The diversity of *V. cholerae* strains in the reed belt was twice as high compared to the other two habitats. In the reed belt 32 haplotypes were identified, compared to 17 and 18 in the open water and the intermediate habitat, respectively. Moreover, 20 of those were unique for this habitat, while in the intermediate habitat only a single unique haplotype was found and none was detected in the open water. Only 6 haplotypes were in common for all three habitats. As the rarefaction curve for the reed belt isolates suggests the

analysis of 10 additional isolates will reveal additional haplotypes, in contrast to the other two habitats where the rarefaction curve was almost saturated and 20 additional isolates would lead to the detection of one additional haplotype.

In contrast to the finding, that *V. cholerae* is inhibited by humic substances (46), the highest diversity of haplotypes was found in the reed belt. Obviously, this habitat offers more ecological niches than the open water area, where winds are continuously homogenizing the water column of this shallow lake. What this high genetic diversity of *V. cholerae* strains in the reed belt means in terms of phenetic diversity is completely unknown. From recent literature it is known that environmental *V. cholerae* is a highly diverse species (23) with a high genetic plasticity (8), but it has only poorly been investigated how this genetic realm translates to differences in environmental performance (42). For other species it has been reported that sub-species level diversity (microdiversity) can be relevant with highly different environmental behaviour of different clones (37). It can thus be assumed that similar may apply also to *V. cholerae*.

The outcomes of different studies dealing with Vibrios in a similar aspect were very diverse, depending on the different examined species and strains: either their main objective was to differentiate Vibrio species (80-82, 84) or if *Vibrio cholerae* was used they were dealing with O1/O139 strains (27, 51, 58, 79) or the used loci and used methods diverged too much for comparison (62, 72, 78). Nevertheless two studies have to be discussed briefly.

A study from 2011 (72) titled “Ecology and Genetic Structure of a Northern Temperate *Vibrio cholerae* Population Related to Toxigenic Isolates” analysed in total 31 *V. cholerae* non-O1/non-O139 isolates, 12 collected in 2008 and 19 in 2009, whereof 17 originated from water, 10 from oysters and 4 from the sediment from Great Bay Estuary in New Hampshire. Schuster, et al. (72) analysed the diversity by MLSA (based on the genes *gapA*, *gyrB*, *pyrH*, *recA* and *topA*) and reported to have revealed a high diversity in this geographically confined ecosystem. They reported that 75 % of their isolates correlated with site but no isolate correlated with year. The strongest correlating factor in the study was major rainfall events (72). Based on the finding of this study it seems clear that to capture the whole diversity of a *V. cholerae* population one might have to use more than 31 isolates from three different ecological niches. At least estimations of the appropriate sample number and the microdiversity in each specific habitat have to be performed. For predicting ecological correlations more than a handful of collected isolates are needed for convincing results.

Another study of 2010 (40) “Recombination Shapes the Structure of an Environmental *Vibrio cholerae* Population” used MLSA (based on the genes *gyrB*, *mdh*, *recA*, *idh*, *asd*,

dnaE, *nagB*) to characterize homologous recombination and structure among 152 environmental *Vibrio cholerae* isolates, 13 other putative *Vibrio* isolates from coastal waters and sediments in central California and four clinical *Vibrio cholerae* isolates. Keymer and Boehm (40) reported that frequent recombination and not necessarily weak selection may be responsible for the extensive intra-species diversity but without knowledge of potentially niche-adaptive genes or phenotypes in coastal ecosystems. Thus, putting observations of genetic or phenotypic diversity into the context of *Vibrio cholerae* ecology and evolution is a difficult task (40). Those 152 examined environmental *Vibrio cholerae* isolates originated from coastal waters and sediments in central California - without specific description of sampling sites and no differentiation to habitat or time of sampling, and can therefore only display a snapshot of the total *Vibrio cholerae* population. To depict the microdiversity of *Vibrio cholerae* populations, intra-species richness estimations were indispensable.

Clusters of reference strains and clinical isolates.

By the used MLSA approach two haplotypes were observed, that grouped together clinical isolates with environmental isolates, indicating that patients that developed disease in earlier years (2008-2011) had been infected with *V. cholerae* strains that were still present in the lake in 2012. A highly interesting clade was formed by three O1 positive reference strains (ElTor serotypes) and one clinical isolate from Austria, with an identical position in all four single loci trees and the concatenated tree. That the fourth reference strain *V. cholerae* O395 forms a separate cluster in the concatenated tree is due to the fact that this is a Classical *V. cholerae* strain (in contrast to the three ElTor strains). Its distant position is caused by its different *recA* and *toxR* sequences.

Comparison of the concatenated tree with single loci trees.

The loci *toxR*, *recA* and *gyrB* mainly affect the arrangement of clades of the concatenated tree. Locus *pyrH* had the lowest effect on the concatenated tree. Some very distant clades are created clearly by one single locus. Such a situation is given for several cases of *toxR* and *recA*; a few but very distant clades were caused by *gyrB* and *pyrH*. Several of the formed clades are present equally in all trees, but most equal clades are formed by *toxR* and *recA*.

The analysis of polymorphic sites showed that half of the parsimony informative sites of the concatenated sequence originate from the *toxR* sequences. Additionally, the number of segregation sites of *toxR* is more than double compared to the other loci. In this respect the primary structure of the corresponding protein, the amino acid sequence, would be interesting to look at in detail, and furthermore, if there are any functional

changes in the protein. In the present study, *toxR* seems to be the most diagnostic locus for the investigated environmental isolates. 28 different haplotypes were identified considering only this locus. With the concatenated set of four genes, 41 haplotypes were revealed. The locus *gyrB* acted as a further differentiation locus by grouping the isolates into 23 different clades. *RecA* revealed the most haplotypes, 30, but mirrored mostly the topology formed by *toxR*.

The aim to establish a suitable multilocus sequence analysis procedure for determining the diversity of *Vibrio cholerae* isolates in Lake Neusiedler See was accomplished by this MLSA, whereas one locus was excluded and maybe a replacement with another locus will reveal further discrimination or further ensure the topology of the concatenated tree. The environmental *Vibrio cholerae* isolates clustered in the phylogram to 41 different clades whereas 40% of all isolates were assigned to the same haplotype. *V. cholerae* strains which caused infections in 2008-2011 were indicated to have been still in the lake in 2012 because of clustering with environmental samples. Briefly, habitat “reed belt” showed a higher diversity than the other two sites and richness estimations verified a sufficient sampling size to characterize the entire diversity of the investigated habitats; at “open water” and the “intermediate habitat” even fewer isolates would be necessary.

OUTLOOK

To relate the observed diversity of *Vibrio cholerae* isolates of Lake Neusiedler See to isolates from other regions in Europe, *V. cholerae* isolates originating from different countries and different ecosystems would be needed. Therefore, further isolates (provided by international cooperation partners) will be included in the present analysis. By this, it could be determined, whether the *V. cholerae* population of the lake Neusiedler See represents an endemic local population or whether similar strains can be found across Europe (e.g. from coastal Italy, the Baltic Sea, the North Sea, the Danube Delta) which have been transmitted by specific vectors (birds, humans).

Since up to now, no O1 or O139 *V. cholerae* strain has been isolated from the lake and none of the examined isolates contained the *ctx* or *tcp* genes (46), the probability that cholera causing strains are present in the lake is very low (44). Nevertheless, in times of climate warming and increased travellers mobility the probability that cholera causing strains invade into the lake is increasing (44). Also, migrating birds present in the bird sanctuary of the national park Neusiedler See – Seewinkel can act as sources of cholera causing strains from epidemic regions like Africa or Asia (44) and may add to the diversity of the *Vibrio cholerae* strains in the lake.

5 SUMMARY

Vibrio cholerae is a natural inhabitant of aquatic ecosystems and a human pathogen causing the acute life-threatening diarrhoeal disease cholera. The pathogenesis of *V. cholerae* is mainly based on the expression of the virulence genes *ctx* and *tcp*. Mediated by the toxin co-regulated pili (TCP) ingested *V. cholerae* can colonize the epithelium of the small intestine where production of cholera toxin (CTX) disrupts the ion transport followed by a watery diarrhoea which leads to death if not treated. However, only the two serogroups O1 and O139 out of 200 serogroups are able to cause epidemic cholera. All other serogroups, collectively named non-O1/non-O139, are associated with mild to severe watery diarrhoea, or wound, ear and blood infections.

In Austria, 13 infections with non-O1/non-O139 which caused otitis media or otitis externa and also one lethal sepsis were reported between 2000 and 2005; five of them were associated with Lake Neusiedler See. In 2001 a surveillance program was started to observe the ecology, genetic diversity and potential pathogenicity of *V. cholerae* by Kirschner and colleagues. A previous genotyping approach based on pulsed-field gel electrophoresis identified 64 different genotypes out of 84 isolates of Lake Neusiedler See which fostered further investigations of the genetic diversity of *Vibrio cholerae* in the lake.

In this study, the genetic diversity of *V. cholerae* in three habitats of Lake Neusiedler See was assessed after establishment of a suitable multilocus sequence analysis (MLSA) procedure. 300 *V. cholerae* isolates were collected from three different habitats in the lake. These isolated presumptive *V. cholerae* were confirmed as *V. cholerae* by detection of a specific *V. cholerae* outer membrane protein – *ompW* – via PCR. Additionally, 8 clinical isolates of *Vibrio cholerae* and four O1 reference strains were included. For the MLSA four housekeeping genes, *gyrB*, *recA*, *pyrH*, *rpoA* and one virulence-associated gene, *toxR*, were chosen, based on the research of recent literature. The used primer sequences for PCR based gene amplification originated from different previous analyses. For accurate amplification, different PCR conditions, like variations in annealing temperature or primer concentration, were successively tested, and the amplified PCR products sequenced at a commercial sequencing facility. The obtained sequences were manually processed. They were first checked for quality, forward and reverse sequences were assembled, trimmed to equal length, concatenated and finally aligned by the program “Muscle”. For phylogenetic analysis, Metropolis-coupled Markov-chain-Monte-Carlo sampling was done by the program “MrBayes” and nucleotide properties and some specific details of the phylogenetic analysis of the different loci were computed by “MrBayes” and “DnaSP”. This analysis showed that *toxR*

contributed 50 % of the parsimony informative sites of the concatenated sequences, and the number of polymorphic sites in this locus was double compared to the other loci. The computed phylograms revealed that locus *rpoA* wasn't applicable for differentiation of environmental *Vibrio cholerae* isolates.

Nevertheless, MLSA of the concatenated sequence set of the remaining four loci resolved the *V. cholerae* population in the three different habitats of Lake Neusiedler See in detail and revealed 41 clades of environmental *V. cholerae* with different distributions within the three different habitats. 40 % of all isolates were grouped together to one haplotype of *V. cholerae* which occurred more frequently in the open water and the intermediate habitat, and the remaining 60 % of the environmental samples were separated into 40 different haplotypes. The diversity of *V. cholerae* strains in the reed belt was twice as high compared to open water and the intermediate habitat. 20 of the revealed haplotypes were unique for the habitat "reed belt", while in the intermediate habitat only a single unique haplotype occurred and no one was unique for "open water". The remaining 20 different haplotypes were distributed in all three habitats. Interrelation of clinical isolates with the environmental isolates of the lake revealed that *V. cholerae* strains which were causing diseases in earlier years (2008-2011) were still present in the lake in 2012. By comparing the concatenated tree with the single loci trees it was obvious that *toxR*, *recA* and *gyrB* mainly affected the concatenated tree and caused some very distant clades concordant with the concatenated tree. For another comparison of *Vibrio cholerae* diversity between the different habitats, the parsimony informative sites within the habitats were compared.

To estimate the intra-species richness of *Vibrio cholerae* in the lake, rarefaction curves were generated, indicating that the number of isolates analysed was sufficient to depict the diversity of *Vibrio cholerae* in the different habitats of Lake Neusiedler See.

6 ZUSAMMENFASSUNG

Vibrio cholerae ist ein natürlicher Bestandteil aquatischer Ökosysteme aber auch ein humanpathogenes Bakterium, das besonders in Entwicklungsländern wo sauberes Trinkwasser nicht verfügbar und Hygienestandards mangelhaft sind, Ausbrüche bis hin zu Epidemien verursachen kann.

V. cholerae ist der Erreger der lebensbedrohlichen Krankheit Cholera die bei einer Dunkelziffer von 100.000 Toten jährlich rangiert von denen nur ein Bruchteil der WHO gemeldet werden. Die Pathogenität von *V. cholerae* basiert hauptsächlich auf der Expression der Virulenzgene *ctx* und *tcp*. Nach der Aufnahme von mit *Vibrio cholerae* kontaminiertem Wasser oder fester Nahrung sind die Bakterien mithilfe des „Toxin Coregulated Pilus“ (TCP) in der Lage das Epithel des Dünndarms zu besiedeln. Die Produktion von Cholera Toxin (CTX) führt zu einer pathologischen Veränderung des Ionen Haushalts. Das hat wässrigen Durchfall zur Folge, der unbehandelt zum Tod führen kann. Von den derzeit 200 bekannten Serogruppen von *V. cholerae* sind jedoch nur zwei, O1 und O139, Auslöser der endemischen Cholera. Vertreter der anderen Serogruppen werden zusammengefasst als nicht-O1/nicht-O139, und können Durchfallserkrankungen mit milderem Verlauf, sowie Wund-, Ohr- oder Blutinfektionen auslösen.

In den Jahren 2000 bis 2005 traten in Österreich 13 Infektionen durch nicht-O1/nicht-O139 *V. cholerae* auf - für fünf davon wurden Aufenthalte am Neusiedler See belegt. Die Patienten litten an Mittelohrentzündung, Entzündungen des äußeren Ohres und einer tödlich verlaufenden Sepsis eines immunsupprimierten Fischers. Aufgrund dieser Vorkommnisse wurde 2001 ein Monitoringprogramm gestartet um die Ökologie, genetische Diversität und potentielle Pathogenität von *V. cholerae* zu erfassen. In einer späteren Studie wurden mittels Pulsfeldgelelektrophorese 84 *Vibrio cholerae* Isolate zu 64 unterschiedlichen Genotypen differenziert.

Die sich aufwerfenden Fragen nach der genetischen Diversität von *V. cholerae* im See wurden nun in dieser Studie mit einer Multilocus Sequenz Analyse beantwortet. Beruhend auf dem Vergleich der DNA-Sequenz ausgewählter, langsam evolvierender Haushaltsgene, wird durch minimale Allelvariationen Auskunft über die Zusammensetzung der Population gegeben. Hierfür wurden 299 Isolate von drei repräsentativen Habitaten gesammelt, kultiviert und mittels *ompW* Polymerasekettenreatkion als *Vibrio cholerae* identifiziert. Vier Haushaltsgene plus ein Virulenzgen und die dazugehörigen Primer wurden basierend auf vorangegangenen Studien ausgewählt, die spezifischen Polymerasekettenreaktionen entwickelt und schließlich 1500 Amplifikate produziert. Weiters wurden 8 klinische *V. cholerae* Isolate

und vier O1 Referenzsequenzen mit in die Studie aufgenommen. Die Amplifikate wurden zur kommerziellen Sequenzierung verschickt, die erhaltenen Sequenzen manuell kontrolliert, zusammengehörende Vorwärts- und Rückwärtssequenzen zusammengefügt, die Sequenzen eines Locus auf gleiche Länge zugeschnitten, die vier Gen-Fragmente eines Isolats zu einer Sequenz verkettet und alle 311 Sequenzen mittels der Programme „MEGA“ und dem „MUSCLE“- Algorithmus „aligned“ - ausgerichtet. Für die phylogenetische Analyse wurde mit dem Programm „MrBayes“ das Markov-Chain-Monte-Carlo-Verfahren angewendet. Mittels „MrBayes“ und „DnaSP“ konnten Details der einzelnen Loci berechnet und verglichen werden.

Die errechneten Phylogramme ließen erkennen, dass *rpoA* nicht zur Differenzierung verschiedener *Vibrio cholerae* herangezogen werden kann. Das Phylogramm basierend aus den Sequenzen der verketteten vier Gen-Fragmente schlüsselte die *V. cholerae* Population der drei untersuchten Habitate auf und deckte 41 unterschiedliche Haplotypen auf, wobei 40 % aller Isolate nur einem einzigen Haplotyp zugeordnet wurden und 60 % der Isolate 40 verschiedenen Haplotypen angehörten. Das Habitat Schilfgürtel wies die höchste Diversität an unterschiedlichen *Vibrio cholerae* auf. 20 der aufgedeckten Haplotypen waren einzigartig für diesen Standort und nur ein einziger Haplotyp war spezifisch für den Probenpunkt „Intermediäres Habitat“ und keiner für den Probenpunkt „Offener See“. Die verbleibenden 20 unterschiedlichen *V. cholerae* Haplotypen kamen an allen drei Probenstellen vor. Manche miteinbezogene klinische *V. cholerae* Isolate wurden zu den Seeisolaten gruppiert und zeigten so, dass die krankheitsauslösenden *V. cholerae* Stämme aus früheren Jahren (2008-2011) 2012 noch immer im See vorhanden waren. Beim Vergleich des erstellten Multi-Locus-Phylogramms mit den Einzel-Loci-Phylogrammen wird deutlich, dass die Genfragmente *toxR*, *recA* und *gyrB* den meisten Einfluss hatten und außerdem einige deutlich weiter entfernte, unterschiedlichere, Isolat-Konstellationen bewirkten, die im Multi-Locus Phylogramm übernommen wurden.

Um die innerartliche Vielfalt der *V. cholerae* im See abzuschätzen, wurden „Rarefaction“-Kurven erstellt, welche ergaben, dass die Zahl der analysierten Isolate ausreichte um die *Vibrio cholerae* Population in drei unterschiedlichen Habitaten des Neusiedler Sees zu erfassen.

7 SUPPLEMENTARY

Table S 1. Different environmental parameters of Lake Neusiedler See, summer 2012.

Date	Sampling site	Water temp. [°C]	CFU *V.c.* [CFU/ml]		SPC *V.c.* [cells/ml]		Total bacterial cell nr		pH	Conductivity [µS/cm]	P_{tot} [µg/l]	Chl-a [µg/l]	DOC [mg/l]	Humics Ratio	TSS [mg/l]	Secchi [cm]
			MW	STABW	MW	STABW	[cells/ml]	SE								
11.06.2012	Intermediate habitat	19.6	6.1	1.0	5.5	4.1	6.10E+06	7.02E+05	8.9	2200	30.0	5.8	17.8		40.0	28.0
11.06.2012	Open water	19.5	0.5	0.1	2.8	3.5	1.58E+07	5.85E+05	8.9	2000	24.0	6.2	15.5		40.0	28.0
05.06.2012	Reed belt	17.6	62.3	8.0	226.4	51.6	2.08E+07	2.15E+06	8.6	2200	56.0	17.5	23.1	5.8	35.3	32.0
18.06.2012	Intermediate habitat	26.0	10.9	1.4	11.5	3.2	8.26E+06	1.01E+06	8.9	2000	21.0	4.7	19.2	14.2	30.0	38.0
18.06.2012	Open water	25.3	7.5	0.6	8.6	3.0	1.04E+07	8.37E+05	8.8	2000	61.0	22.6	15.2	15.9	60.6	18.0
18.06.2012	Reed belt	27.6	58.0	15.5	169.3	22.9	1.78E+07	1.68E+06	8.6	2600	45.0	7.1	35.8	9.9	34.0	33.0
02.07.2012	Intermediate habitat	27.0	46.3	4.0	24.9	5.1	1.20E+07	1.10E+06	8.9	2400	31.0	6.6	17.6	15.1	48.1	23.0
02.07.2012	Open water	27.5	11.7	1.2	3.8	3.8	1.56E+07	1.56E+06	8.8	2200	38.0	6.9	14.5	15.1	50.2	22.0
02.07.2012	Reed belt	28.0	84.0	24.3	160.0	39.2	2.03E+07	1.19E+06	8.7	2600	45.0	9.8	27.7	10.2	27.3	42.0
16.07.2012	Intermediate habitat						no sampling									
16.07.2012	Open water	18.1	22.4	4.2	41.1	14.5	6.70E+06	2.31E+05	8.9	1900	100.0	29.5	14.7	19.5	326.9	3.0
16.07.2012	Reed belt	18.3	47.0	7.2	79.1	20.8	1.26E+07	8.53E+05	8.7	2300	41.0	13.3	24.8	9.3	21.2	55.0
30.07.2012	Intermediate habitat	23.6	57.3	7.0	86.7	26.4	9.29E+06	7.71E+05	8.9	2100	82.0	20.8	15.8	15.0	105.3	10.0
30.07.2012	Open water	23.5	39.7	3.2	42.6	19.5	5.94E+06	6.54E+05	8.9	2000	99.0	15.3	13.1	17.9	105.3	10.0
30.07.2012	Reed belt	24.4	27.3	10.7	72.9	31.2			8.7	2300	48.0	14.2	23.9	7.7	21.9	53.0
06.08.2012	Intermediate habitat	26.4	46.0	18.4	40.3	9.6	8.53E+06	4.82E+05	8.8	1800	16.0	4.9	18.6	18.4	35.3	32.0
06.08.2012	Open water	27.0	21.3	5.7	30.5	16.6	8.53E+06	6.07E+05	8.8	1700	22.0	6.0	14.7	8.3	27.3	42.0
06.08.2012	Reed belt	28.3	30.3	12.5	57.5	25.6	1.24E+07	1.14E+06	8.7	1900	45.0	12.9	23.9	9.7	16.5	72.0
28.08.2012	Intermediate habitat	19.4	26.4	13.2	50.3	3.1	1.10E+07	8.93E+05	8.8	2600	132.0		16.1	6.5	40.0	17.0
28.08.2012	Open water	19.5	47.0	17.1	54.0	11.3	8.32E+06	8.18E+05	8.8	2500	122.0		13.5		34.3	15.0
28.08.2012	Reed belt	20.2	42.7	3.8	241.6	42.7	8.42E+06	8.06E+05	8.7	2700			31.5	6.4	14.9	80.0
10.09.2012	Intermediate habitat	21.3	90.0	2.6	83.0	11.5	1.16E+07	1.58E+06	8.9	2500	56.0	6.4	16.8	21.7	105.3	10.0
10.09.2012	Open water	21.8	54.0	4.4	53.0	6.0	8.64E+06	1.23E+06	8.9	2100	44.0	11.5	14.0		50.2	14.0
10.09.2012	Reed belt	22.8	59.7	6.0	169.9	44.0	1.31E+07	1.12E+06	8.8	2600	35.0	8.9	27.9	11.0	18.7	21.0
	Intermediate habitat	23.3	40.4		43.2		9.54E+06		8.9	2229	52.6	8.2	17.4	15.1	57.7	22.6
MW	Open water	22.8	25.5		29.6		9.99E+06		8.9	2050	63.8	14.0	14.4	15.3	86.8	19.0
	Reed belt	23.4	51.4		147.1		1.51E+07		8.7	2400	45.0	12.0	27.3	8.7	23.7	48.5
	Intermediate habitat	3.3	29.0		32.4		2.12E+06		0.0	287	41.8	6.2	1.3	5.1	33.0	10.8
SD	Open water	3.6	19.4		21.6		3.77E+06		0.1	233	38.4	9.1	0.8	4.3	99.9	11.9
	Reed belt	4.4	18.6		70.3		4.64E+06		0.1	273	6.4	3.6	4.4	1.9	7.7	20.4

Table S 2. Distribution of OTU's in three ecosystems of Lake Neusiedler See in July 2012. Continued for Maximum Parsimony tree.

Haplotype ID	Locus	Intermediate habitat	Open water	Reed belt	Total this study	References	Patients	Total
146	concatenated	54	61	6	121			121
147	concatenated	2	2	2	6			6
148	concatenated	5			5			5
149	concatenated			2	2			2
150	concatenated	4	4	2	10			10
151	concatenated	6	6	4	16			16
152	concatenated			2	2	1		3
153	concatenated	2	3	1	6			6
154	concatenated			3	3			3
155	concatenated						1	1
156	concatenated						1	1
157	concatenated						1	1
158	concatenated			1	1			1
159	concatenated			1	1			1
160	concatenated			7	7			7
161	concatenated	7	3	1	11			11
162	concatenated	1	1		2			2
163	concatenated			3	3			3
164	concatenated			11	11			11
165	concatenated						1	1
166	concatenated	2	2	10	14			14
167	concatenated			1	1			1
168	concatenated					3	1	4
169	concatenated	4	4		8			8
170	concatenated	1	1		2			2
171	concatenated			10	10			10
172	concatenated			9	9			9
173	concatenated	2	1	10	13		2	15
174	concatenated		2	5	7			7
175	concatenated	1	1	6	8			8
176	concatenated						1	1
177	concatenated	9	8	3	20			20

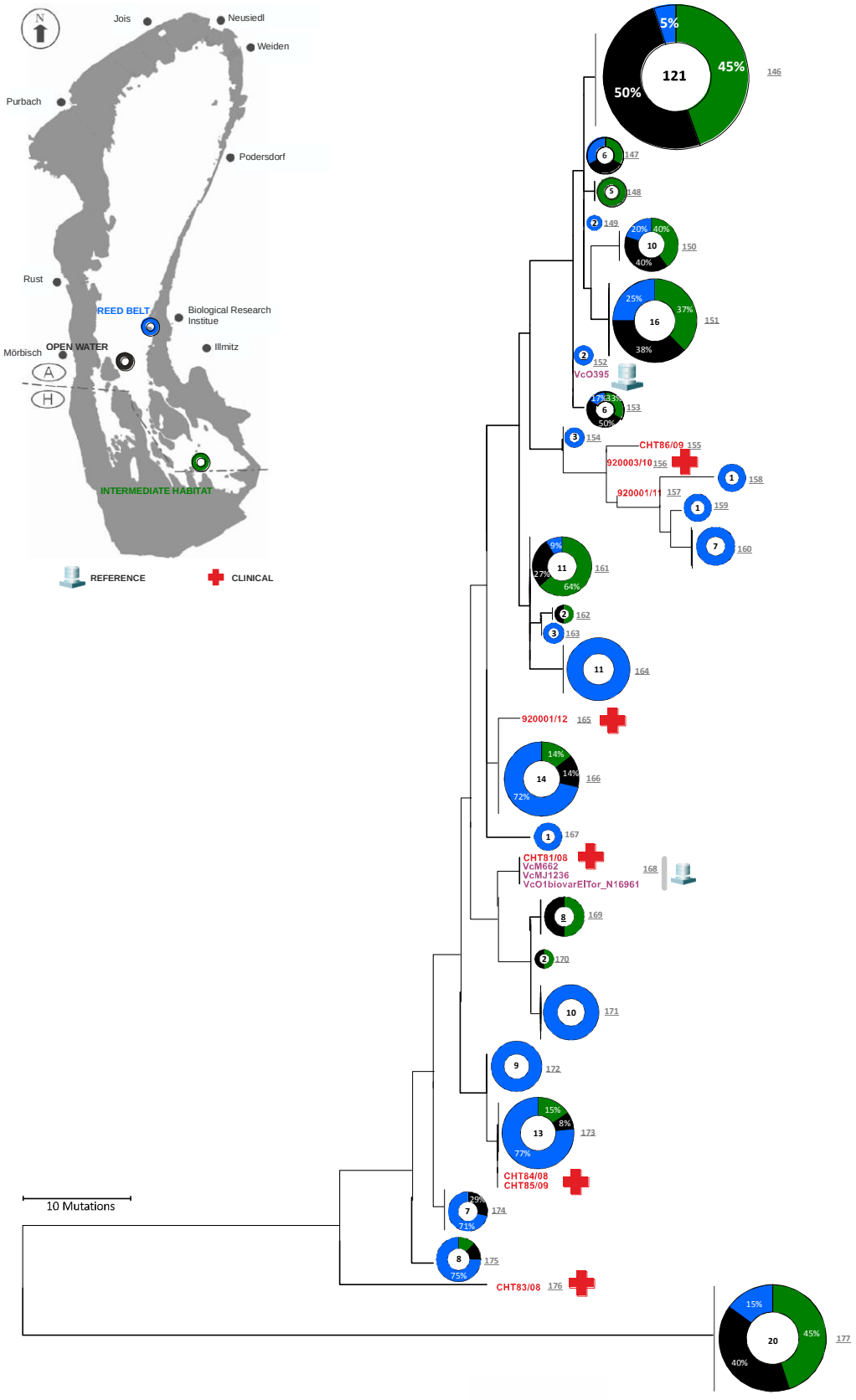


Figure S 1. Maximum Parsimony multilocus tree of *V. cholerae* isolates of Lake Neusiedler See.

Maximum parsimony tree resulting from analysis of the concatenated sequences comprising of *gyrB, recA, pyrH* and *toxR* [2,270 bp] of 299 *Vibrio cholerae* isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Haplotype ID's are given in grey colored numbers, according to **Table S 2**. Piecharts indicate the distribution of each clade among sampling sites, numbers of isolates per sampling site declared in percent and the total number of isolates is centered. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

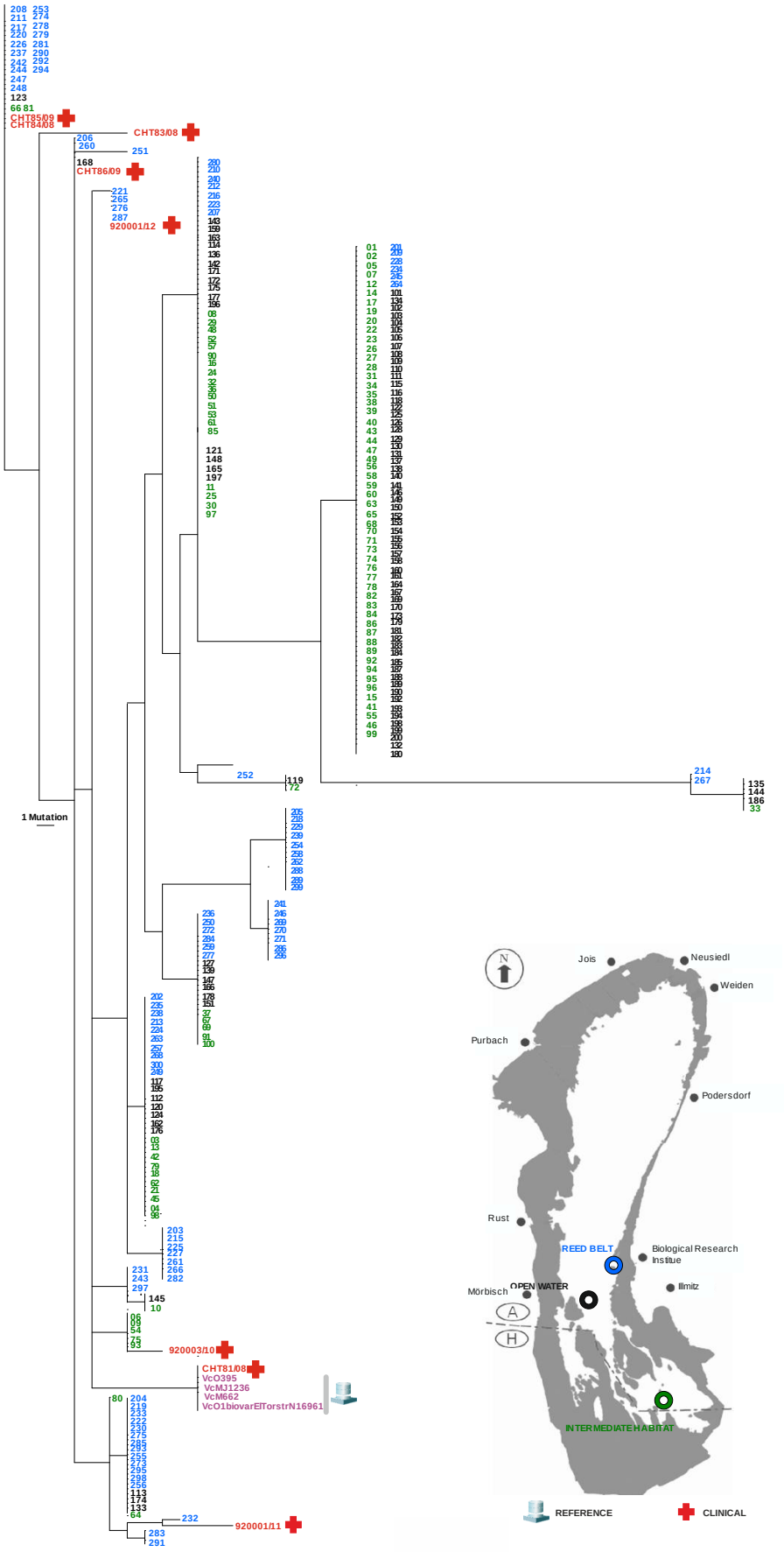


Figure S 2. Maximum Parsimony single loci tree: *gyrB*
1 out of 100 possible Maximum Parsimony trees resulting from analysis of the *gyrB* sequence fragment of 299 isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

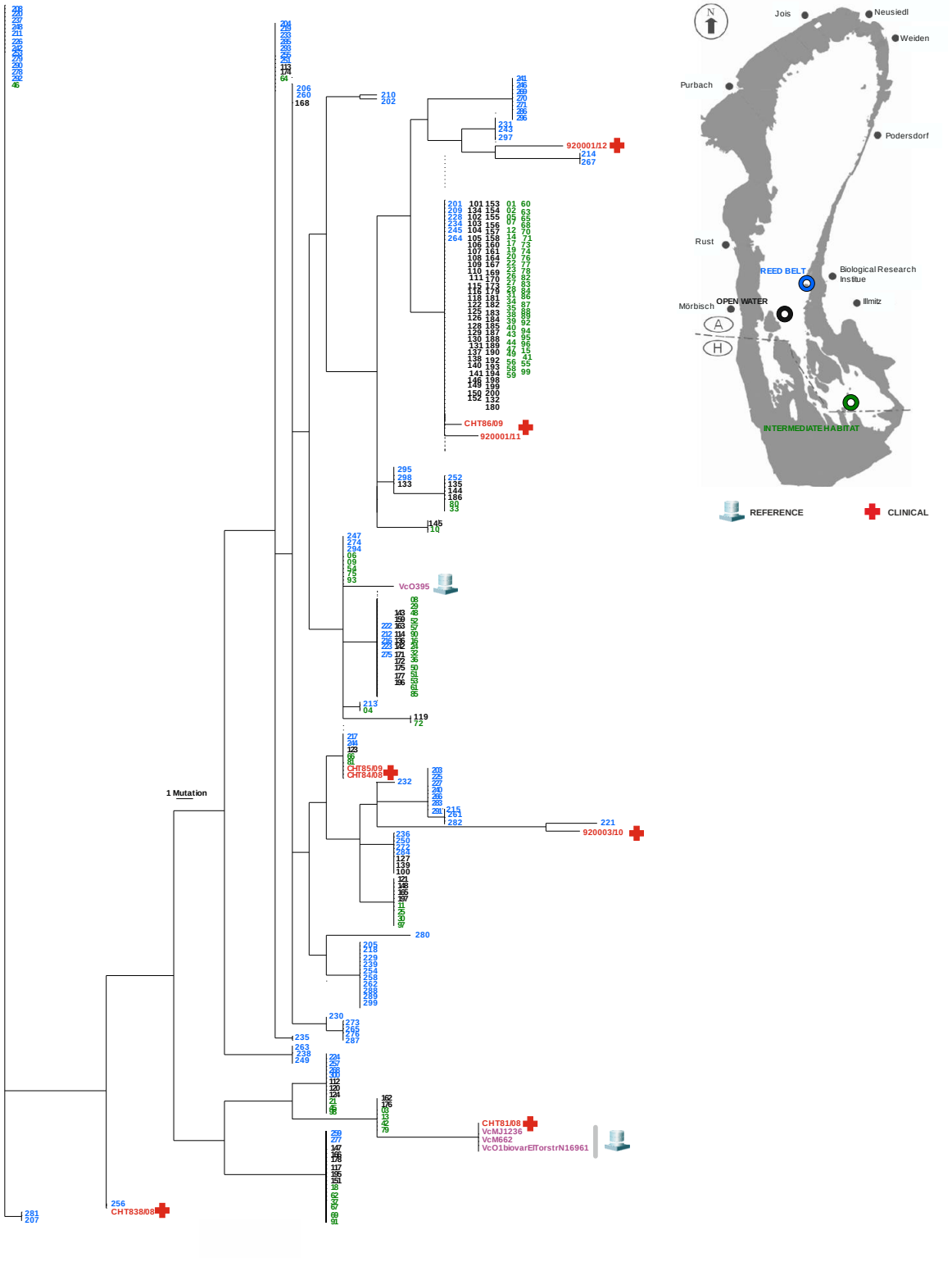


Figure S 3. Maximum Parsimony single loci tree: *recA*

1 out of 100 possible Maximum Parsimony trees resulting from analysis of the *recA* sequence fragment of 299 isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

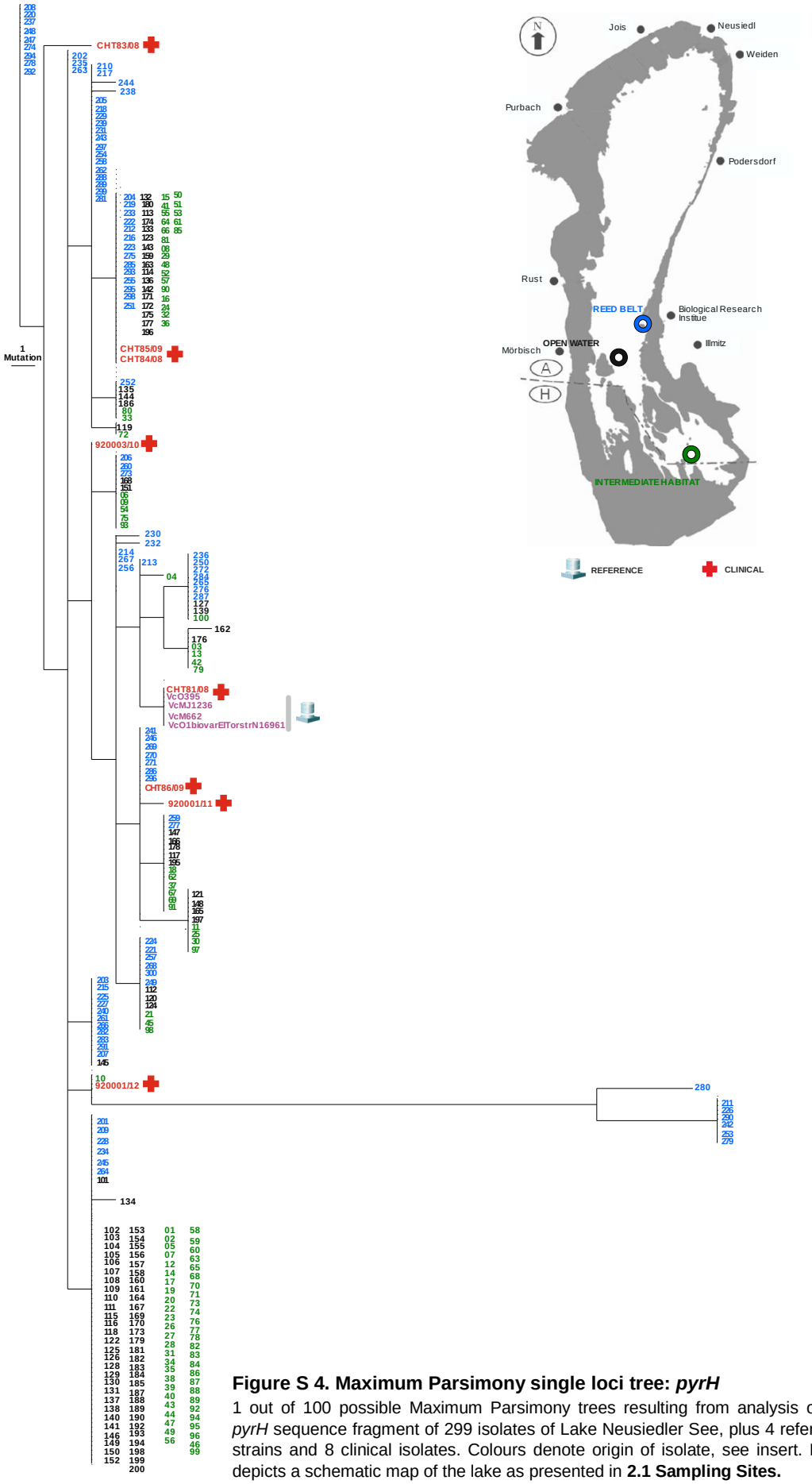


Figure S 4. Maximum Parsimony single loci tree: *pyrH*

1 out of 100 possible Maximum Parsimony trees resulting from analysis of the *pyrH* sequence fragment of 299 isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

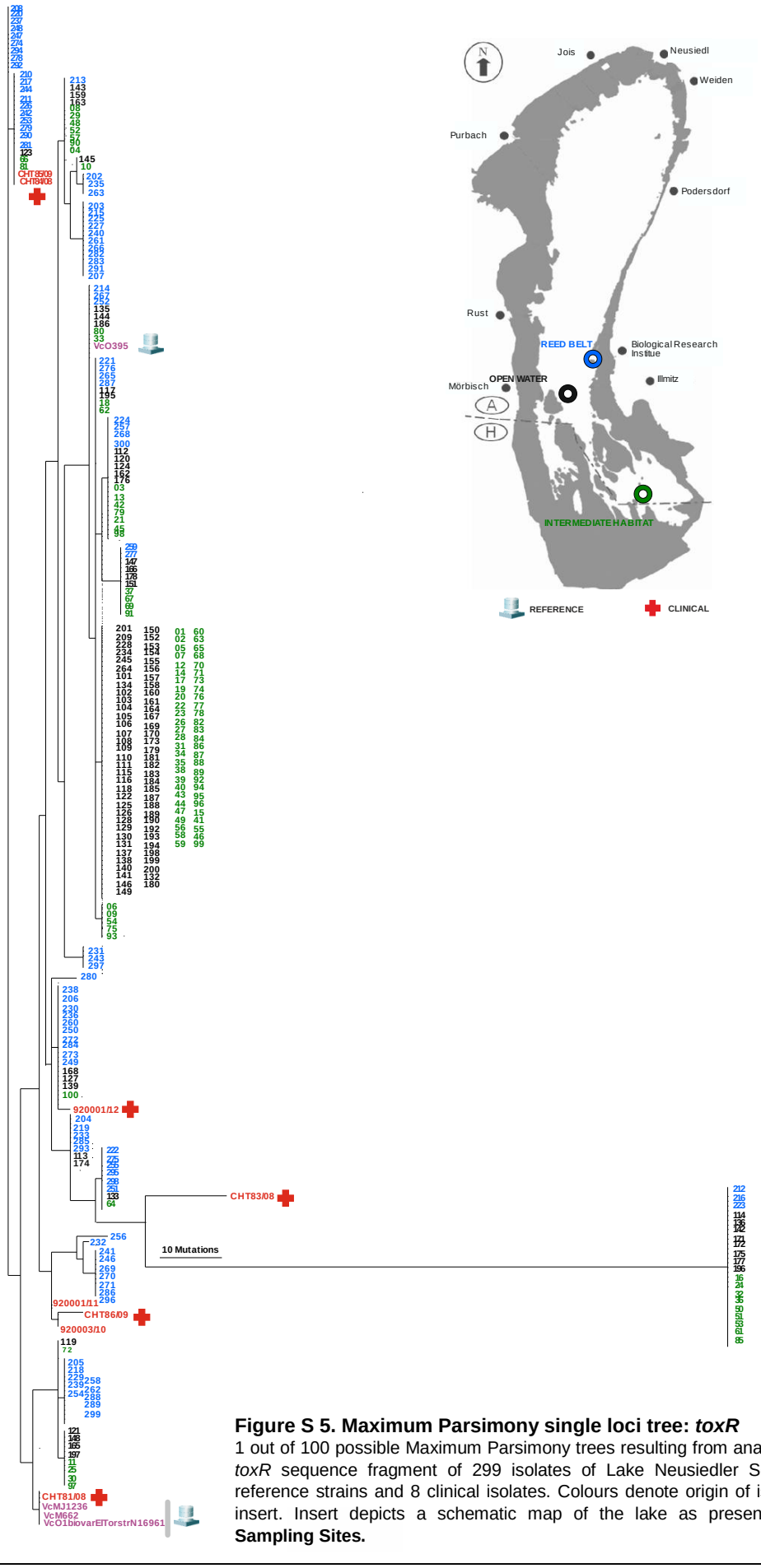


Figure S 5. Maximum Parsimony single loci tree: *toxR*
1 out of 100 possible Maximum Parsimony trees resulting from analysis of the *toxR* sequence fragment of 299 isolates of Lake Neusiedler See, plus 4 reference strains and 8 clinical isolates. Colours denote origin of isolate, see insert. Insert depicts a schematic map of the lake as presented in **2.1 Sampling Sites.**

7.1 RECIPES

7.1.1 MEDIA

Nutrient Agar without NaCl, [1L]

3.0 g	Meat extract (Lot: VM545479, 1.03979.0500, Merck KGaA, Darmstadt, Germany)
5.0 g	Meat Peptone (Art Nr. 2366.3, Charge 069103465, Roth, Karlsruhe, Germany)
15.0 g	Agar (Agar Bacteriological, LP0011, Agar NO.1, Lot 1168585-02, Oxoid, Hampshire, GB)

→ Adjust volume to 1L with dH_2O. → Sterilize by autoclaving

TCBS Agar, [1L]

VWR International GmbH, Vienna, Austria.
Recipe source: http://www.bd.com/ds/productCenter/221872.asp
According to manufacturer´s instructions: Quality control procedures: BBL, TCBS Agar, L007414. Rev. 07. October 2006

5.0 g	Yeast Extract	20.0 g	Sucrose
5.0 g	Pancreatic Digest of Casein	10.0 g	Sodium Chloride
5.0 g	Peptic Digest of Animal Tissue	1.0 g	Ferric Citrate
10.0 g	Sodium Citrate	0.04 g	Thymol Blue
10.0 g	Sodium Thiosulfate	0.04 g	Bromthymol Blue
5.0 g	Oxgall. dehydrated	14.0 g	Agar
3.0 g	Sodium Cholate		

→ Adjust volume to 1L with dH_2O

7.1.2 GEL-ELECTROPHORESIS

Agarose gel, [2 %, 150 ml]

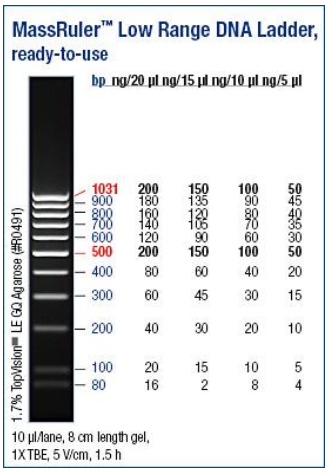
3.0 g	Agarose ultra Pure (Lot: 15510-027, Life Technologies, Paisley, Scotland)

→ Adjust volume to 150 ml with 1×TAE

Marker

MassRuler Low Range DNA ladder,

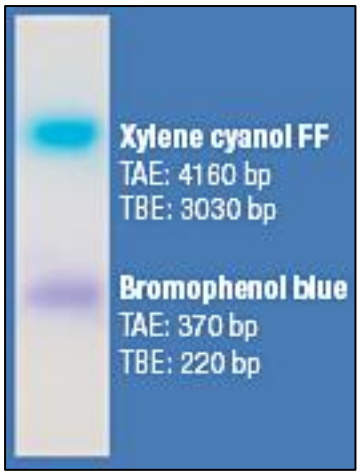
Fermentas.



Loading dye

6× loading dye solution,

Fermentas.



8 REFERENCES

8.1 REFERENCES

1. **Almeida LA, Araujo R.** 2013. Highlights on molecular identification of closely related species. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **13:**67-75.
2. **Bik EM, Bunschoten AE, Gouw RD, Mooi FR.** 1995. Genesis of the novel epidemic Vibrio cholerae O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. The EMBO journal **14:**209-216.
3. **Bik EM, Gouw RD, Mooi FR.** 1996. DNA fingerprinting of Vibrio cholerae strains with a novel insertion sequence element: a tool to identify epidemic strains. Journal of clinical microbiology **34:**1453-1461.
4. **Bina J, Zhu J, Dziejman M, Faruque S, Calderwood S, Mekalanos J.** 2003. ToxR regulon of Vibrio cholerae and its expression in vibrios shed by cholera patients. Proceedings of the National Academy of Sciences of the United States of America **100:**2801-2806.
5. **Bliem.** 2009. Invasion Potential of *Vibrio cholerae* O1/O139 in Lake Neusiedler See and Three Alkaline Saline Lakes. Diploma thesis. Medical University Vienna, Vienna.
6. **Butte AJ, Dzau VJ, Glueck SB.** 2001. Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues". Physiological genomics **7:**95-96.
7. **Cheasty T, Said B, Threlfall EJ.** 1999. V cholerae non-O1: implications for man? Lancet **354:**89-90.
8. **Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, Taviani E, Jeon YS, Kim DW, Lee JH, Brettin TS, Bruce DC, Challacombe JF, Detter JC, Han CS, Munk AC, Chertkov O, Meincke L, Saunders E, Walters RA, Huq A, Nair GB, Colwell RR.** 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. Proceedings of the National Academy of Sciences of the United States of America **106:**15442-15447.
9. **Chun J, Huq A, Colwell RR.** 1999. Analysis of 16S-23S rRNA intergenic spacer regions of Vibrio cholerae and Vibrio mimicus. Applied and environmental microbiology **65:**2202-2208.
10. **Collins MD, Ash C, Farrow JA, Wallbanks S, Williams AM.** 1989. 16S ribosomal ribonucleic acid sequence analyses of lactococci and related taxa. Description of Vagococcus fluvialis gen. nov., sp. nov. The Journal of applied bacteriology **67:**453-460.
11. **Colwell RK.** 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9 and earlier. User's Guide and application. Published at: http://purl.oclc.org/estimates.
12. **Colwell RR, Kaper J, Joseph SW.** 1977. Vibrio cholerae, Vibrio parahaemolyticus, and other vibrios: occurrence and distribution in Chesapeake Bay. Science (New York, N.Y.) **198:**394-396.
13. **Comstock LE, Johnson JA, Michalski JM, Morris JG, Jr., Kaper JB.** 1996. Cloning and sequence of a region encoding a surface polysaccharide of Vibrio cholerae O139 and characterization of the insertion site in the chromosome of Vibrio cholerae O1. Molecular microbiology **19:**815-826.
14. **Cooper JE, Feil EJ.** 2004. Multilocus sequence typing--what is resolved? Trends in microbiology **12:**373-377.
15. **Cottingham KL, Chiavelli DA, Taylor RK.** 2003. Environmental microbe and human pathogen: the ecology and microbiology of Vibrio cholerae. Frontiers in Ecology and the Environment **1:**80-86.
16. **Cox MM.** 2003. The bacterial RecA protein as a motor protein. Annual Review of Microbiology **57:**551-577.

17. **Druzhinina IS, Kubicek CP, Komon-Zelazowska M, Mulaw TB, Bissett J.** 2010. The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. BMC evolutionary biology **10:**94.

18. **Ebright RH, Busby S.** 1995. The Escherichia coli RNA polymerase alpha subunit: structure and function. Current opinion in genetics & development **5:**197-203.

19. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics **5:**113.

20. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research **32:**1792-1797.

21. **Eisenberg E, Levanon EY.** 2003. Human housekeeping genes are compact. Trends in genetics : TIG **19:**362-365.

22. **Faruque SM, Albert MJ, Mekalanos JJ.** 1998. Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae. Microbiology and molecular biology reviews : MMBR **62:**1301-1314.

23. **Faruque SM, Chowdhury N, Kamruzzaman M, Dziejman M, Rahman MH, Sack DA, Nair GB, Mekalanos JJ.** 2004. Genetic diversity and virulence potential of environmental Vibrio cholerae population in a cholera-endemic area. Proceedings of the National Academy of Sciences of the United States of America **101:**2123-2128.

24. **Faruque SM, Mekalanos JJ.** 2003. Pathogenicity islands and phages in Vibrio cholerae evolution. Trends in microbiology **11:**505-510.

25. **Faruque SM, Nair GB.** 2008. Vibrio cholerae : genomics and molecular biology. Caister Academic Press, Norfolk.

26. **Feng L, Reeves PR, Lan R, Ren Y, Gao C, Zhou Z, Ren Y, Cheng J, Wang W, Wang J, Qian W, Li D, Wang L.** 2008. A recalibrated molecular clock and independent origins for the cholera pandemic clones. PloS one **3:**e4053.

27. **Garg P, Aydanian A, Smith D, J. Glenn M J, Nair GB, Stine OC.** 2003. Molecular epidemiology of O139 Vibrio cholerae: mutation, lateral gene transfer, and founder flush. Emerging infectious diseases **9:**810-814.

28. **Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J.** 2005. Opinion: Re-evaluating prokaryotic species. Nature reviews. Microbiology **3:**733-739.

29. **Gimplinger J.** 2011. Bestimmung der Diversiät von Vibrio cholerae Stämmen aus dem Neusiedler See mittels Pulsfeldgelelektrophorese (PFGE). Bakkalaureatsarbeit. Medical University Vienna, Vienna.

30. **Goel AK, Ponmariappan S, Kamboj DV, Singh L.** 2007. Single multiplex polymerase chain reaction for environmental surveillance of toxigenic-pathogenic O1 and non-O1 Vibrio cholerae. Folia microbiologica **52:**81-85.

31. **Guidolin A, Manning PA.** 1987. Genetics of Vibrio cholerae and its bacteriophages. Microbiological reviews **51:**285-298.

32. **Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM.** 2000. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature **406:**477-483.

33. **Heymann DL.** 2004. Control of Communicable Diseases Manual. American Public Health Association Publications.

34. **Hornick RB, Music SI, Wenzel R, Cash R, Libonati JP, Snyder MJ, Woodward TE.** 1971. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. Bulletin of the New York Academy of Medicine **47:**1181-1191.

35. **Huhulescu SI, Alexander; Feierl, Gebhard; Stoeger, Anna; Ruppitsch, Werner; Sarkar, Banwarial; Allerberger, Franz.** 2007. Occurrence of Vibrio cholerae serogroups other than O1 and O139 in Austria Wiener klinische Wochenschrift **119:**235-241.

36. **Informationszentrum-Neusiedler-See-Seewinkel.** Date Accessed 2012-12-14. http://www.nationalpark-neusiedlersee-seewinkel.at/.

37. **Jaspers E, Overmann J.** 2004. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. Applied and environmental microbiology **70:**4831-4839.

38. **Johnson JA, Salles CA, Panigrahi P, Albert MJ, Wright AC, Johnson RJ, Morris JG, Jr.** 1994. Vibrio cholerae O139 synonym bengal is closely related to Vibrio cholerae El Tor but has important differences. Infection and immunity **62:**2108-2110.

39. **Kaper JB, Morris JG, Jr., Levine MM.** 1995. Cholera. Clinical microbiology reviews **8:**48-86.

40. **Keymer DP, Boehm AB.** 2011. Recombination shapes the structure of an environmental Vibrio cholerae population. Applied and environmental microbiology **77:**537-544.

41. **Keymer DP, Lam LH, Boehm AB.** 2009. Biogeographic patterns in genomic diversity among a large collection of Vibrio cholerae isolates, p. 1658-1666, Applied and environmental microbiology, 2009/01/14 ed, vol. 75.

42. **Keymer DP, Miller MC, Schoolnik GK, Boehm AB.** 2007. Genomic and phenotypic diversity of coastal Vibrio cholerae strains is linked to environmental factors. Applied and environmental microbiology **73:**3705-3714.

43. **Khodursky AB, Peter BJ, Schmid MB, DeRisi J, Botstein D, Brown PO, Cozzarelli NR.** 2000. Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. Proceedings of the National Academy of Sciences of the United States of America **97:**9419-9424.

44. **Kirschner AKT, Farnleitner AH, Herzig A, Sommer R.** 2009. Vibrio cholerae im Neusiedler See: Ein Grund zur Beunruhigung? . Schriften Verein zur Verbreitung naturwissenschaftlicher Kenntnisse **147:**67-90.

45. **Kirschner AKT, Schauer S, Steinberger B, Wilhartitz I, Grim CJ, Huq A, Colwell RR, Herzig A, Sommer R.** 2011. Interaction of Vibrio cholerae non-O1/non-O139 with copepods, cladocerans and competing bacteria in the large alkaline lake Neusiedler See, Austria. Microbial ecology **61:**496-506.

46. **Kirschner AKT, Schlesinger J, Farnleitner AH, Hornek R, Suss B, Golda B, Herzig A, Reitner B.** 2008. Rapid growth of planktonic Vibrio cholerae non-O1/non-O139 strains in a large alkaline lake in Austria: dependence on temperature and dissolved organic carbon quality. Applied and environmental microbiology **74:**2004-2015.

47. **Kita-Tsukamoto K, Oyaizu H, Nanba K, Simidu U.** 1993. Phylogenetic relationships of marine bacteria, mainly members of the family Vibrionaceae, determined on the basis of 16S rRNA sequences. International journal of systematic bacteriology **43:**8-19.

48. **Knoop V, Müller K.** 2008. Gene und Stammbäume: Ein Handbuch Zur Molekularen Phylogenetik. Spektrum Akademischer Verlag GmbH.

49. **Kotetishvili M, Stine OC, Chen Y, Kreger A, Sulakvelidze A, Sozhamannan S, Morris JG, Jr.** 2003. Multilocus sequence typing has better discriminatory ability for typing Vibrio cholerae than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. Journal of clinical microbiology **41:**2191-2196.

50. **Leache AD, Reeder TW.** 2002. Molecular systematics of the Eastern Fence Lizard (Sceloporus undulatus): a comparison of Parsimony, Likelihood, and Bayesian approaches. Systematic biology **51:**44-68.

51. **Lee JH, Han KH, Choi SY, Lucas ME, Mondlane C, Ansaruzzaman M, Nair GB, Sack DA, von Seidlein L, Clemens JD, Song M, Chun J, Kim DW.** 2006. Multilocus sequence typing (MLST) analysis of Vibrio cholerae O1 El Tor isolates from Mozambique that harbour the classical CTX prophage. Journal of medical microbiology **55:**165-170.

52. **Li W, Raoult D, Fournier PE.** 2009. Bacterial strain typing in the genomic era. FEMS microbiology reviews **33:**892-916.

53. **Lloyd AT, Sharp PM.** 1993. Evolution of the Reca Gene and the Molecular Phylogeny of Bacteria. Journal of Molecular Evolution **37:**399-407.

54. **Madigan MT, Martinko JM.** 2006. Brock Biology of Microorganisms. Pearson Prentice Hall.

55. **Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Sciences of the United States of America **95:**3140-3145.

56. **Margos G, Castillo-Ramirez S, Hoen AG.** 2012. Phylogeography of Lyme borreliosis-group spirochetes and methicillin-resistant Staphylococcus aureus. Parasitology **139:**1952-1965.

57. **McLaughlin JC.** 1995. Vibrio. In: Manual of clinical microbiology. Pages 465-476. ASM Press.

58. **Mohapatra SS, Ramachandran D, Mantri CK, Colwell RR, Singh DV.** 2009. Determination of relationships among non-toxigenic Vibrio cholerae O1 biotype El Tor strains from housekeeping gene sequences and ribotype patterns. Research in microbiology **160:**57-62.

59. **Morris JG, Jr.** 1990. Non-O group 1 Vibrio cholerae: a look at the epidemiology of an occasional pathogen. Epidemiologic reviews **12:**179-191.

60. **NCBI.** Date Accessed 2012-2013. http://www.ncbi.nlm.nih.gov/. National Center for Biotechnology Information.

61. **Nesweda JM.** 2010. Zeitlich und räumlich hochaufgelöste Tagesgänge des Phytoplanktons in einem trüben Flachsee am Beispiel des Neusiedler Sees. Diploma thesis. Univ. Wien.

62. **Octavia S, Salim A, Kurniawan J, Lam C, Leung Q, Ahsan S, Reeves PR, Nair GB, Lan R.** 2013. Population structure and evolution of non-O1/non-O139 Vibrio cholerae by multilocus sequence typing. PloS one **8:**e65342.

63. **Oliveros JC.** 2007. VENNY. An interactive tool for comparing lists with Venn Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/.

64. **Riepl M, Schauer S, Knetsch S, Holzhammer E, Farnleitner AH, Sommer R, Kirschner AK.** 2011. Applicability of solid-phase cytometry and epifluorescence microscopy for rapid assessment of the microbiological quality of dialysis water. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association **26:**3640-3645.

65. **Rivera IN, Chun J, Huq A, Sack RB, Colwell RR.** 2001. Genotypes associated with virulence in environmental isolates of Vibrio cholerae. Applied and environmental microbiology **67:**2421-2429.

66. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-1574.

67. **Rozas J.** 2009. DNA sequence polymorphism analysis using DnaSP. Methods in molecular biology **537:**337-350.

68. **Ruimy R, Breittmayer V, Elbaze P, Lafay B, Boussemart O, Gauthier M, Christen R.** 1994. Phylogenetic analysis and assessment of the genera Vibrio, Photobacterium, Aeromonas, and Plesiomonas deduced from small-subunit rRNA sequences. International journal of systematic bacteriology **44:**416-426.

69. **Sack DA, Sack RB, Nair GB, Siddique AK.** 2004. Cholera. Lancet **363:**223-233.

70. **Sack DA, Tacket CO, Cohen MB, Sack RB, Losonsky GA, Shimko J, Nataro JP, Edelman R, Levine MM, Giannella RA, Schiff G, Lang D.** 1998. Validation of a volunteer model of cholera with frozen bacteria as the challenge. Infection and immunity **66:**1968-1972.

71. **Schauer S, Sommer R, Farnleitner AH, Kirschner AKT.** 2012. Rapid and sensitive quantification of Vibrio cholerae and Vibrio mimicus cells in water samples by use of catalyzed reporter deposition fluorescence in situ hybridization combined with solid-phase cytometry. Applied and environmental microbiology **78:**7369-7375.

72. **Schuster BM, Tyzik AL, Donner RA, Striplin MJ, Almagro-Moreno S, Jones SH, Cooper VS, Whistler CA.** 2011. Ecology and genetic structure of a northern temperate Vibrio cholerae population related to toxigenic isolates. Applied and environmental microbiology **77:**7568-7575.

73. **Singleton FL, Attwell RW, Jangi MS, Colwell RR.** 1982. Influence of Salinity and Organic Nutrient Concentration on Survival and Growth of Vibrio-Cholerae in Aquatic Microcosms. Applied and environmental microbiology **43:**1080-1085.

74. **Spratt BG, Maiden MC.** 1999. Bacterial population genetics, evolution and epidemiology. Philosophical transactions of the Royal Society of London. Series B, Biological sciences **354:**701-710.

75. **Swofford DL.** 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Sinauer Associates, Sunderland, MA.

76. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular biology and evolution **24:**1596-1599.

77. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular biology and evolution **28:**2731-2739.

78. **Teh CSJ, Chua KH, Thong KL.** 2011. Genetic variation analysis of Vibrio cholerae using multilocus sequencing typing and multi-virulence locus sequencing typing. Infection Genetics and Evolution **11:**1121-1128.

79. **Thompson CC, Freitas FS, Marin MA, Fonseca EL, Okeke IN, Vicente AC.** 2011. Vibrio cholerae O1 lineages driving cholera outbreaks during seventh cholera pandemic in Ghana. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **11:**1951-1956.

80. **Thompson CC, Thompson FL, Vandemeulebroecke K, Hoste B, Dawyndt P, Swings J.** 2004. Use of recA as an alternative phylogenetic marker in the family Vibrionaceae. International journal of systematic and evolutionary microbiology **54:**919-924.

81. **Thompson CC, Thompson FL, Vicente AC.** 2008. Identification of vibrio cholerae and vibrio mimicus by multilocus sequence analysis (MLSA). International journal of systematic and evolutionary microbiology **58:**617-621.

82. **Thompson CC, Vicente AC, Souza RC, Vasconcelos AT, Vesth T, Alves N, Jr., Ussery DW, Iida T, Thompson FL.** 2009. Genomic taxonomy of Vibrios. BMC evolutionary biology **9:**258.

83. **Thompson FL, Austin B, Swings JG, American Society for Microbiology.** 2006. The biology of vibrios. ASM Press, Washington, D.C.

84. **Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, Munn CB, Swings J.** 2005. Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. Applied and environmental microbiology **71:**5107-5115.

85. **Tu JL, Chin KH, Wang AH, Chou SH.** 2009. Unique GTP-binding pocket and allostery of uridylate kinase from a gram-negative phytopathogenic bacterium. Journal of molecular biology **385:**1113-1126.

86. **Waldor MK, Mekalanos JJ.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science (New York, N.Y.) **272:**1910-1914.

87. **Wang D, Wang H, Zhou Y, Zhang Q, Zhang F, Du P, Wang S, Chen C, Kan B.** 2011. Genome sequencing reveals unique mutations in characteristic metabolic pathways and the transfer of virulence genes between V. mimicus and V. cholerae. PloS one **6:**e21299.

88. **Watnick PI, Fullner KJ, Kolter R.** 1999. A role for the mannose-sensitive hemagglutinin in biofilm formation by Vibrio cholerae El Tor. Journal of Bacteriology **181:**3606-3609.

89. **Wilgenbusch JC, Warren DL, Swofford DL.** 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. http://ceb.csit.fsu.edu/awty.

90. **Woese CR.** 1987. Bacterial evolution. Microbiological reviews **51:**221-271.

91. **World-Health-Organization.** Date Accessed 2013-01-15. http://www.who.int/features/factfiles/cholera/facts/en/index1.html.

92. **World-Health-Organization.** Date Accessed 2013-02-20. http://www.who.int/wer/2010/wer8531.pdf.

93. **World-Health-Organization.** Date Accessed 2013-02-20. http://www.who.int/wer/2011/wer8631.pdf.

94. **World-Health-Organization.** Date Accessed 2013-02-20. http://www.who.int/wer/2012/wer8731_32.pdf.

95. **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL.** 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC bioinformatics **13:**134.

96. **Zeigler DR.** 2003. Gene sequences useful for predicting relatedness of whole genomes in bacteria. International journal of systematic and evolutionary microbiology **53:**1893-1900.

8.2 ADDITIONAL USED LITERATURE

Byun. R.. L. D. Elbourne. R. Lan. and P. R. Reeves. 1999. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. Infect Immun **67:**1116-1124.

Cameron. D. N.. F. M. Khambaty. I. K. Wachsmuth. R. V. Tauxe. and T. J. Barrett. 1994. Molecular characterization of *Vibrio cholerae* O1 strains by pulsed-field gel electrophoresis. J Clin Microbiol **32:**1685-1690.

Danin-Poleg. Y.. L. A. Cohen. H. Gancz. Y. Y. Broza. H. Goldshmidt. E. Malul. L. Valinsky. L. Lerner. M. Broza. and Y. Kashi. 2007. *Vibrio cholerae* strain typing and Phylogeny study based on simple sequence repeats. J Clin Microbiol **45:**736-746.

Dashtbani-Roozbehani. A.. B. Bakhshi. and M. R. Pourshafie. 2013. Genetic Relatedness of Clinical and Environmental *Vibrio cholerae* Isolates Based on Triple Housekeeping Gene Analysis. Curr Microbiol **67:**15-20.

Sawabe. T.. K. Kita-Tsukamoto. and F. L. Thompson. 2007. Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. J Bacteriol **189:**7932-7936.

Sozhamannan. S.. Y. K. Deng. M. Li. A. Sulakvelidze. J. B. Kaper. J. A. Johnson. G. B. Nair. and J. G. Morris. Jr. 1999. Cloning and sequencing of the genes downstream of the wbf gene cluster of *Vibrio cholerae* serogroup O139 and analysis of the junction genes in other serogroups. Infect Immun **67:**5033-5040.

Stine. O. C.. S. Sozhamannan. Q. Gou. S. Zheng. J. G. Morris. Jr.. and J. A. Johnson. 2000. Phylogeny of *Vibrio cholerae* based on recA sequence. Infect Immun **68:**7180-7185.

Teh. C. S.. K. H. Chua. and K. L. Thong. 2010. Multiple-locus variable-number tandem repeat analysis of *Vibrio cholerae* in comparison with pulsed field gel electrophoresis and virulotyping. J Biomed Biotechnol **2010:**817190.

“Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung der Bilder in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.“

CURRICULUM VITAE

CARINA PRETZER

ADRESS:	Kollmayergasse 18/10, 1120 Vienna, Austria.
PLACE AND DATE OF BIRTH:	St.Pölten, 09.04.1987
CITIZENSHIP:	Austria

EDUCATION

Aug. 2012 – Sept. 2013	Diploma thesis: „Establishment of a Suitable Multilocus Sequence Analysis Procedure for Determining the Diversity of *Vibrio cholerae* Strains in Lake Neusiedler See“ Medical University of Vienna and University of Vienna
2006-2013	Diploma study in biology/ genetics-microbiology. University of Vienna. Austria
2001-2006	“HLA für Umwelt und Wirtschaft im Yspertal” Final examination: 20.06.2006
1999-2001	High school, Rabenstein/Pielach, Austria
1997-1999	High school, Grünau, Austria
1993-1997	Primary school, Rabenstein/Pielach, Austria

EXPERIENCE

Sept. 2011	Univ.-Prof. Dr.med.univ. Rudolf Valenta, Medical University Vienna Course: *“Molecular Medicine – Expression of recombinant cow’s milk proteins for a customized Microarray”*
July 2011	Ass.-Prof.Dipl.-Biol. Dr. Alexander Loy, University of Vienna Course: *“DNA Microarrays - Hochparallele molekulare Analyse von Mikroorganismen”*
May – June 2011	Ass.-Prof.Dipl.-Biol. Dr. Alexander Loy, University of Vienna Course: *“Growth curve of a complex microbial community in natural mineral water after bottling”*
May 2010	Privatdoz. Dipl.-Biol. Dr. Stefanie Wienkoop, University of Vienna, Course: *“Metabolomics - Changes in metabolic profile of Medicago truncatula after inoculation with Sinorhizobium melioti”*

LANGUAGE

German	mother tongue
English	fluent
French	basic level

PERSONAL SKILLS AND COMPETENCES:

Microbiological methods:	PCR, FISH, Micro array, ELISA, Protein expression & purification
Computer knowledge:	MS Office, CorelDRAW. MEGA, CLC Main Workbench, EstimateS, DnaSP, TreeView.