



Fakultät Wissenschaftszentrum Weihenstephan  
für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Mikrobiologie

Entwicklung von *Thermus thermophilus* als Wirt für  
(Meta)Genomanalysen

Benedikt Leis

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:

Univ.-Prof. Dr. E. Grill

Prüfer der Dissertation:

1. Univ.-Prof. Dr. W. Liebl
2. Univ.-Prof. Dr. V. Sieber

Die Dissertation wurde am 10.06.2014 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 14.09.2014 angenommen.



*Für meine Familie*



## Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abkürzungsverzeichnis.....	V
Abbildungsverzeichnis.....	VIII
Tabellenverzeichnis.....	X
I. Einleitung.....	1
1. Die Metagenom-Analyse.....	1
1.1. Die Vielfalt mikrobiellen Lebens.....	1
1.2. Metagenomik als Quelle neuer Biokatalysatoren für industrielle Prozesse.....	2
1.3. Technologien und Anwendungen in der modernen Metagenomik.....	3
1.4. Neue Wirte für metagenomische Durchmusterungen.....	6
2. Der thermophile Modellorganismus <i>Thermus thermophilus</i> HB27.....	7
2.1. Regulation der Carotinoid-Biosynthese: das Car-Cluster.....	8
2.2. <i>T. thermophilus</i> als Expressionswirt für Genbanken.....	9
II. Zielsetzung der Arbeit.....	10
III. Materialien und Methoden.....	11
1. Verwendete Stämme und Wachstumsbedingungen.....	11
1.1. Stämme und Plasmide.....	11
1.2. Verwendete Medien.....	15
1.3. Anzuchtbedingungen, Zellernte, Gewinnung von Zellextrakten.....	17
2. Molekulargenetische Methoden.....	18
2.1. Nukleinsäureanalytik.....	18
2.1.1. Isolation und Aufreinigung von (meta)genomischen Nukleinsäuren und Plasmiden.....	18
2.1.2. Gelelektrophorese.....	20
2.1.3. DNA-Quantifizierung.....	20
2.2. Enzymatische Modifikationen.....	21
2.2.1. Polymerase-Kettenreaktion (PCR).....	21
2.2.2. Restriktionsverdau.....	22
2.2.3. Dephosphorylierung und <i>blunting</i> .....	22
2.2.4. Ligationen.....	22
2.3. Transformation von Bakterien.....	23
2.3.1. Chemische Transformation von <i>E. coli</i> .....	23
2.3.2. Elektroporation von <i>E. coli</i> .....	24
2.3.3. Natürliche Transformation von <i>T. thermophilus</i> HB27.....	25
2.4. Strategien für genetische Manipulationen in <i>T. thermophilus</i> HB27.....	25

2.4.1. Transposon-Mutagenese .....	25
2.4.2. Gezielte Genunterbrechungen .....	25
2.4.3. Genaustauschmutanten und Deletionen.....	26
2.4.4. Gendeletionen ohne verbleibenden Marker .....	27
2.5. Southern Blot-Analyse .....	29
3. Konstruktion und funktionelle Analyse von Genbanken.....	30
3.1. Verwendete Genbanken und Umweltproben.....	30
3.2. Erstellung von Fosmidbanken aus (meta)genomischer DNA.....	31
3.3. Erstellung von Plasmid-Genbanken .....	32
3.3.1. Subklonieren von Fosmid-DNA.....	32
3.3.2. 16S rDNA Sequenzen.....	32
3.4. Funktionelle Analyse und Aktivitätsnachweise der Genbanken .....	32
3.4.1. Übertragung der Fosmidbanken in <i>T. thermophilus</i> .....	32
3.4.2. Direkte Detektion .....	33
3.4.3. Heterologe Komplementationsanalyse .....	35
4. Sequenzierungen und bioinformatische Analyse.....	36
4.1. Sanger-Sequenzierung und <i>Next-Generation Pyrosequencing</i> (454-NGS).....	36
4.2. Prozessierung und Analyse der Sequenzen .....	37
5. Proteinbiochemische Analytik.....	38
5.1. Allgemeine Methoden .....	38
5.1.1. Quantifizierung der Proteinkonzentration .....	38
5.1.2. Aufkonzentrieren und Dialyse von Protein .....	38
5.1.3. Formulierung der Proteinlösungen .....	38
5.2. Polyacrylamid-Gelelektrophorese .....	39
5.3. Membrantransfer und Immunodetektion von Protein (Western Blot).....	39
5.4. Proteinaufreinigung .....	41
5.4.1. Hitzefällung von Zellextrakten.....	41
5.4.2. Aufreinigung mittels Fast Protein Liquid Chromatography (FPLC).....	41
5.4.3. Aufreinigung mittels immobilisierter Metallionen-Affinitätschromatographie .....	42
5.5. Biochemische Charakterisierungen .....	42
5.5.1. $\beta$ -Glucosidase Test .....	42
5.5.2. Bestimmung reduzierender Zuckerenden mit DNSA-Test.....	43
5.5.3. Qualitative Analyse der Hydrolyseprodukte durch Dünnschichtchromatographie .....	44
5.5.4. Zymogramm zur Detektion enzymatischen Abbaus von Polysacchariden .....	45
5.5.5. Bindestudien von Proteinen an unlösliche Polysaccharide.....	45
5.5.6. Bestimmung lipolytischer Aktivitäten.....	46

---

6. Datendokumentation, -verarbeitung und statistische Auswertungen .....	47
IV. Ergebnisse .....	48
1. Verbesserung des Wirt-Vektor-Systems in <i>Thermus</i> .....	48
1.1. Hintergrundaktivitäten in <i>T. thermophilus</i> HB27.....	48
1.2. Entwicklung eines licht-induzierbaren Expressionssystems auf Basis des Car-Clusters .....	49
1.2.1. Deletion des Car-Clusters ( $\Delta car$ ) in <i>Thermus</i> .....	50
1.2.2. Konstruktion von Expressionsvektoren auf Basis des Carotinoid-Clusters .....	52
1.2.3. Vergleich der Expressionsvektoren im Reportergenassay .....	54
1.3. Erstellung einer Transposonmutagenese-Bank in <i>T. thermophilus</i> HB27 .....	55
1.3.1. <i>In vivo</i> Transposonmutagenese .....	55
1.3.2. <i>In vitro</i> Transposonmutagenese .....	59
1.4. Deaktivierung der lipolytischen Hintergrundaktivität in <i>T. thermophilus</i> und Herstellung einer Multideletionsmutante BL03 als Expressionswirt für Genbanken.....	61
1.4.1. Deaktivierung ausgewählter Gene zur Verminderung der <i>T. thermophilus</i> -eigenen Esterase-Aktivität.....	61
1.4.2. Markerlose, multiple Gendelektionen in <i>Thermus</i> .....	64
1.4.3. Phänotypische Charakterisierungen der Multideletionsmutante BL03 .....	68
2. Identifizierung und Charakterisierung von biotechnologisch relevanten Enzymen aus diversen metagenomischen Fosmidbanken in <i>E. coli</i> .....	71
2.1. Durchmusterung psychrophiler, mesophiler und thermophiler Metagenombanken .....	71
2.2. Sequenzierung und <i>in silico</i> Analyse der aktiven Fosmide.....	74
2.3. Charakterisierung einer neuartigen, archaeellen und hyperthermostabilen Endoglucanase Cel12E.....	78
2.3.1. Analyse des Fosmids HA-cmc-1 und des beteiligten ORFs .....	78
2.3.2. Heterologe Expression und Aufreinigung.....	80
2.3.3. Biochemische Charakterisierung von Cel12E.....	83
3. Vergleichende Durchmusterung und Analyse von Fosmid-Genbanken in <i>E. coli</i> und <i>T. thermophilus</i> HB27 .....	90
3.1. Analyse einer Genbank von <i>T. Brockianus</i> .....	90
3.1.1. Vergleichende Durchmusterung in beiden Wirten.....	90
3.1.2. Bioinformatische Analyse .....	92
3.2. Konstruktion einer metagenomischen Genbank aus heißen Quellen in <i>E. coli</i> .....	93
3.2.1. Analyse des Probenmaterials .....	93
3.2.2. Gewinnung von metagenomischer DNA .....	95
3.2.3. Erstellung der pCT3FK-Fosmidbank .....	97
3.2.4. Phylogenetische Analyse mittels 16S rDNA-Sequenzen .....	98
3.3. Funktionelle Durchmusterung thermophiler Metagenombanken.....	99
3.3.1. Übertragung der Fosmide in <i>T. thermophilus</i> .....	100

3.3.2.	Komplementation der Esterase-defizienten <i>T. thermophilus</i> Mutante BL03 .....	101
3.3.3.	Aktivität von Rohextrakten aus positiven Klonen auf pNP-Substraten .....	103
3.4.	Sequenzierung und <i>in silico</i> Analyse der Fosmide der positiven Klone des Screenings.....	104
3.5.	Expression und vorläufige Charakterisierung ausgewählter Esterasen .....	109
3.5.1.	Expression in <i>T. thermophilus</i> .....	109
3.5.2.	pET21a(+)-Expression in <i>E. coli</i> .....	112
3.5.3.	Biochemische Charakterisierung .....	114
V.	Diskussion .....	116
1.	Verbesserung des <i>Thermus</i> Wirt-Vektor-Systems.....	116
1.1.	Entwicklung eines licht-induzierbaren Expressionsystems auf Basis des Car-Genclusters....	116
1.2.	Erstellung einer Transposonmutagenese-Bank in <i>T. thermophilus</i> HB27.....	122
1.3.	Deaktivierung der lipolytischen Hintergrundaktivität in <i>Thermus</i> und Herstellung einer Multideletionsmutante BL03 als Expressionswirt für Genbanken .....	125
2.	Identifizierung und Charakterisierung von biotechnologisch relevanten Enzymen aus diversen metagenomischen Fosmidbanken in <i>E. coli</i> .....	130
2.1.	Funktionelle Durchmusterung von Metagenombanken.....	131
2.2.	Sequenzierung und bioinformatische Auswertung .....	135
2.2.1.	Assemblierung und <i>mapping</i> .....	135
2.2.2.	Identifizierung von Kandidatengenen .....	136
2.3.	Identifizierung und Charakterisierung einer archaeellen Endoglucanase Cel12E.....	141
2.3.1.	Durchmusterung und <i>in silico</i> -Analyse des Fosmids HA-cmc-1 .....	141
2.3.2.	Expression und biochemische Charakterisierung .....	144
3.	Vergleichende Durchmusterung thermophiler Metagenombanken in <i>E. coli</i> und <i>T. thermophilus</i> .....	147
3.1.	Analyse der Genbank von <i>T. Brockianus</i> .....	148
3.2.	Konstruktion einer metagenomischen Fosmidbank aus heißen Quellen in <i>E. coli</i> .....	149
3.3.	Funktionelle Durchmusterung thermophiler Metagenombanken in beiden Wirten .....	150
3.4.	Sequenzierung und <i>in silico</i> Analyse der Fosmide aus positiv getesteten Klonen .....	153
3.5.	Expression und vorläufige biochemische Charakterisierung ausgewählter Esterasen .....	159
4.	<i>T. thermophilus</i> als neuartiger Expressionswirt für Metagenomanalysen .....	162
VI.	Zusammenfassung .....	170
VII.	Im Rahmen der Dissertation entstandene Veröffentlichungen .....	172
VIII.	Literaturangaben .....	173
IX.	Anhang .....	185
X.	Danksagungen .....	216
XI.	Lebenslauf .....	217

## Abkürzungsverzeichnis

Δ	Unterbrechung, Deletion oder Austausch einer Gensequenz
°C	Grad Celsius
× g	Vielfaches der Erdbeschleunigung
μ-	Mikro- ( $1 \times 10^{-6}$ )
A	Ampere
BAC	<i>bacterial artificial chromosome</i>
BCI-	5-Bromo-4-Chloro-3-Indoxyl-
bp	Basenpaar
c-	Centi- ( $1 \times 10^{-2}$ )
C-terminal	carboxyterminal
ca.	circa
Car	Carotinoid-Cluster, Teil des Carotinoid-Biosynthese ( <i>crt</i> ) Clusters in <i>T. thermophilus</i> (kodierende <i>car</i> -Gene sind TT_P0055 bis TT_P0058)
CBM	Kohlenhydrat-Bindemodul
CMC	Carboxymethyl-Cellulose
CV	Säulenvolumen ( <i>column volume</i> )
d.h.	das heißt
Da	Dalton
ddH <sub>2</sub> O	doppelt demineralisiertes, deionisiertes Wasser
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DNSA	Dinitrosalicylsäure-Assay, kolorimetrisches Nachweisverfahren zur Bestimmung reduzierender Zuckerenden
dNTP	Desoxyribonukleotidtriphosphat
EC	Nummer der Enzym-Kommission
EDTA	Ethylendiamintetraacetat
<i>et al.</i>	und weitere ( <i>et alii</i> )
FPLC	<i>fast protein liquid chromatography</i>
g	Gramm
gDNA	genomische DNA
GC	Prozentualer Gehalt an Guanin und Cytosin in DNA: $(G+C)/(A+T+G+C) \times 100$
GO	<i>Gene ontology</i> , Vereinheitlichung der Annotationen von Genen und Genprodukten, Ashburner <i>et al.</i> (2000)
h	Stunde
His-tag	Hexahistidin(6xHis)-tag
IMAC	Immobilisierte Metallionen-Affinitäts-Chromatographie
IPTG	Isopropyl-β-D-thiogalactopyranosid
k-	Kilo- ( $1 \times 10^3$ )
<i>kat</i>	Kanamycin-Resistenzkassette, thermostabile Kanamycin-Adenyltransferase (Kat)
KBE	Kolonie-bildende Einheit(en)

---

L	Liter
LB	<i>lysogeny broth</i> (auch als Luria-Bertani Medium bekannt)
m-	Milli- ( $1 \times 10^{-3}$ )
M-	Mega- ( $1 \times 10^6$ )
m	Meter
M	Molar ( $\text{mol} \times \text{L}^{-1}$ )
MCS	Multiple Klonierungsstelle ( <i>multiple cloning site</i> )
MES	2-(N-Morpholino)ethansulfonsäure
min	Minute(n)
n-	Nano- ( $1 \times 10^{-9}$ )
N-terminal	aminoterminal
NCBI	<i>National Center for Biotechnology Information</i>
nt	Nukleotide
o.g.	oben genannt
OD	Optische Dichte
ORF	Offener Leserahmen ( <i>open reading frame</i> )
P <sub>x</sub>	Promotor des Gens X
PAGE	Polyacrylamid Gelelektrophorese
PCR	Polymerase-Kettenreaktion ( <i>polymerase chain reaction</i> )
pH	negativer dekadischer Logarithmus der Protonenkonzentration
pNP-	<i>para</i> -Nitrophenyl-
R	Resistenz
RBS	Ribosomale Bindestelle
RNA	Ribonukleinsäure
RT	Raumtemperatur
s	Sekunde
SDS	Natriumlaurylsulfat
SOE-PCR	<i>Splicing by overlap extension</i> PCR (Heckman und Pease, 2007)
sp.	Spezies
TB	<i>Thermus broth</i>
Tris	Tris-(hydroxymethyl)-aminomethan
U	Einheit der Enzymaktivität ( <i>Unit</i> )
u.a.	unter anderem
ÜN	über Nacht
Upm	Umdrehungen pro Minute
UV	Ultraviolett
V	Volt
VE	Vollentsalztes Wasser
v/v	Volumen pro Volumen
w/v	Gewicht pro Volumen

---

Abkürzungen für Nukleinsäuren:

A	Adenin
C	Cytosin
T	Thymin
G	Guanin

Abkürzungen für Aminosäuren:

A	Ala	Alanin
C	Cys	Cystein
D	Asp	Aspartat
E	Glu	Glutamat
F	Phe	Phenylalanin
G	Gly	Glycin
H	His	Histidin
I	Ile	Isoleucin
K	Lys	Lysin
L	Leu	Leucin
M	Met	Methionin
N	Asn	Asparagin
P	Pro	Prolin
Q	Gln	Glutamin
R	Arg	Arginin
S	Ser	Serin
T	Thr	Threonin
V	Val	Valin
W	Trp	Tryptophan
Y	Tyr	Tyrosin
X	Xaa	beliebige Aminosäure

## Abbildungsverzeichnis

Abbildung 1: Flussdiagramm mit den wesentlichen Schritten der Metagenomanalyse .....	5
Abbildung 2: Eigenschaften des Car-Clusters in <i>T. thermophilus</i> HB27 .....	8
Abbildung 3: Grundlegende Strategien zur Erstellung von gezielten genetischen Manipulationen zur Deaktivierung, zum Austausch oder zur vollständigen Deletion der Zielgene in <i>T. thermophilus</i> .....	28
Abbildung 4: Charakterisierung der markerlosen Deletionsmutante $\Delta car$ in <i>T. thermophilus</i> $\Delta bgl$ .....	51
Abbildung 5: Flussdiagramm zur Konstruktion der <i>Thermus</i> -Expressionsvektoren basierend auf dem Car-Cluster in <i>T. thermophilus</i> HB27 .....	53
Abbildung 6: Aktivitätsdaten der Expressionsplasmide in <i>T. thermophilus</i> .....	54
Abbildung 7: Schematische Darstellung der Konstruktion von pTnB2 und pTnC2 für <i>in vivo</i> Transposon-Mutagenese in <i>T. thermophilus</i> HB27 .....	56
Abbildung 8: Analyse der genomischen DNA zehn verschiedener Transposonmutanten in <i>T. thermophilus</i> HB27 aus einer <i>in vivo</i> Mutagenese mit pTnB2 .....	58
Abbildung 9: Analyse von Mutanten in <i>T. thermophilus</i> nach <i>in vitro</i> Transposonmutagenese.....	60
Abbildung 10: Charakterisierung einzelner Mutanten mit verringerter lipolytischer Aktivität .....	63
Abbildung 11: Zusammenfassung der wesentlichen Arbeitsschritte zur Erstellung von Markerlosen Gendeletionen in <i>T. thermophilus</i> .....	65
Abbildung 12: Visualisierung der Detektions-PCR zum Nachweis der Präsenz der Gendeletionsmutante $\Delta TT\_C1787$ in <i>T. thermophilus</i> .....	66
Abbildung 13: Schematische Darstellung zur Identifizierung der Gendeletionsmutanten im 96-Well Block mittels kombinatorischem PCR-Nachweis .....	67
Abbildung 14: Bestätigung des Genotyps der Vierfachmutante <i>T. thermophilus</i> BL03 ( $\Delta bgl \Delta 1787 \Delta 0340-1 \Delta 0904$ ) mittels PCR-Nachweis und Southern Blot Analyse .....	68
Abbildung 15: Zusammenfassung der lipolytischen Aktivitäten in <i>T. thermophilus</i> HB27 Wildtyp und den jeweiligen Mutanten $\Delta bgl$ , BL01, BL02 und BL03 .....	69
Abbildung 16: Untersuchung des Wachstumsverhaltens bei 70 °C und der Ausprägung der lipolytischen Aktivität über die Wachstumsphase.....	70
Abbildung 17: Ausgewählte Fosmidklone mit Aktivitäten auf unterschiedlichen Substraten .....	73
Abbildung 18: Statistik der 454-Sequenziererergebnisse der Fosmid-Bank <i>ANG48</i> .....	74
Abbildung 19: Vorhersage der Domänenzusammensetzung von Xyn5A .....	77
Abbildung 20: Darstellung des sequenzierten Fosmids HA-cmc-1, welches für cellulolytische Aktivität kodiert .....	79
Abbildung 21: Vorhersage der Domänenzusammensetzung von Cel12E.....	80
Abbildung 22: Fragmentierung von Cel12E während der Proteinaufreinigung .....	81
Abbildung 23: Visualisierung der Aufreinigungsschritte von Cel12E.....	82
Abbildung 24: Zusammenfassung der Aktivitätsprofile von Cel12E.....	84
Abbildung 25: Einfluss verschiedener Salze, Metallionen und ausgewählter Detergenzien auf die Aktivität von Cel12E.....	85
Abbildung 26: Einfluss von zweiwertigen Kobalt- und Manganionen und EDTA auf die Aktivität von Cel12E.....	86
Abbildung 27: Substratspektrum von Cel12E auf verschiedenen getesteten Substraten. ....	87
Abbildung 28: Dünnschichtchromatographische Analyse von Hydrolyseprodukten aus Polysacchariden.....	88
Abbildung 29: Bindeverhalten von Cel12E auf verschiedenen Polysacchariden. ....	88
Abbildung 30: Enzymkinetischer Verlauf der Aktivität von Cel12E mit steigender Substratkonzentration an CMC .....	89

Abbildung 31: Beispielhafte Zusammenstellung ausgewählter Fosmidklone, die in <i>T. thermophilus</i> auf Substrat- und Mikrotiterplatten Aktivitäten zeigten.....	92
Abbildung 32: Mikroskopie-Aufnahmen der Proben von heißen Quellen .....	94
Abbildung 33: Qualität der gewonnenen metagenomischen DNA im Agarosegel.....	96
Abbildung 34: Analyse der Diversität der Metagenombank, beispielhaft gezeigt an der Wasserprobe AZ3 .....	97
Abbildung 35: Identifizierung positiver Fosmidklone in <i>E. coli</i> aus den durchmusterten thermophilen Metagenombanken der Azoren und aus Kompost .....	100
Abbildung 36: Komplementationsexperiment in <i>T. thermophilus</i> nach 3 Tagen Inkubation bei 60 °C auf SH Minimalmedium mit 1,0 % (v/v) Tributyrin supplementiert .....	101
Abbildung 37: Aktivitäten einzelner Fosmidklone nach funktioneller Durchmusterung der thermophilen Metagenombanken in <i>T. thermophilus</i> BL03 .....	102
Abbildung 38: Zusammenfassung der lipolytischen Aktivitäten der Fosmidklone, die in den beiden Expressionwirten mittels <i>pNP</i> -Assay bestimmt wurden.....	104
Abbildung 39: Übersicht der GO-Klassifizierungen aller kodierten Proteine, die aus den Metagenomsequenzen vorhergesagt werden konnten.....	105
Abbildung 40: Übersicht aller sequenzierten und assemblierten Fosmide aus den vergleichenden Durchmusterungen in <i>E. coli</i> und <i>T. thermophilus</i> .....	107
Abbildung 41: Aktivitätsdaten der getesteten Esterasen im pJ-car05-RBS Expressionsvektor.....	110
Abbildung 42: Expression der metagenomischen Esterasen EstA2 und EstB1 mit pMK18 Plasmid in <i>T. thermophilus</i> BL03 nach 14 Stunden konstitutiver Expression bei 60 °C.....	112
Abbildung 43: Visualisierung der Aufreinigungsschritte der beiden metagenomischen Esterasen EstA2 sowie EstB1 mittels SDS-PAGE und Western-Blot .....	113
Abbildung 44: Enzymatische Charakterisierung und Aktivitätsparameter der aufgereinigten metagenomischen Esterasen EstA2 und EstB1 .....	115
Abbildung 45: Schematische Abb. der Regulation an der Promotorregion des <i>crtB</i> -Genclusters.....	118
Abbildung 46: Schematischer Aufbau des pJ-car05-RBS Vektors und Sequenz der beiden synthetischen Promotoren <i>P<sub>litR</sub></i> und <i>P<sub>crtB</sub></i> vor dem Zielprotein der Expression (Bgl)....	120
Abbildung 47: Aktivitätsprofil der wichtigsten extrazellulär aktiven Esterasen und Lipasen in <i>T. thermophilus</i> HB27.....	129
Abbildung 48: Zusammenfassung der Durchmusterungen verschiedener Genbanken in <i>E. coli</i> .....	134
Abbildung 49: Funktionelle Klassifikation aller kodierten Proteine, die auf assemblierten Fosmidsequenzen (siehe Tabelle 24) identifiziert werden konnten .....	137
Abbildung 50: Bioinformatische Analyse der beiden Fosmide der Klone TB01-A10 und TB-xyl4, welchen keine direkten Kandidatengene vorhergesagt werden konnten.....	140
Abbildung 51: Multiples Sequenz-Alignment der katalytischen Zentren ausgewählter Glycosidhydrolasen der Familie 12 und Kohlenhydratbindemodule der Familie 2.....	143
Abbildung 52: Graphische Repräsentation der vergleichenden Durchmusterungsergebnisse für lipolytisch-aktive Fosmide zwischen <i>E. coli</i> , <i>T. thermophilus</i> BL03 und bioinformatischen Methoden (schematisches Mengendiagramm).....	154
Abbildung 53: Sequenzvergleich der putativen Kandidatenproteine sowie experimentell bestätigter Esterase-aktiver Enzyme.....	155
Abbildung 54: Eingruppierung klassifizierter Esterase- und Lipase-Familien aufgrund Sequenzähnlichkeiten ( <i>similarity tree</i> ) .....	156
Abbildung 55: Homologieverteilung von Proteinsequenzen metagenomischer Fosmide, die in <i>E. coli</i> und <i>T. thermophilus</i> identifiziert werden konnten .....	165
Abbildung 56: Vergleich vorhergesagter und konservierter Promotorsequenzen in <i>T. thermophilus</i> . 167	
Abbildung 57: Promotorsequenzen, welche vor den für Esterasen kodierenden Kandidatengenen vorhergesagt werden konnten .....	168

## Tabellenverzeichnis

Tabelle 1: Verwendete Stämme.....	11
Tabelle 2: Verwendete Plasmide.....	12
Tabelle 3: In dieser Arbeit konstruierte und verwendete Vektoren.....	13
Tabelle 4: Zusammensetzung LB- und TB-Medium.....	16
Tabelle 5: SH-Minimalmedium.....	16
Tabelle 6: Medienzusätze und Antibiotika.....	16
Tabelle 7: Liste angewandeter Methoden (Kits) zur Extraktion metagenomischer DNA.....	19
Tabelle 8: Lösungen zur Herstellung chemisch kompetenter Zellen.....	23
Tabelle 9: Herstellung von YENB-Medium für elektrokompente Zellen.....	24
Tabelle 10: Arbeitsschritte für Southern Blot-Prozedur.....	29
Tabelle 11: Liste der durchmusterten Metagenombanken in <i>E. coli</i> .....	30
Tabelle 12: Verfügbares Probenmaterial, das für die Erstellung der Genbanken verwendet wurde.....	31
Tabelle 13: Liste der Substrate für funktionelle Durchmusterungen.....	34
Tabelle 14: Verwendete Programme zur bioinformatischen Auswertung der Sequenzdaten.....	37
Tabelle 15: Zusammensetzung der Lösungen für SDS-PAGE.....	40
Tabelle 16: Verwendete Säulen zur Aufreinigung von Protein Cel12E mittels FPLC (ÄKTA).....	41
Tabelle 17: Polysaccharid-Substrate für biochemische Charakterisierung von Cel12E.....	43
Tabelle 18: Zusammensetzung des DNSA-Reagens.....	44
Tabelle 19: Liste der beobachteten Hintergrundaktivitäten in <i>E. coli</i> und <i>T. thermophilus</i> HB27.....	49
Tabelle 20: Zusammenfassung der <i>in vivo</i> Transposonmutagenese in <i>T. thermophilus</i> .....	57
Tabelle 21: Zusammenfassung der <i>in vitro</i> Transposonmutagenese in <i>T. thermophilus</i> .....	59
Tabelle 22: <i>In silico</i> Proteom-Datensatz vorhergesagter lipolytischer Proteine in <i>T. thermophilus</i> HB27.....	61
Tabelle 23: Zusammenfassung aller aktiver <i>E. coli</i> Fosmidklone aus funktionellen Durchmusterungen. .....	72
Tabelle 24: Zusammenfassung aller sequenzierten und assemblierbaren metagenomischen Fosmide.....	76
Tabelle 25: Aufreinigungstabelle aus Präparation der heterolog exprimierten Glycosidhydrolase Cel12E.....	83
Tabelle 26: Auswertung der vergleichenden Durchmusterungsergebnisse in <i>E. coli</i> und <i>T. thermophilus</i> HB27.....	91
Tabelle 27: Sequenzdaten und Kandidatengene aus <i>T. brockianus</i> Genombank.....	93
Tabelle 28: Eigenschaften der Proben von heißen Quellen (Azoren).....	94
Tabelle 29: Ausbeuten der isolierten metagenomischen Nukleinsäuren aus heißen Quellen.....	96
Tabelle 30: Generierte einzelne Fosmidklone aus Metagenomprobe (Azoren).....	98
Tabelle 31: Phylogenetische 16S rRNA-Analyse der Azorenprobe.....	99
Tabelle 32: Bioinformatische Analyse der sequenzierten und assemblierten Fosmidklone aus vergleichenden Durchmusterungen in beiden Wirten.....	108
Tabelle 33: Vergleich der in <i>T. thermophilus</i> verfügbaren Expressionssysteme mit pJ-car05-RBS... .....	121
Tabelle 34: Durchmusterungsergebnisse anderer Metagenomstudien im Vergleich zu dieser Arbeit.....	133
Tabelle 35: Beispielhafte Übersicht charakterisierter bakterieller und archaeeller, thermostabiler Glycosidhydrolasen der Familie 12.....	146
Tabelle 36: Beispiele für thermostabile Esterasen und Lipasen, die aus metagenomischen Durchmusterungen gewonnen werden konnten, im Vergleich mit dieser Arbeit.....	151
Tabelle 37: Auswahl (meta)genomischer, thermostabiler Esterasen und Lipasen mit biochemischen Daten im Vergleich zu EstA2 und EstB1.....	161
Tabelle 38: Zusammenfassung vergleichender funktioneller Metagenomanalysen in <i>E. coli</i> und anderen Bakterien.....	163

Tabelle 39: Klonierung der <i>Suicide</i> -Plasmide zur Inaktivierung der lipolytischen Aktivität in <i>T. thermophilus</i> HB27.....	188
Tabelle 40: PCR-Nachweise der erfolgten Genunterbrechungen im <i>T. thermophilus</i> Chromosom. ..	188
Tabelle 41: Liste der klassifizierten Esterase/Lipase-Familien.....	189
Tabelle 42: Verwendete Oligonukleotide und PCR-Primer.....	191
Tabelle 43 : Konstruktion der Marker-freien Deletionsvektoren und Mutantendetektion.....	193
Tabelle 44: Assemblierungsstatistik der 454-Sequenzierung <i>EST6</i> .....	194
Tabelle 45: Sequenzanalyse des Fosmids HA-cmc-1 ( <i>contig</i> 4).....	195
Tabelle 46: Sequenzanalyse des Fosmids ES-6-B5 ( <i>contig</i> 8).....	196
Tabelle 47: Sequenzanalyse des Fosmids TB-est-7 ( <i>contig</i> 9).....	197
Tabelle 48: Sequenzanalyse des Fosmids TB-xyl-3 ( <i>contig</i> 5).....	198
Tabelle 49: Sequenzanalyse des Fosmids TB01-A10 ( <i>contig</i> 1).....	199
Tabelle 50: Sequenzanalyse des Fosmids TN-9-B5 ( <i>contig</i> 6).....	200
Tabelle 51: Sequenzanalyse des Fosmids ES-7-E10 ( <i>contig</i> 7).....	201
Tabelle 52: Sequenzanalyse <i>contig</i> 10 (MB-est-2/3/5 und 6).....	202
Tabelle 53: Sequenzanalyse <i>contig</i> 25 (MB-est- 7 und 8). .....	202
Tabelle 54: Sequenzanalyse der nicht zuordenbaren <i>contigs</i> .....	203
Tabelle 55: Sequenzanalyse des <i>contigs</i> 12. ....	204
Tabelle 56: Sequenzanalyse des <i>contigs</i> 17 . ....	204
Tabelle 57: Sequenzanalyse des Fosmids AZ2-4-B6.....	205
Tabelle 58: Sequenzanalyse des <i>shotgun</i> -Klons aus Fosmid M12-4-D9. ....	205
Tabelle 59: Sequenzanalyse des Fosmids AZ3-14-D2.....	206
Tabelle 60: Sequenzanalyse des Fosmids AZ3-14-D11.....	207
Tabelle 61: Sequenzanalyse des Fosmids AZ3-32-E5.....	208
Tabelle 62: Sequenzanalyse des Fosmids AZ3-33-C12.....	209
Tabelle 63: Sequenzanalyse des Fosmids M12-4-G10. ....	210



# I. Einleitung

## 1. Die Metagenom-Analyse

### 1.1. Die Vielfalt mikrobiellen Lebens

Mikroorganismen kommen ubiquitär in der Umwelt vor und besiedeln eine Vielzahl natürlicher und anthropogen geprägter Habitats, von der Tiefsee bis zu den Gletschern, von der Wüstenregion bis zum Polarkreis, vom Tiefengestein bis zu den Wolken in der Atmosphäre. Selbst an den – für menschliches Verständnis – unwirtlichsten Standorten konnten Organismen nachgewiesen werden, welche man schließlich als Extremophile bezeichnete (MacElroy, 1974). In den letzten Jahrzehnten wurde klar, dass das Auffinden solcher Mikroorganismen eher die Regel als die Ausnahme darstellt. Aufgrund ihrer Verbreitung und enormen Anpassungsfähigkeit stellen Prokaryoten mit geschätzten  $4$  bis  $6 \times 10^{30}$  Zellen und  $10^6$  bis  $10^8$  verschiedenen Spezies die größte Artenvielfalt des Lebens auf der Erde dar (Simon und Daniel, 2009; Whitman *et al.*, 1998). Ihre Überlebensstrategien spiegeln sich in der genetischen Ausstattung wider, die ihrerseits aus der evolutionären Anpassung an die gegebenen physikochemischen Bedingungen der Habitats resultieren. Diese Habitats können in Bezug auf hohe oder niedrige Temperaturen, pH, Druck, energiereiche Strahlung, Salzkonzentrationen, limitierte Nährstoffanbindung, Verfügbarkeit von Wasser und andere abiogene Faktoren mitunter extrem ausfallen (Gomes und Steiner, 2004). Die Nische, in der sich eine Zelle gegenüber anderen Mikroorganismen befindet, kann ebenfalls sehr spezifisch und äußerst komplex sein. Nach heutigen Schätzungen geht man davon aus, dass in einem Gramm Boden mehrere Tausend verschiedene mikrobielle Genome der dort lebenden Konsortien zu finden sind (Daniel, 2005; Torsvik *et al.*, 2002). Bis zum heutigen Tag sind in der Genomdatenbank des *National Center for Biotechnology Information* (NCBI, Stand 11. Januar 2014) 4.418 prokaryotische und 266 archaeelle Genome abgelegt. Diese Zahl stellt jedoch lediglich einen geringen Bruchteil aller in der Natur vorhandenen Spezies und mikrobieller Konsortien dar, da von der immensen Vielfalt an Mikroorganismen nur etwa 0,1 bis 1,0 % als kultivierbar gelten (Amann *et al.*, 1995; Torsvik *et al.*, 1990). Dieses Phänomen ist auch als „*the great plate count anomaly*“ bekannt geworden (Staley und Konopka, 1985), da die spezifischen Ansprüche der einzelnen Organismen unter Laborbedingungen bisher oft nur unzureichend nachgeahmt werden können.

Um diesem Problem zu begegnen wurden verschiedene Strategien entwickelt, damit diese „nicht kultivierbaren“ Mikroorganismen trotzdem charakterisiert werden können. Einerseits kommen verbesserte Kultivierungstechniken zum Einsatz, andererseits wird versucht, durch Analyse der

16S ribosomalen DNA (rDNA) die mikrobielle Vielfalt kultivierungsunabhängig zu erfassen. Eine weitere, kultivierungsunabhängige Strategie um die Mehrheit der Umweltorganismen trotzdem untersuchen zu können, ist die Metagenomik (Handelsman *et al.*, 1998).

In der Metagenomik wird die gesamtheitliche genetische Information der Organismen einer Umweltprobe (das Metagenom) gewonnen, um phylogenetische und funktionelle Eigenschaften der Umweltorganismen zu analysieren. Seit der ersten Extraktion und Analyse von Nukleinsäuren mikrobieller Konsortien aus Umweltproben und der funktionellen Charakterisierung von Genbanken (Healy *et al.*, 1995; Olsen *et al.*, 1986; Schmidt *et al.*, 1991; Torsvik und Goksøyr, 1978) hat sich die Metagenomik zu einem eigenständigen und wichtigen Forschungsbereich entwickelt.

## **1.2. Metagenomik als Quelle neuer Biokatalysatoren für industrielle Prozesse**

Die Metagenomanalyse ist für die angewandte Forschung von immer größerer Bedeutung, da in vielen biotechnologischen und pharmazeutischen Prozessen die Nachfrage nach neuen Enzymen groß ist. Im Zuge der Abkehr von traditionellen, chemisch-katalysierten Reaktionen hin zu einer enzymatischen oder fermentativen Prozesslösung, werden stereochemisch selektive Enzyme gesucht, die unter bestimmten Bedingungen (z.B. extreme Temperaturen oder pH-Werte) optimal funktionieren (Lorenz und Eck, 2005). Von besonderem Interesse sind dabei thermostabile Enzyme, die gegenüber herkömmlichen Enzymen eine Reihe vorteilhafter Eigenschaften aufweisen und daher aussichtsreiche Kandidaten für biotechnologische Anwendungen sein können (nach Sterner und Liebl, 2001): Durch höhere Prozesstemperaturen wird die Gefahr von mikrobiellen Kontaminationen vermindert und gleichzeitig die Reaktionsraten durch verbesserte Löslichkeit von Substrat und Produkt, durch erhöhte Diffusionsraten und erniedrigte Viskositäten gesteigert. Biotechnologisch relevante Enzyme können für eine Vielzahl an industriellen Prozessen eingesetzt werden, beispielsweise in der Papier- und Textilindustrie (Xylanasen, Cellulasen), in der Lebensmittelproduktion (z.B. Amylasen) und Landwirtschaft sowie in der Waschmittelindustrie (Proteasen, Lipasen) (Steele *et al.*, 2009). Seit vielen Jahren werden nützliche Gene und kodierte Enzyme aus Umweltproben isoliert, wie z.B. aus natürlich erhitztem vulkanischen Boden (Mientus *et al.*, 2013), unterschiedlichsten Erdproben (Wang *et al.*, 2012), arktischen Sedimenten (Fu *et al.*, 2013), heißen Quellen (Graham *et al.*, 2011), Kuh-Pansen (Ferrer *et al.*, 2012) oder dem Darm von „höheren“ Termiten (Nimchua *et al.*, 2012).

Die Zahl der in metagenomischen Studien üblich identifizierten Gene und Genprodukte fällt jedoch in Relation zum Umfang der bisher untersuchten Sequenzinformationen gering aus, trotz der mittlerweile verbesserten Detektionstechnologien in der funktionellen Metagenomik und dem Einsatz neuartiger

Hochdurchsatz-Sequenzierungstechnologien (*next-generation sequencing*, NGS). Diese Tatsache wurde auch als „*the great screen anomaly*“ beschrieben (Ekkers *et al.*, 2012).

### 1.3. Technologien und Anwendungen in der modernen Metagenomik

Zu Beginn jeder metagenomischen Analyse müssen Mikroorganismen aus der Umweltprobe herausgelöst werden, um sie effizient lysieren zu können (z.B. durch Filtration von Wasserproben, Homogenisieren von Bodenpartikeln, Abschaben von anhaftenden Organismen). Neben der direkten Lyse (Ogram *et al.*, 1987) der Zellen aus den unveränderten Probenmaterialien (Abbildung 1-A) können gezielte Anreicherungsstrategien (Dumont und Murrell, 2005) sowie Zellseparationstechniken mittels Nicodenz-Dichtezentrifugation (Berry *et al.*, 2003) und Kationenaustauscher (Jacobsen und Rasmussen, 1992) zum Einsatz kommen (Abbildung 1-B). Als Alternative können einzelne Zellen aus unbekanntem Konsortien isoliert werden. Beispielhafte Techniken hierfür sind optische Pinzetten (sogenannte *optical tweezers*) (Huber *et al.*, 1995) oder miniaturisierte, mikrofluidische Kultivierung mittels Gel-Einschlussröpfchen, welche zur parallelen Kultivierung von Umweltorganismen verwendet werden (Zengler *et al.*, 2002). Gebildete Mikrokolonien können durchflusszytometrisch detektiert und isoliert werden.

Kombinationen aus physikalischen, chemischen und enzymatischen Lysemethoden sind für die optimale Extraktion der metagenomischen DNA notwendig (Abbildung 1-C). Falls zu geringe Mengen chromosomaler Nukleinsäuren vorhanden sind, werden die Ausbeuten in einer PCR-basierten Genomamplifikation (*whole genome amplification*, WGA) durch die DNA-Polymerase des Bakteriophagen  $\Phi 29$  erhöht (zusammengefasst in Binga *et al.*, 2008). Diese Technik kommt ebenfalls bei der Sequenzierung von einzelnen Zellen (*single cell sequencing*) zum Einsatz.

Je nach Reinheit und Größe der erhaltenen Nukleinsäuren werden diese in Plasmide, Cosmide, Fosmide oder BACs kloniert (Abbildung 1-D) und in einem geeigneten Expressions- und/oder Klonierungsstamm kultiviert (Abbildung 1-E). Generell lassen sich metagenomische Analysetechniken in Sequenz- und Funktions-basierte Methoden einteilen. Die Sequenz-abhängige Metagenomanalyse bedient sich der Sequenzinformation von Nukleinsäuren und kodierten Proteinen. Durch konservierte Sequenzbereiche in den Zielgenen oder Proteindomänen lassen sich in selbst unbekanntem metagenomischen Sequenzen DNA-Fragmente mittels PCR amplifizieren und so vollständige Gen- und Proteinsequenzen mittels *Primer-Walking* erhalten (Abbildung 1-H; Chandrasekharaiah *et al.*, 2012; Jiang *et al.*, 2006; Tekedar und Sanli-Mohamed, 2011). Andere Hybridisierungstechniken basieren auf der Anreicherung und Identifizierung von homologen Sequenzabschnitten aus dem Gemisch der Umwelt-DNA durch Sonden mittels selektiver

Separationsmethoden (*magnetic bead separation* oder *subtractive hybridization*) und Umwelt-Microarray-Technologien (siehe Abbildung 1-I; Iwai *et al.*, 2011; Meyer *et al.*, 2007; Park *et al.*, 2008; Pathak und Gärtner, 2010). NGS-Sequenziermethoden sind mittlerweile etablierte Technologien und erlauben äußerst hohe Sequenzertiefen von *a priori* unbekanntem Nucleinsäuren aus Umweltproben (Abbildung 1-J; Shokralla *et al.*, 2012).

Mittels moderner Bioinformatik werden die sequenzierten DNA-Fragmente assembliert (*assembly*), phylogenetisch klassifiziert (*binning*), Gensequenzen vorhergesagt (*ORF prediction*) und ihre Funktion auf Homologie-Basis mit bekannten Genen (*annotations*) und ihren Genprodukten der öffentlich zugänglichen Datenbanken (DB) abgeglichen (Abbildung 1-1 bis -4; Thomas *et al.*, 2012; Wooley *et al.*, 2010). Im Zeitalter sinkender Sequenzierungskosten und ständiger technologischer Verbesserungen ist die Sequenz-basierte Analyse ein äußerst wichtiges Werkzeug in der heutigen Metagenomik. Sie erlaubt Einblicke in das Verständnis selbst komplexer mikrobieller Konsortien, ihrer Stoffwechselvielfalt und ökologischen Funktionen. Ein großer Nachteil ist jedoch, dass wahrhaft neuartige Gene und Genprodukte ohne ausreichende Übereinstimmung zu Proteinsequenzen in den Datenbanken (mit entsprechenden Annotationen und Funktionszuordnungen) durch Sequenz-basierte Methoden alleine nicht identifizierbar sind und daher übersehen werden können (Liebl, 2011; Ferrer *et al.*, 2012).

Die funktionelle Metagenomanalyse (Abbildung 1-F) basiert im Gegensatz zu Sequenz-abhängigen Verfahren auf der Identifizierung von aktiven Proteinen in einem Expressionswirt. Die häufigste genutzte Technik ist die phänotypische Detektion positiver Klone auf Indikatormedien. Dabei wird durch die Expression von funktionellem Protein ein chromogenes Substrat verändert oder abgebaut, sodass in der Folge die Freisetzung des Farbstoffes sichtbar wird oder aber sich Hofbildungen durch den Abbau der Substrate auf den Agarplatten feststellen lassen. Die modulierte Detektion nutzt neben dem Expressionswirt ein weiteres Reporter-System zur Identifizierung positiver Metagenomklone (Bijtenhoorn *et al.*, 2011; Lakhdari *et al.*, 2010), oder aber das Wachstum des Wirts wird durch heterologe Komplementation eines genetischen und metabolischen Defektes erzielt (z.B. durch Verstoffwechslung des normalerweise nicht abbaufähigen Substrates) (Parachin und Gorwa-Grauslund, 2011; Schallmey *et al.*, 2011). Die sogenannte Substrat- oder Produkt-basierte Detektion positiver Klone beruht auf der Bildung, Veränderung oder dem Abbau eines Induktors, welcher ein Reporter-System (z.B. auf Basis des *green fluorescent protein*-Gens) aktiviert (Uchiyama *et al.*, 2005; Uchiyama und Miyazaki, 2010; Williamson *et al.*, 2005). Mittels Durchflusszytometrie können fluoreszierende Zellen durch automatisierte Zellsortierung (*fluorescence-activated cell sorting* abgekürzt: FACS) im Hochdurchsatzformat von nicht aktiven Klonen getrennt werden. Neu identifizierte Gene und kodierte Genprodukte können schließlich heterolog exprimiert und biochemisch auf die Anwendbarkeit für biotechnologisch relevante Prozesse getestet werden (Abbildung 1-G).

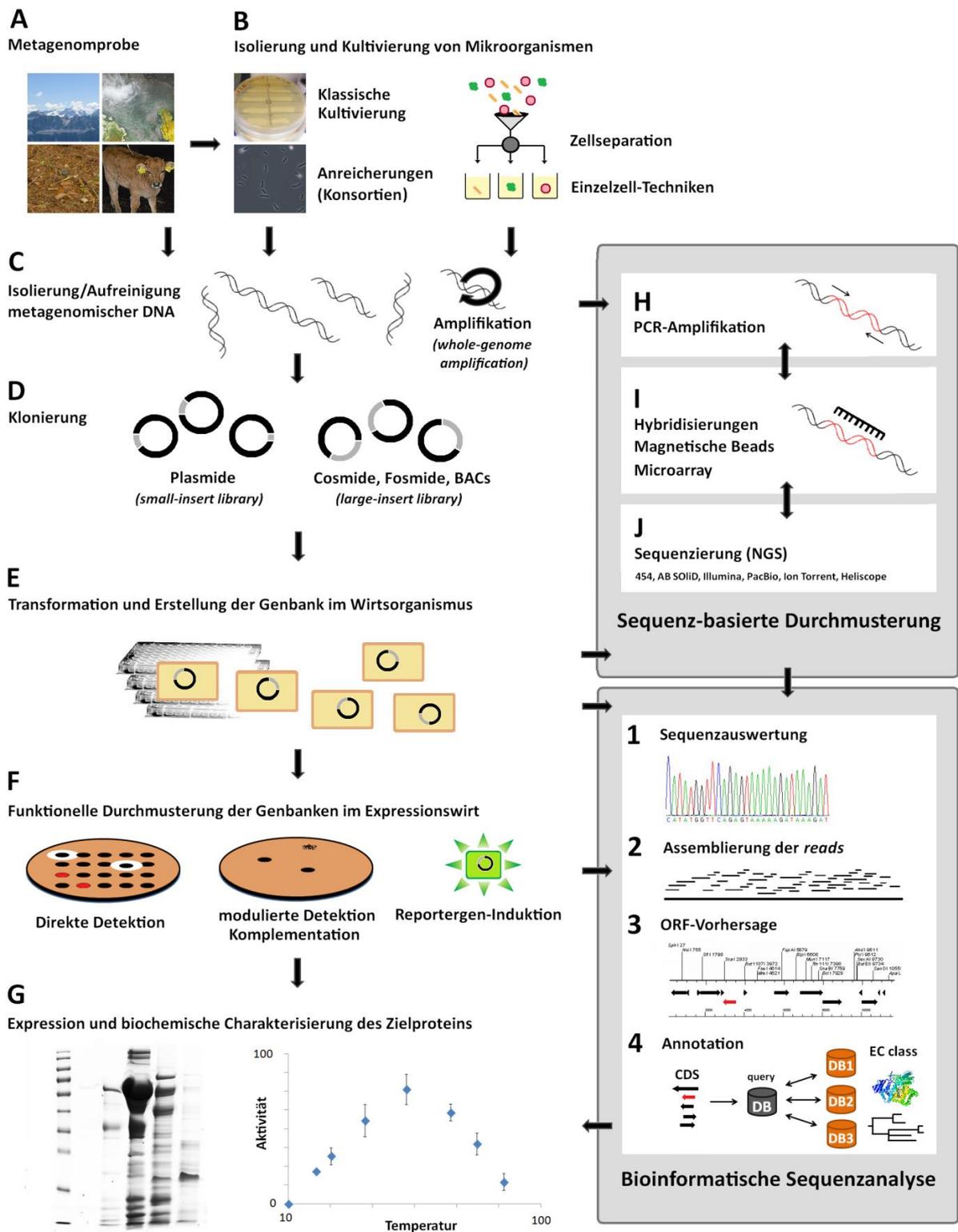


Abbildung 1: Flussdiagramm mit den wesentlichen Schritten der Metagenomanalyse. Ausgehend von unbekanntem mikrobiellen Spezies und Konsortien aus Umweltproben werden metagenomische Nucleinsäuren mit unterschiedlichen Strategien und Technologien analysiert, um neuartige Proteine zu identifizieren und nützliche Biokatalysatoren zu gewinnen (gezeigt ist die PDB Proteinstruktur 1QTM der *Taq*-Polymerase 1 aus *Thermus aquaticus*). Die einzelnen Schritte werden im Text ausführlicher erläutert.

#### 1.4. Neue Wirte für metagenomische Durchmusterungen

Da die homologe Expression von Proteinen aus natürlichen Wirten sehr schwierig ist, greift man auf die Nutzung von heterologen Expressionswirten wie *Escherichia coli* zurück. *E. coli* ist bis heute der am häufigsten verwendete heterologe Wirt zur Herstellung, Expression und funktionellen Durchmusterung von metagenomischen Genbanken (Ekkers *et al.*, 2012; Simon und Daniel, 2011; Taupp *et al.*, 2011). Diese Tatsache ist auf die zahlreichen (historischen) Entwicklungen auf dem Gebiet der Molekularbiologie zurückzuführen, welche dieses Bakterium zu einem wichtigen Modellorganismus gemacht haben (Snyder und Champness, 2007). Das Bakterium ist genetisch gut untersucht, lässt sich einfach und günstig kultivieren, wächst zügig zu relativ hohen Zelldichten und verfügt über ein großes Repertoire an etablierten Expressionsvektoren zur Herstellung rekombinanter Proteine aus Fremd-DNA (Sørensen und Mortensen, 2005a,b).

Trotz allen diesen Entwicklungen ist *E. coli* nicht immer der optimale Wirt. Die verhältnismäßig niedrige Zahl aktiver Enzyme aus metagenomischen Durchmusterungen weist darauf hin, dass *E. coli* häufig nur einen Teil der verfügbaren Sequenzinformation aus Fremdgenen in funktionelles Protein übersetzen kann. Die Ursachen hierfür sind vielseitig und noch nicht ausreichend erforscht, betreffen aber alle beteiligten Ebenen von der Klonierung, Transkription und Proteinbiosynthese bis hin zum funktionellen Zielprotein (zusammengefasst in Leis *et al.*, 2013): (i) die Duldung von Fremd-DNA im Wirt (Klonierbarkeit) (Gabor *et al.*, 2004; Sorek *et al.*, 2007), (ii) der GC-Gehalt (Holben *et al.*, 2004), (iii) die Erkennung von heterologen Transkriptionssignalen (Fremd-Promotoren) (Warren *et al.*, 2008), (iv) die Stabilität der messenger RNA (mRNA) (Kudla *et al.*, 2009), (v) die Präsenz der ribosomalen Bindestelle (RBS) (Sørensen und Mortensen, 2005a), (vi) der Codonpräferenz zur Initiation der Translation (Villegas und Kropinski, 2008), und (vii) der Verteilungshäufigkeit der verwendeten Basentriplets (*codon usage*) (Ekkers *et al.*, 2012). Nach der Translation wird die korrekte Faltung des gebildeten Polypeptids durch (viii) einen ausgewogenen Expressionsgrad des Zielproteins, (ix) durch etwaige post-translationale Proteinmodifikationen, (x) durch molekulare Faltungshelfer wie Chaperone ermöglicht, (xi) aber auch die Zusammensetzung der Signalpeptide zur Proteinsekretion ist dabei erwähnenswert (Mergulhão *et al.*, 2005; Sørensen und Mortensen, 2005a). (xii) Schließlich hängt es ebenfalls von den physiologischen und physikochemischen Parametern ab, ob ein Enzym in einem Wirt aktiv ist oder nicht (z.B. Präsenz von Cofaktoren, optimale Bedingungen von Temperatur und pH-Wert).

Es gibt bereits einige Beispiele für die Nutzung anderer Expressionswirte als *E. coli* für Metagenomanalysen: Darunter sind neben diversen bakteriellen Wirten (Aakvik *et al.*, 2009; Courtois *et al.*, 2003; Craig *et al.*, 2010; Martinez *et al.*, 2004; McMahan *et al.*, 2012), auch

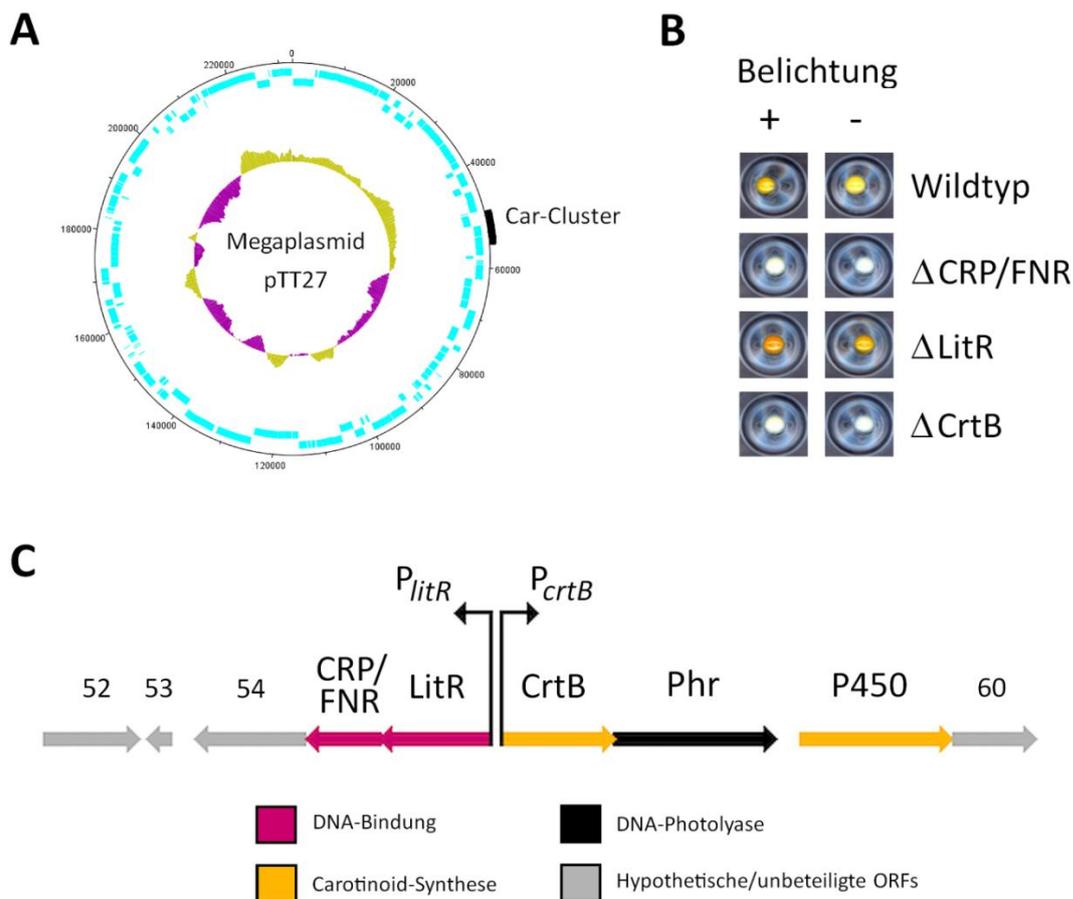
eukaryotische Expressionssysteme (Hefen) beschrieben, welche für funktionelle metagenomische Durchmusterungen erfolgreich angewendet werden konnten (Parachin und Gorwa-Grauslund, 2011; Kellner *et al.*, 2011; Damon *et al.*, 2011).

## 2. Der thermophile Modellorganismus *Thermus thermophilus* HB27

*Thermus thermophilus* wurde erstmals von Oshima und Imahori im Jahr 1974 aus heißen Quellen auf der japanischen Izu-Halbinsel isoliert und bezeichnet ein Gram-negatives, nicht Sporenbildendes und nicht motiles Bakterium. Als stäbchenförmiger Organismus ist es in der Lage, Sphärenförmige, multizelluläre Körper zu bilden (sogenannte *rotund bodies*). Die Vertreter der Gattung *Thermus* sind stammesgeschichtlich gesehen sehr alt und gehören der phylogenetisch tiefverzweigten *Deinococcus-Thermus* Entwicklungslinie an. Es werden regelmäßig neue *Thermus*-Arten isoliert, die aus verschiedensten thermophilen Habitaten stammen. Deren Wachstumsoptimum liegt zwischen 62 und 75 °C, jedoch sind manche *Thermus*-Arten auch in der Lage, bei Temperaturen bis zu 85 °C zu wachsen, was sie als extrem thermophile Organismen auszeichnet (Cava *et al.*, 2009). Mit einer Generationszeit von etwa 40 Minuten erreicht *T. thermophilus* darüber hinaus sehr hohe Zelldichten in kurzer Zeit. Seit der Sequenzierung des Genoms von *T. thermophilus* HB27 im Jahr 2004 (Henne *et al.*, 2004), liegen drei weitere *T. thermophilus*-Stämme bis heute laut NCBI-Datenbank (Stand November 2013) vollständig sequenziert vor. Jüngste Untersuchungen haben gezeigt, dass womöglich alle *T. thermophilus*-Stämme polyploid sind (Ohtani *et al.*, 2010). Das Genom von *T. thermophilus* HB27 besteht aus einem Chromosom mit 1,89 Mbp Größe und einem Megaplasmid (pTT27) mit 232 kbp. Der GC-Gehalt beträgt 69,4 mol % und etwa 95 % des Genoms kodieren für die insgesamt 2.200 Proteine, welche zum Teil biotechnologisch interessant sind (Liebl, 2004). Das Bakterium ist natürlich transformierbar und verfügt über einen äußerst effizienten DNA-Aufnahmeapparat, welcher durch hohe Transformationsfrequenzen ( $10^{-1}$  bis  $10^{-2}$  Transformanten pro Lebendzellzahl) und hohe Aufnahmegeschwindigkeiten charakterisiert ist (pro Zelle werden bis zu 40 kbp Nukleinsäure pro Sekunde aufgenommen) (Koyama *et al.*, 1986; Schwarzenlander und Averhoff, 2006). Zahlreiche etablierte genetische Werkzeuge zur Transformation, Selektion und Mobilisierung sowie zur gezielten genetischen Manipulation sind vorhanden (zusammengefasst in Cava *et al.*, 2009). Darüberhinaus existiert bereits ein kommerziell erhältliches Expressionssystem auf Basis des *nar*-Operons in *T. thermophilus* HB8 (Moreno *et al.*, 2005). *T. thermophilus* wird für systembiologische Untersuchungen als Modellorganismus verwendet, da die thermostabilen Enzyme besonders einfach zu kristallisieren sind (Jenney und Adams, 2008; Yokoyama *et al.*, 2000). Durch 3D-Strukturaufklärung aller bekannten Proteinfamilien sollen schließlich die biologischen Prozesse in diesem thermophilen Bakterium studiert werden können (*Structural and Functional Whole-Cell Project for T. thermophilus* HB8, <http://www.thermus.org>).

## 2.1. Regulation der Carotinoid-Biosynthese: das Car-Cluster

Die Fähigkeit von *T. thermophilus* Carotinoide bilden zu können führte zur vorläufigen Klassifikation als Flavobakterium (ehemals *Flavobacterium thermophilus*). Die Regulation des beteiligten lichtabhängigen Carotinoid-Clusters (Car- bzw. *crt*-Clusters, siehe Abbildung 2A-C) und der relevanten Gene wurde bereits intensiv studiert (A. Angelov, unveröffentlichte Ergebnisse; Ortiz-Guerrero *et al.*, 2011; Takano *et al.*, 2011).



**Abbildung 2: Eigenschaften des Car-Clusters in *T. thermophilus* HB27.** A: Das Car-Cluster (TT\_P0055 bis TT\_P0059) ist auf dem Megaplasmid pTT27 zwischen Position 49,24 bis 55,12 kbp lokalisiert. B: Vorarbeiten zu Mutationsanalysen in *T. thermophilus* HB27 erlaubten die phänotypische Charakterisierung beteiligter Gene des Car-Clusters und seiner Regulation in Abhängigkeit des Belichtungszustandes (zusammengefasst in Geissler, 2011; Angelov, unveröffentlichte Ergebnisse). Durch Lichtexposition wird in Wildtyp-Zellen eine Zunahme der Carotinoid-Bildung gegenüber unbelichteten Zellen beobachtet. Die Inaktivierung des transkriptionellen Aktivators CRP/FNR (TT\_P0055) oder der Phytoen-Synthase CrtB (TT\_P0057) führt zu einem Carotinoid-defizienten Phänotyp (weiße Kolonien). Die Deletion des lichtabhängigen Repressors LitR (TT\_P0056) hingegen resultiert in einem dereprimierten Phänotyp, welcher Carotinoide in einem höheren Ausmaß akkumuliert als der Wildtyp (unabhängig vom Belichtungszustand). C: Schematischer Aufbau des Car-Clusters (Darstellung nach Takano *et al.*, 2011). Zur Regulation des bidirektionalen Promotors  $P_{litR}$  und  $P_{crtB}$  ist der Aktivator CRP/FNR (TT\_P0055) sowie der lichtabhängige Repressor LitR (TT\_P0056) vorhanden, welche die Expression der Phytoen-Synthase CrtB unter Belichtung aktivieren. Weitere Elemente des Car-Clusters beinhalten eine DNA-Photolyase (TT\_P0058) sowie TT\_P0059 ( $\beta$ -Carotin Hydroxylase, P450 Superfamilie). Die TT\_P Gennummern sind über den Pfeilen angegeben.

Die Regulation beruht auf einem Liganden-abhängigen CarA/LitR Regulator (*light-induced transcriptional regulator*) namens LitR (kodiert von ORF TT\_P0056), welcher bei Dunkelheit die beiden Gene *crtB* (Phytoen-Synthase) und *litR* bidirektional reprimiert. In Anwesenheit des Cofaktors 5'-Deoxyadenosylcobalamin (Ado-B<sub>12</sub>) tetramerisiert LitR zu einem aktiven Proteinkomplex und inhibiert den Promotor für die Transkription des Schlüsselenzyms der Carotinoidbiosynthese. Durch die Einwirkung von Licht wird die Bindung zwischen LitR und dem Cobalamin unterbrochen, sodass der Repressor am *crtB*-Operon abfällt und die Phytoen-Synthase für die vermehrte Carotinoid-Bildung exprimiert wird (Ortiz-Guerrero *et al.*, 2011). Damit konnte erstmals gezeigt werden, dass Ado-B<sub>12</sub> als Cofaktor für die lichtabhängige transkriptionelle Regulation in dem thermophilen Bakterium fungiert. Dies führt zu einer deutlichen Orangefärbung der Zellen nach erfolgter Belichtung.

## 2.2. *T. thermophilus* als Expressionswirt für Genbanken

Zum Zwecke der metagenomischen Durchmusterungen wurde bereits im Vorfeld dieser Arbeit ein *E. coli*-*Thermus Shuttle*-Fosmid konstruiert (pCT3FK), welches die Klonierung von großen DNA-Fragmenten von etwa 40 kb Sequenzinformationen ermöglicht (Angelov *et al.*, 2009). Der Vektor basiert auf dem kommerziell erhältlichen pCC1FOS Fosmid (Epicentre, Madison, USA). Die Klonierung von selbst schwer klonierbaren DNA-Sequenzen wird durch einen F-Faktor basierten Replikationsursprung (*single-copy*) sichergestellt. Durch Induktion mittels Arabinose-Supplementation aktiviert das chromosomal lokalisierte Gen *trfA* des *E. coli* EPI300-Stammes den Multi-copy origin *oriV*. Durch die Erhöhung der Kopienzahl in der Zelle können selbst gering transkribierte und translatierte Genprodukte nachgewiesen werden. Zusätzliche genetische Elemente für *Thermus* erlauben die chromosomale Integration des Fosmids in den *pyr*-Locus. Eine thermostabile Kanamycin-Adenyltransferase (Kat) aus dem *E. coli*-*Thermus Shuttle*-Plasmid pMK18 (de Grado *et al.*, 1999) stellt die Selektion unter Anwesenheit von Kanamycin sicher. Um die Funktionalität des Systems zu testen, wurde eine Genbank des frei lebenden saccharolytischen Bakteriums *Spirochaeta thermophila* vergleichend in *E. coli* und *Thermus* durchmustert (Angelov *et al.*, 2009). Von den insgesamt 192 Fosmidklonen konnten mehr funktionell exprimierte Xylanase-Gene in dem thermophilen Wirt als im mesophilen Bakterium *E. coli* aufgefunden werden. Die besondere phylogenetische Stellung und physiologische Ausstattung von *T. thermophilus* erlaubt es daher, DNA-Banken fremder Organismen hinsichtlich interessanter Aktivitäten zu durchmustern, welche in *E. coli* sonst nicht identifizierbar sind. Die Erweiterung dieses Ansatzes auf metagenomische Genbanken dürfte die Ausbeuten für neue, biotechnologisch interessante Enzyme erhöhen.

## II. Zielsetzung der Arbeit

### **Verbesserung von *T. thermophilus* HB27 als neuartigen Expressionswirt für Metagenomanalysen:**

Zur funktionellen Durchmusterung von metagenomischen Fosmidbanken ist es unerlässlich, *T. thermophilus* auf seine Anwendbarkeit hin zu untersuchen. Die hohe lipolytische Hintergrundaktivität des Bakteriums verhindert jedoch eine metagenomische Durchmusterung nach Esterasen und Lipasen. Durch die Anwendung eines neuartigen markerlosen Deletionssystems sollte die lipolytische Aktivität des Wirts gezielt reduziert werden und eine multiple Deletionsmutante auf ihre Verwendung für metagenomische *Screenings* geprüft werden. Die Etablierung einer Transposonmutagenese-Bank erlaubt die Ausarbeitung neuer molekularbiologischer Werkzeuge zur zufälligen Mutagenese für diesen thermophilen Modellorganismus. Zur Expression metagenomisch identifizierter Proteine sollte ein neuartiges und induzierbares Expressionssystem auf Basis des lichtabhängigen Car-Clusters in *T. thermophilus* konstruiert und angewendet werden.

### **Funktionelle Durchmusterung diverser metagenomischer Fosmidbanken auf biotechnologisch relevante, neuartige Enzyme in *E. coli*:**

Die zur Verfügung stehenden metagenomischer Fosmidbanken stammten aus psychrophilen, mesophilen und thermophilen Habitaten und sollten funktionell u.a. auf cellulolytische, hemicellulolytische, proteolytische, lipolytische und amylolytische Aktivitäten in *E. coli* als Standardwirt untersucht werden. Die aufgefundenen Fosmidklone sollten mittels *next-generation sequencing* (NGS) sequenziert werden. Aufgrund bioinformatischer Analyse können die biotechnologisch interessanten Kandidatengene identifiziert und exprimiert werden. Die heterolog produzierten Enzyme sollten bezüglich ihrer biochemischen Aktivitäten charakterisiert werden.

### **Konstruktion und vergleichende Durchmusterung thermophiler Metagenombanken in *E. coli* und *T. thermophilus* HB27:**

Die metagenomische DNA von Probenmaterialien (Wasser und Sedimente) heißer Quellen von den Azoren sollten mit optimierten Extraktionsmethoden präpariert werden, um so die Erstellung einer möglichst großen und diversen metagenomischen Fosmidbank in *E. coli* zu ermöglichen. Die Klonierung in das pCT3FK-*Shuttle*-Fosmid erlaubt die parallele Durchmusterung der Banken in dem Standardexpressionswirt *E. coli* und in dem thermophilen Modellorganismus *T. thermophilus*. Die Nutzung von *Thermus* als neuartigen Expressionswirt ermöglicht die Identifizierung neuartiger Gene und Enzyme, die mit dem Standardwirt *E. coli* nicht identifiziert werden können. Dies gilt insbesondere für Metagenombanken aus thermophilen Konsortien. Beispielhaft sollte eine Genombank aus *Thermus brockianus* vergleichend durchmustert werden.

### III. Materialien und Methoden

#### 1. Verwendete Stämme und Wachstumsbedingungen

##### 1.1. Stämme und Plasmide

Die in dieser Arbeit verwendeten bakteriellen Stämme sind in Tabelle 1 zusammengefasst.

Tabelle 1: Verwendete Stämme.

Stamm	Genotyp, weitere Information	Referenz
<i>Escherichia coli:</i>		
XL1-Blue	<i>recA<sup>-</sup>, thi, hsdR1, supE44, relA1, lacF<sup>'</sup>, proAB, lacI<sup>q</sup>, lacZΔM15, Tn10[Tet]</i>	Stratagene (La Jolla, USA)
XL10-GOLD®	<i>endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet<sup>R</sup> F'[proAB lacI<sup>q</sup>ZΔM15 Tn10(Tet<sup>R</sup> Amy Cm<sup>R</sup>)]</i>	Stratagene
DH10B	<i>F<sup>-</sup> endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ<sup>-</sup></i>	Invitrogen (Carlsbad, USA)
JM109	<i>endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB<sup>+</sup> Δ(lac-proAB) e14- [F<sup>'</sup> traD36 proAB<sup>+</sup> lacI<sup>q</sup> lacZΔM15] hsdR17(rK:mK<sup>+</sup>)</i>	Promega (Fitchburg, USA)
EPI300-T1R	<i>F<sup>-</sup> mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara,leu)7697 galU galK λ<sup>-</sup> rpsL (Str<sup>R</sup>) nupG trfA tonA</i>	Epicentre (Madison, USA)
BL21 (DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B intr::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>	New England Biolabs (Ipswich, USA)
<i>Thermus thermophilus:</i>		
HB27	Wildtyp Laborstamm Göttingen, DSMZ 7039	Oshima und Imahori (1974); Deutsche Sammlung von Mikroorganismen und Zellkulturen
Δ <i>bgl</i>	ΔTT_P0042	Ohta <i>et al.</i> (2006); H. Li, (unveröffentlicht)
HB27:: <i>nar</i>	<i>nar</i> -Operon	Ramírez-Arcos <i>et al.</i> (1998b)
TL01	ΔTT_P0042 Δ <i>crtB</i> :: <i>bgl</i>	A. Lichev, Bachelorthesis (2011)
Δ <i>car</i>	Δ <i>bgl</i> , ΔTT_P0055 bis ΔTT_P0058	diese Arbeit

**Tabelle 1, Fortsetzung**

<b>Stamm</b>	<b>Genotyp, weitere Information</b>	<b>Referenz</b>
Δ340-1::kat	ΔTT_C0340-1::kat, kat-Austauschmutante	diese Arbeit
Δ749	ΔTT_C0749:kat, pBKat-Insertionsmutante	diese Arbeit
Δ797	ΔTT_C0797:kat, pBKat-Insertionsmutante	diese Arbeit
Δ824	ΔTT_C0824:kat, pBKat-Insertionsmutante	diese Arbeit
Δ904	ΔTT_C0904:kat, pBKat-Insertionsmutante	diese Arbeit
Δ904::kat	ΔTT_C0904::kat, kat-Austauschmutante	diese Arbeit
Δ1085	ΔTT_C1085:kat, pBKat-Insertionsmutante	diese Arbeit
Δ1341	ΔTT_C1341:kat, pBKat-Insertionsmutante	diese Arbeit
Δ1452	ΔTT_C1452:kat, pBKat-Insertionsmutante	diese Arbeit
Δ1787	ΔTT_C1787:kat, pBKat-Insertionsmutante	diese Arbeit
Δ1787::kat	ΔTT_C1787::kat, kat-Austauschmutante	diese Arbeit
BL01	ΔTT_P0042 ΔTT_C1787, Markerfreie Zweifachmutante	diese Arbeit
BL02	ΔTT_P0042 ΔTT_C1787 ΔTT_C0340-0341, Markerfreie Dreifachmutante	diese Arbeit
BL03	ΔTT_P0042 ΔTT_C1787 ΔTT_C0340-0341 ΔTT_C0904, Vierfachmutante	diese Arbeit

Die Liste der verwendeten Plasmide ist in Tabelle 2 abgebildet.

**Tabelle 2: Verwendete Plasmide.**

<b>Plasmid</b>	<b>Größe (kb)</b>	<b>Beschreibung</b>	<b>Referenz</b>
pJET 1.2	2,97	Blunt-end Klonierungsvektor, rep(pMB1), bla(Amp <sup>R</sup> ), P <sub>lac</sub> und P <sub>T7</sub> , eco47IR, MCS	Fermentas (Vilnius, Litauen)
pUC18	2,69	Klonierungsvektor, rep(pMB1), bla(Amp <sup>R</sup> ), lacPOZ', MCS	Fermentas (Vilnius, Litauen)
pMK18	5,04	E. coli-Thermus Shuttle-Vektor für Klonierungen und konstitutive Proteinexpression, oriEc, repA, P <sub>slp</sub> , kat(Kan <sup>R</sup> ), αLacZ, MCS	de Grado <i>et al.</i> (1999)
pET21a(+)	5,44	Expressionsvektor, pBR322 ori, bla(Amp <sup>R</sup> ), P <sub>T7</sub> , MCS, His-tag, lacI	Merck-Millipore (Billerica, USA)
pS18a	3,43	Deletionsvektor für Thermus, rep(pMB1), bla(Amp <sup>R</sup> ), rpsLI(Str <sup>R</sup> ), lacPOZ'	Blas-Galindo <i>et al.</i> (2007)
pMKE2	7,42	Expressionsvektor für Thermus-Stamm mit nar-Operon, oriE, repA, kat(Kan <sup>R</sup> ), P <sub>nar</sub>	Moreno <i>et al.</i> (2005)

Tabelle 2, Fortsetzung

Plasmid	Größe (kb)	Beschreibung	
pCC1FOS	8,14	Fosmidvektor für Klonierung, Induzierung und Expression von (meta)genomischer DNA, <i>repE</i> , <i>oriV</i> , <i>ori2</i> , <i>parABC</i> , <i>Chl<sup>R</sup></i> , <i>lacZ</i> , <i>P<sub>T7</sub></i> , <i>redF</i> , <i>cos</i> , <i>loxP</i>	Epicentre
pCT3FK	11,6	<i>E. coli-Thermus Shuttle</i> -Fosmidvektor für Klonierung, Induzierung und Expression von (meta)genomischer DNA, <i>repE</i> , <i>oriV</i> , <i>ori2</i> , <i>parABC</i> , <i>Chl<sup>R</sup></i> , <i>kat(Kan<sup>R</sup>)</i> , <i>lacZ</i> , <i>P<sub>T7</sub></i> , <i>redF</i> , <i>cos</i> , <i>loxP</i>	Angelov <i>et al.</i> (2009)
pCR2.1-XL-TOPO	3,93	Klonierungsvektor für PCR-Produkte mit 3'-dA-Überhängen, <i>f1 ori</i> , <i>rep(pMB1)</i> , <i>bla(Amp<sup>R</sup>)</i> , <i>kan(Kan<sup>R</sup>)</i> , <i>P<sub>lac</sub></i> , <i>lacZa</i>	Invitrogen
pMOD3-kat	3,97	Transposon Klonierungsvektor mit thermostabilem Resistenzmarker für <i>Thermus</i> , <i>pBR322 ori</i> , <i>R6Kγori</i> , <i>bla(Amp<sup>R</sup>)</i> , <i>kat(Kan<sup>R</sup>)</i> , <i>P<sub>slp</sub></i> , mosaic ends (ME)	Angelov <i>et al.</i> (2013)
pMarB	8,32	<i>In vivo</i> Transposonvektor mit induzierbarer <i>HimarI</i> Transposase für <i>Bacillus</i> , <i>oriEc</i> , <i>ori<sup>T</sup></i> , <i>repG<sup>+</sup>ts</i> , <i>bla(Amp<sup>R</sup>)</i> , ( <i>Erm<sup>R</sup></i> ), <i>kan(Kan<sup>R</sup>)</i> , <i>P<sub>B</sub></i> , <i>HimarI</i> , <i>ITRs</i> , <i>TnYLB-1</i>	Le Breton <i>et al.</i> (2006)
pBKat	4,49	<i>Suicide</i> -Plasmid für <i>Thermus</i> , <i>oriEc</i> ( <i>ColE1</i> ), <i>lacPOZ'</i> , <i>bla(Amp<sup>R</sup>)</i> , <i>kat(Kan<sup>R</sup>)</i> , <i>P<sub>slp</sub></i> , <i>MCS</i>	A. Wiezer, Diplomarbeit (2001)
pJ-car05-RBS-Bgl	6,42	Licht-induzierbares Expressionsplasmid auf Basis des Carotinoid-Clusters in <i>Thermus</i> , <i>rep(pMB1)</i> , <i>bla(Amp<sup>R</sup>)</i> , <i>kat(Kan<sup>R</sup>)</i> , <i>P<sub>slp</sub></i> , <i>P<sub>litR</sub></i> , <i>P<sub>crbB</sub></i> , <i>litR</i> , <i>CRP/FNR</i> , <i>bgl</i>	F. Terzi, Masterthesis (2012)
pJΔ57::bgl	6,22	<i>Suicide</i> -Plasmid für <i>Thermus</i> zum Genaustausch der β-Glucosidase ( <i>bgl</i> ) mit <i>crbB</i> (TT_P0057), <i>rep(pMB1)</i> , <i>bla(Amp<sup>R</sup>)</i> , <i>P<sub>litR</sub></i> , <i>P<sub>crbB</sub></i> , <i>litR</i> , <i>bgl</i>	A. Lichev, Bachelorthesis (2011)

Tabelle 3 fasst alle Vektorkonstrukte zusammen, die in dieser Arbeit entstanden sind.

Tabelle 3: In dieser Arbeit konstruierte und verwendete Vektoren.

Plasmid	Größe (kb)	Beschreibung
<b>Plasmide für Gendisruption und –deletion in <i>T. thermophilus</i>:</b>		
pBKat-Δ0749	5,19	Vektor zur Genunterbrechung von ORF TT_C0749 (Putative Esterase), Flanke kloniert über <i>SacI</i>
pBKat-Δ0797	4,98	Vektor zur Genunterbrechung von ORF TT_C0797 (Patatin-ähnliche Phospholipase), Flanke kloniert über <i>SacII</i>
pBKat-Δ0824	5,03	Vektor zur Genunterbrechung von ORF TT_C0824 (Protein mit Phospholipase D-ähnlicher Domäne), Flanke kloniert über <i>NotI</i>
pBKat-Δ1085	5,01	Vektor zur Genunterbrechung von ORF TT_C1085 (Protein mit Phospholipase D-ähnlicher Domäne), Flanke kloniert über <i>NotI</i>

**Tabelle 3, Fortsetzung**

<b>Plasmid</b>	<b>Größe (kb)</b>	<b>Beschreibung</b>
pBKat- $\Delta$ 1341	5,02	Vektor zur Genunterbrechung von ORF TT_C1341 ( $\alpha/\beta$ -Hydrolase), Flanke kloniert über <i>SacII</i>
pBKat- $\Delta$ 1452	5,01	Vektor zur Genunterbrechung von ORF TT_C1452 ( $\alpha/\beta$ -Hydrolase), Flanke kloniert über <i>NotI</i>
pCR- $\Delta$ 0340-1	5,44	Vektor zur markerlosen Gendeletion von ORFs TT_C0340 und TT_C0341 ( $\alpha/\beta$ -Hydrolasen)
pCR- $\Delta$ 0340-1::kat	6,66	Vektor zum Genaustausch von ORFs TT_C0340 und TT_C0341 ( $\alpha/\beta$ -Hydrolasen) mit <i>kat</i> -Kassette
pS18a- $\Delta$ 0904::kat	6,65	Vektor zum Genaustausch von ORF TT_C0904 (Pectinacetylerase) mit <i>kat</i> -Kassette
pS18a- $\Delta$ 0904	5,43	Vektor zur markerlosen Gendeletion von ORF TT_C0904 (Pectinacetylerase)
pS18a- $\Delta$ 1787::kat	6,65	Vektor zum Genaustausch von ORF TT_C1787 (Carboxylesterase) mit <i>kat</i> -Kassette
pS18a- $\Delta$ 1787	5,43	Vektor zur markerlosen Gendeletion von ORF TT_C1787 (Carboxylesterase)
pUC18- $\Delta$ 0904	4,68	Vektor zur markerlosen Gendeletion von ORF TT_C0904 (Pectinacetylerase)
pUC18- $\Delta$ 1787	5,43	Vektor zur markerlosen Gendeletion von ORF TT_C1787 (Carboxylesterase)
<b><i>E. coli</i> Expressionsplasmide:</b>		
pET21a-HACMC1	7,00	Expressionsvektor zur biochemischen Charakterisierung der archaeellen metagenomischen Endoglucanase Cel12E
pET21a-spHACMC1	6,92	Expressionsvektor zur biochemischen Charakterisierung der archaeellen metagenomischen Endoglucanase Cel12E mit nativem Signalpeptid
pET21a-EstA2	5,87	Expressionsvektor zur biochemischen Charakterisierung der archaeellen, metagenomischen Esterase EstA2
pET21a-EstA2his	5,89	Expressionsvektor zur biochemischen Charakterisierung der archaeellen, metagenomischen Esterase EstA2 mit <i>His-tag</i>
pET21a-EstB1	7,16	Expressionsvektor zur biochemischen Charakterisierung der bakteriellen, metagenomischen Esterase EstB1
pET21a-EstB1his	7,17	Expressionsvektor zur biochemischen Charakterisierung der bakteriellen, metagenomischen Esterase EstB1 mit <i>His-tag</i>
pET21a-ES6B5-18his	6,64	Expressionsvektor zur Bestätigung der Aktivität von Chi18A
pET21a-ES6B5-23chis	6,88	Expressionsvektor zur Bestätigung der Aktivität von Xyn5A
pET21a-TBxyl3-9c	6,51	Expressionsvektor zur Bestätigung der Aktivität von Xyn18A
pJ-car05-RBS-EstS9	6,76	Licht-induzierbarer Expressionsvektor für EstS9
pJ-car05-RBS-EstA2	5,58	Licht-induzierbarer Expressionsvektor für EstA2
pJ-car05-RBS-EstB1	6,87	Licht-induzierbarer Expressionsvektor für EstB1
pMK18-EstA2his	5,55	Konstitutiver Expressionsvektor für EstA2 mit <i>His-tag</i>

Tabelle 3, Fortsetzung

Plasmid	Größe (kb)	Beschreibung
pMK18-EstB1his	6,83	Konstitutiver Expressionsvektor für EstB1 mit His- <i>tag</i>
pMKE2-EstA2	7,85	Nitrat-induzierbarer Expressionsvektor für EstA2
pMKE2-EstB1	9,16	Nitrat-induzierbarer Expressionsvektor für EstB1
<b>Transposon Mutagenese-Plasmide:</b>		
pHimarB	5,47	Zwischenkonstrukt aus pMarB nach <i>EcoRI</i> -Verdau
pHimarB <sup>tv</sup>	6,09	Zwischenkonstrukt aus pHimarB mit thermostabiler Kanamycin-Resistenz ( <i>kat</i> )
pTnC2	6,09	Kontrollvektor für die <i>in vivo</i> Transposonmutagenese
pTnB2	4,97	Expressionsvektor der <i>Himar1</i> Transposase <i>in vivo</i>
<b>Erstellung eines Expressionssystems auf Basis des Car-Clusters:</b>		
pJ-car	8,64	Zwischenkonstrukt des Car-Clusters mit ORFs TT_P0055 bis TT_P0085 und je 1 kb äußeren Flanken
pJ $\Delta$ car	4,97	Vektor zur markerlosen Deletion des Car-Clusters in <i>T. thermophilus</i> HB27 $\Delta$ <i>bgl</i>
pJ-TL01	7,17	Zwischenkonstrukt zum Klonieren des ORFs TT_P0055 aus dem Reporterstamm TL01
pJ-car::bgl	6,81	Zwischenkonstrukt zur Wiederherstellung des vollständigen Car-Clusters mit den Elementen von TT_P0055 bis TT_P0058, <i>crtB::bgl</i> als Reportergen ersetzt
pMK18-car01	8,92	<i>Thermus bgl</i> -Expressionsvektor mit Car-Cluster nach <i>slp</i> -Promotor
pMK18-car02	8,92	<i>Thermus bgl</i> -Expressionsvektor mit Car-Cluster vor <i>slp</i> -Promotor
pCar01	6,93	Integratives <i>bgl</i> -Expressionsplasmid mit Car-Cluster nach <i>slp</i> -Promotor
pCar02	6,93	Integratives <i>bgl</i> -Expressionsplasmid mit Car-Cluster vor <i>slp</i> -Promotor

## 1.2. Verwendete Medien

Die verwendeten Medien wurden mit demineralisiertem Wasser (VE-Wasser) oder Mineralwasser (Aqua Purania, TSI GmbH & Co. KG, Zeven) hergestellt. Die Sterilisation fand unter Überdruckbedingungen bei 121 °C für 20 Minuten statt (VARIOKLAV, HP-Medizintechnik GmbH,

Oberschleißheim). Als Geliermittel zur Herstellung von Festmedien wurde 1,5 % (w/v) Agar-Agar verwendet. Zusätze wie Vitamine, Antibiotika und Zucker wurden zuerst gelöst, sterilfiltriert (0,20 µm Minisart, Sartorius, Göttingen) und den autoklavierten Medien bei etwa 60 °C zugesetzt. Pro Petrischale (Greiner Bio-One International AG, Kremsmünster, Österreich) wurden etwa 25 mL des gießfertigen Mediums bei Raumtemperatur ausgehärtet und unter der Sterilbank getrocknet. Medienzusammensetzungen für *lysogeny broth* (LB-), *Thermus broth* (TB-) (Tabelle 4) und SH-Minimalmedium (Tabelle 5) sowie verwendete Medienzusätze und Antibiotika (Tabelle 6).

**Tabelle 4: Zusammensetzung LB- und TB-Medium.**

	LB-Medium <sup>1</sup> ( <i>E. coli</i> )	TB-Medium <sup>2</sup> ( <i>T. thermophilus</i> )
Trypton	10 g/L	8,0 g/L
Hefeextrakt	5,0 g/L	4,0 g/L
NaCl	5,0 g/L	3,0 g/L
Wasserart	VE-Wasser	Mineralwasser
pH-Einstellungen	keine	pH 7,5 (NaOH)

<sup>1</sup>: Nach Sambrook *et al.* (1989); 2x LB mit doppelter Menge Trypton und Hefeextrakt bei gleichbleibender NaCl-Zugabe

<sup>2</sup>: *Thermus broth* nach Ramírez-Arcos *et al.* (1998b)

**Tabelle 6: Medienzusätze und Antibiotika.**

Zusatz	Abkürzung	Endkonzentration
Ampicillin	Amp	100 µg/mL
Chloramphenicol	Chl	12,5 µg/mL
Kanamycin	Kan	20 µg/mL
Streptomycin	Str	100 µg/mL
Glucose	Glc	1,0 % (w/v)
Isopropyl-β-D-thiogalactopyranosid	IPTG	0,1 – 1,0 mM
Fosmid Autoinduction solution (Epicentre)	AIS	1,0 % (v/v)

**Tabelle 5: SH-Minimalmedium.**

SH-Minimalmedium <sup>3</sup> ( <i>T. thermophilus</i> )	
Tributyrin	1,0 % (v/v)
K <sub>2</sub> HPO <sub>4</sub>	0,75 g/L
KH <sub>2</sub> PO <sub>4</sub>	0,25 g/L
NaCl	2,0 g/L
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2,5 g/L
Casamino acids	5,0 g/L
100 µg/ml Biotin	0,1 mg/L
1 mg/ml Thiamin	1,0 mg/L
Agar-Agar	15 g/L
pH-Einstellung pH 7,2 (NaOH)	

Zugabe steriler Spurenelementlösungen (Endkonzentration):

CaCl <sub>2</sub>	170 µM
CoCl <sub>2</sub> × 6 H <sub>2</sub> O	3,4 µM
CuSO <sub>4</sub> × 5 H <sub>2</sub> O	0,1 µM
FeSO <sub>4</sub> × 7 H <sub>2</sub> O	21,6 µM
MgCl <sub>2</sub>	615 µM
MnCl <sub>2</sub> × 4 H <sub>2</sub> O	2,5 µM
Na <sub>2</sub> MoO <sub>4</sub> × 2 H <sub>2</sub> O	5,0 µM
NiCl <sub>2</sub> × 6 H <sub>2</sub> O	0,1 µM
VOSO <sub>4</sub> × 3 H <sub>2</sub> O	0,5 µM
ZnSO <sub>4</sub> × 7 H <sub>2</sub> O	0,2 µM

<sup>3</sup>: nach Protokoll (Lehrstuhl für Mikrobiologie, AG Angelov, TU München, pers. Mitteilung)

### 1.3. Anzuchtbedingungen, Zellernte, Gewinnung von Zellextrakten

*E. coli* wurde unter Standardbedingungen bei 37 °C auf LB-Festmedium oder mit 5 mL LB-Flüssigmedium in sterilen Glasröhrchen schüttelnd inkubiert. Zur Proteinexpression wurden 200 mL Medium pro 1 L Schikanenkolben verwendet und bei 150 - 180 Upm für mindestens 4 Stunden oder ÜN nach Induktor-Zugabe inkubiert.

Die Kultivierung von *T. thermophilus* fand bei 70 °C (oder bei 60 °C unter Kanamycin-Selektion) auf TB-Festmedium oder in 5 mL TB-Flüssigmedium in 15 mL Plastikröhrchen (Greiner Bio-One International AG) statt. Um übermäßige Evaporation von Wasser aus den Medien zu vermeiden wurden die Petrischalen in Plastikfolien eingepackt bzw. Kulturgefäße flüssiger Kulturen mit Zellstoffstopfen oder Schraubverschlüssen verschlossen. Die ein- bis zweitägigen Inkubationen fanden in geeigneten Schüttelinkubatoren (Innova 42, New Brunswick, Edison, USA) statt. Die Belichtung der Kulturen wurde mit einer photosynthetischen Belichtungseinheit (Innova 42) durchgeführt, die Beleuchtungsstärke wurde mit einem Messgerät (PCE-172 LIGHT METER, PCE Instruments, Southampton, UK) bestimmt (in Abhängigkeit von der Distanz zur Lichtquelle betrug die Belichtungsstärke im Schüttler zwischen 2 und 3 kLux). Das Wachstum der Kulturen wurde durch die optische Dichte der Bakteriensuspensionen bei 600 nm Wellenlänge ( $OD_{600nm}$ ) in einem UV/Vis Spektrophotometer (Ultrospec 2100pro, Amersham Bioscience, UK) bestimmt. Zur Zellernte wurden die bewachsenen Kulturen auf Eis abgekühlt, bei 4 °C und  $5.000 \times g$  für 10 Minuten abzentrifugiert und die Zellpellets in 50 mM Tris-HCl Puffer (pH 8,0) gewaschen und resuspendiert. Kleinere Suspensionsvolumina (1-5 mL) wurden mit einem Ultraschall-Aufschlussgerät (Dr. Hielscher Ultrasonics GmbH, Teltow) für mehrere Minuten beschallt. Größere Volumina (z.B. für Proteinexpression) wurden zweimal mit einer French Press Zelle (American Instruments Company, Silver Spring, USA) lysiert. Der Erfolg des Zellaufschlusses wurde mittels abnehmender Trübung und Mikroskopie (Zeiss Axio Scope, Carl Zeiss, Oberkochen, Deutschland) der Zellsuspension kontrolliert. Die Reinheit der in dieser Arbeit verwendeten Stämme wurde durch Vereinzelungsausstriche auf geeigneten Selektionsmedien sichergestellt. Die Integrität der Plasmide aus Klonierungsstämmen wurde durch analytische Restriktionsverdaus überprüft. Die Lagerung der Kulturen im Kühlschrank (8 °C) war über mehrere Wochen in Flüssig- und Festmedien möglich. Zur Langzeitsicherung wurden 0,75 mL frisch gewachsene Kultur mit gleichem Volumen sterilem 50 % (w/v) Glycerol gemischt und bei mindestens -70 °C weggefroren. Genbanken in Mikrotiterplatten wurden in 180 µL Medium unter Zugabe von 10% DMSO gesichert. Die elektronische Dokumentation der abgelegten Stämme geschah durch die Labcollector Lab Management Software (AgileBio, New York, USA).

## 2. Molekulargenetische Methoden

Alle verwendeten Materialien (Gefäße, Pipettenspitzen, Lösungen) wurden autoklaviert (121 °C, 20 Minuten). Nicht-autoklavierbare Waren wurden mit 70 % Ethanol (w/v) abgewaschen oder mit einem Bunsenbrenner abgeflammt, Lösungen von hitzeempfindlichen Substanzen wurden sterilfiltriert. Molekularbiologische Methoden wurden nach Sambrook *et al.* (1989) mit Modifikationen durchgeführt.

### 2.1. Nukleinsäureanalytik

#### 2.1.1. Isolation und Aufreinigung von (meta)genomischen Nukleinsäuren und Plasmiden

Wasser aus Umweltproben wurde mit verschiedenen feinen Filtertypen filtriert. Zunächst wurden grobkörnige Schwebstoffe mit Filterpapier (Selecta, Cellulosenitrat, Schleicher & Schuell Bioscience GmbH, Dassel) abgetrennt. Die vorhandenen Mikroorganismen konnten auf feinporige Filter mit 0,45 µm Porengröße aus Polyethersulfon (Sartorius, Göttingen) und 0,20 µm aus Cellulosenitrat (Whatman GmbH, Dassel) abfiltriert werden. Die metagenomische DNA aus den Filtern sowie aus weiteren feststofflichen Sediment/Biofilm-Proben wurden mit verschiedenen Kits (MO BIO Laboratories, Inc., Carlsbad, USA; Epicentre) zur Präparation von Umwelt-DNA nach Herstellerangaben isoliert (Tabelle 7). In Änderung zu den bestehenden Protokollen und Herstellerangaben wurden die aufgereinigten Nukleinsäuren in sterilem ddH<sub>2</sub>O gelöst.

Für die Isolation von Nukleinsäuren von reinen Kulturen wurden standardmäßig 5 mL Zellkulturen aus Einzelkolonien inokuliert und über Nacht schüttelnd inkubiert. Genomische DNA wurde aus 500 µL der Zellsuspension mit Hilfe des MasterPure Complete DNA and RNA Purification Kit (Epicentre) nach Herstellerangaben isoliert. Plasmide und Fosmide für Klonierungen, Sequenzierungen und weitere Transformationen wurden mit dem Accuprep Plasmid Mini Extraction Kit (Bioneer Corporation, Daejeon, Südkorea) aufgereinigt. Analytische Plasmid- und Fosmidpräparationen wurden mit alkalischer Lyse nach Birnboim und Doly (1979) mit Modifizierungen durchgeführt: Die pelletierten Zellen wurden in 250 µL Resuspensionspuffer (10 mM EDTA, 50 mM Tris-HCl, pH 8,0) supplementiert mit Ribonuklease A (10 µg/mL) aufgenommen und resuspendiert. Zur Zelllyse wurde 250 µL Lysepuffer (0,2 M NaOH und 1,0 % SDS) hinzugegeben und durch mehrmaliges Invertieren gut durchgemischt. Nach Zugabe von 350 µL Neutralisationspuffer (3,2 M Kaliumacetat, pH 5,5) und erneutem Invertieren der Lösung wurden Zellreste für 15 Minuten

in einer Kühlzentrifuge bei 4 °C und mindestens 21.000 × g abzentrifugiert. Der Überstand mit der gelösten DNA wurde durch Zugabe von 0,7 Volumen Isopropanol und kräftigem Schütteln des Ansatzes gefällt, erneut abzentrifugiert und die pelletierte DNA mit 70 % Ethanol (v/v) gewaschen. Löste sich das Pellet vom Reaktionsgefäß, musste der Ansatz erneut kurz abzentrifugiert werden. Der Ethanol wurde mit einer Pipette abgenommen und das Pellet in der Sterilbank luftgetrocknet. Das Pellet wurde schließlich in 35 µL ddH<sub>2</sub>O gelöst.

**Tabelle 7: Liste angewendeter Methoden (Kits) zur Extraktion metagenomischer DNA.**

Produktname	Hersteller	Strategie der Isolation	Angewendet bei Umweltprobe
UltraClean Water DNA Isolation Kit	MO BIO	Filtration, <i>bead beating</i> , chemische Lyse mit Detergenz, DNA-Bindung an Silica-Membran	Wasser
Meta-G-nome DNA Isolation Kit	Epicentre	Filtration, enzymatische Lyse mit Lysozym und Proteinase K, DNA-Fällung mit Isopropanol	Wasser
SoilMaster DNA Extraction Kit	Epicentre	Enzymatische und chemische Lyse mit Proteinase K und Detergenz bei 65°C, DNA-Präzipitation	Sediment/Biofilm
Power(Max) Soil DNA Isolation Kit	MO BIO	<i>bead beating</i> , chemische Lyse mit Detergenz, DNA-Bindung an Silica-Membran	Sediment/Biofilm
UltraClean Fecal DNA Isolation Kit	MO BIO	<i>bead beating</i> , chemische Lyse mit Detergenz, anschließend DNA-Bindung an Silica-Membran	Sediment/Biofilm

Die analytische Plasmidanalyse von einzelnen Kolonien (*colony cracking*) erfolgte durch Resuspendieren der Zellen in 15 µL 10 mM EDTA (pH 8,0) und Zugabe von 10 µL Lyseagens (0,2 M NaOH, 1,0 % SDS). Nach 10 minütigem Aufkochen auf 95 °C und Abkühlen auf Eis, wurden Zellbestandteile durch Zentrifugation pelletiert und der Überstand mit der Plasmid-DNA direkt auf das Agarosegel geladen.

Aufreinigungen von Nukleinsäuren aus enzymatischen Reaktionen, PCR-Reaktionen und Gelelektrophoresen wurden mit Accuprep Gel und PCR Purification Kit (Bioneer) durchgeführt. Aufreinigung und Aufkonzentrieren kleinerer Volumina (auf weniger als 35 µL) war mit SureClean PCR-cleanup (Bioline, London, UK) nach Herstellerangaben möglich.

Für die Standardaufreinigungen und das Aufkonzentrieren von Nukleinsäuren wurde eine Isopropanol-Fällung durchgeführt. Dabei wurden die DNA-haltigen Lösungen mit 0,7 Volumen Isopropanol und 0,1 Volumen 3 M Natriumacetat-Lösung (pH 5,5) versetzt, 10 Minuten auf Eis inkubiert, und präzipitierte DNA durch Zentrifugieren (21.000 × g, 15 Minuten, 4 °C) pelletiert. Nach einem Waschschrift mit 70 % (v/v) Ethanol wurde die DNA in geeignetem Volumen ddH<sub>2</sub>O resuspendiert,

sobald der restliche Ethanol bei RT verdampft war. Optional wurde zusätzlich zur Isopropanolreinigung Glykogen eingesetzt, um höhermolekulare DNA effizient zu pelletieren (Glycogen Molecular Biology Grade, Fermentas). Aufkonzentrieren durch Eindampfen von nukleinsäurehaltigen Lösungen geschah in einem SpeedVac Vakuumpkonzentrator (Bachofer GmbH, Reutlingen).

### **2.1.2. Gelelektrophorese**

Die Größe der DNA-Fragmente aus Restriktionsverdau oder PCR-Amplifikationen wurde mit Agarose-Gelelektrophorese analysiert. Als Laufpuffer wurde Tris-Acetat-EDTA (TAE) Puffer verwendet (Endkonzentration aus 50x Stammlösung: 40 mM Tris-Acetat pH 8,0 und 2 mM EDTA). Ladepuffer (6x DNA Loading Dye) und DNA-Größenstandards (GeneRuler 50 bp, 1 kb und 1 kb plus DNA ladder) wurden lt. Herstellerangaben verwendet (Fermentas). Agarosegele mit 0,5 bis 2,0 % (w/v) trennten die unterschiedlich langen DNA-Fragmente bei 100 bis 170 Volt (Pharmacia EPS Netzteil, Uppsala, Schweden) auf. Nach 10-minütiger Färbung der Gele im Ethidiumbromid-Bad (5 mg/L Endkonzentration) und kurzem Waschen im Wasserbad, wurden die Nukleinsäuren im Gel unter UV-Licht sichtbar. Die Geldokumentation wurde mit einem AlphaImager Mini System (ProteinSimple, Santa Clara, USA) durchgeführt.

### **2.1.3. DNA-Quantifizierung**

Zur Analyse der DNA-Konzentration und -Qualität wurden je 1,5 µL der DNA-Lösungen im Nanodrop ND-1000 Spectrophotometer (NanoDrop Products, Wilmington, USA) eingesetzt. Der Grad der Verunreinigung der Probe durch RNA, Proteine sowie durch andere Kontaminationen konnte durch das Verhältnis der Absorptionen von 260 und 280 nm Wellenlänge abgeschätzt werden. Bei Gemischen aus mehreren DNA-Fragmenten in der Probe ließ sich die Ziel-DNA durch Agarose-Gelelektrophorese quantifizieren, wobei definierte Mengen des Markers (GeneRuler, Fermentas) zum Bestimmen der Nukleinsäuren verwendet worden sind.

## 2.2. Enzymatische Modifikationen

### 2.2.1. Polymerase-Kettenreaktion (PCR)

Die Vervielfältigung von DNA durch die Polymerase-Kettenreaktion diente zu Klonierung und Analyse der Reaktionsprodukte. Ein standardisierter PCR-Ansatz umfasste die Zugabe von Pufferkonzentraten (5x oder 10x), einem Mix von Desoxyribonukleosidtriphosphaten (dNTP-Mischungen bestehend aus cATP, cTTP, cCTP und cGTP), einem Paar flankierender Oligonukleotide (*forward/reverse primer*) komplementär zur Matrizen-DNA (*template DNA*) sowie die DNA-Polymerase. Das PCR-Programm umfasste 30 bis 35 aufeinanderfolgende Zyklen dieser Schritte: Denaturierung der DNA, Anlagerung der Oligonukleotide (*annealing*) bei einer definierten Temperatur oder im Temperaturbereich (Gradienten-PCR) in Abhängigkeit von der Schmelztemperatur ( $T_m$ ) des verwendeten Primer-Paares laut Vorhersage mit Sequenzmanagement-Software (Tabelle 14), und die enzymatische Verlängerung der Nukleinsäurestränge (*elongation*) in Abhängigkeit von Synthesegeschwindigkeit der verwendeten Polymerase und Länge des Zielfragments. Für die Analyse von DNA-Amplifikaten und Produktgrößen wurde die DreamTaq DNA Polymerase (Fermentas) verwendet. Für fehlerfreie Amplifikation kamen *proof-reading* Enzyme wie *Pyrococcus furiosus* (*Pfu*) DNA Polymerase (Fermentas) oder Phusion F530S (Finnzymes Oy, Espoo, Finnland) zum Einsatz, für lange und schwer zu amplifizierende DNA-Abschnitte wurde die LongAmp Taq PCR Polymerase (New England Biolabs) verwendet. Die PCR-Protokolle und Cycling-Bedingungen wurden laut Herstellerangaben durchgeführt. Die Ansätze wurden in einem BioRad MyCycler PCR Thermalcycler System amplifiziert. Die verwendeten PCR-Primer hatten eine Sequenzübereinstimmung von mindestens 18 Basen zur Zielsequenz, wobei ihr GC-Gehalt sofern möglich zwischen 40 - 60 % lag. Eine Liste der verwendeten Primer und Oligonukleotide (alle von Eurofins MWG Operon, Ebersberg, Deutschland) ist in Tabelle 42 zu finden.

In Abwandlung zur Standard-PCR diente die sogenannte Kolonie-PCR (*colony PCR*) zur schnellen Identifizierung rekombinanter *E. coli* Kolonien, bei denen die Klonierung erfolgreich war. Dabei wurden anstelle aufgereinigter Matrizen-DNA Zellmaterial einer Kolonie mit einem Zahnstocher abgepickt und mit dem PCR-Ansatz gemischt.

Eine weitere Vereinfachung der Klonierung zweier genomischer Flanken war mit der *Splicing by Overlap Extension-PCR* (SOE-PCR) möglich (Heckman und Pease, 2007). Mit Hilfe von vier verschiedenen Oligonukleotiden konnte die Amplifikation zweier Flanken in einer Standard-PCR durchgeführt werden. Da die Oligonukleotide an den Enden komplementär zur Sequenz der jeweils anderen Flanke sind, können die beiden Flanken in einem zweiten PCR-Ansatz endständig

miteinander hybridisieren. Mit den hybridisierten Flanken sowie Primern an den äußeren Enden kann die DNA-Polymerase schließlich das fusionierte DNA-Fragment amplifizieren. Die Methode fand vor allem bei der Erstellung von *Knockout*-Vektoren Verwendung.

### 2.2.2. Restriktionsverdau

Für übliche Klonierungen und zu analytischen Zwecken wurden kommerziell erhältliche Restriktionsenzyme (Fermentas und New England Biolabs) verwendet. Je nach eingesetzten Mengen an PCR-Produkten, genomischer und Plasmid-DNA wurde die Zusammensetzung des Restriktionsverdaus lt. Herstellerangaben angepasst. Verdaus mit mehreren Enzymen (sogenannte *double digests*) wurden in kompatiblen Puffern durchgeführt bzw. nach geeigneter Aufreinigung/vorheriger Inaktivierung des Enzyms nacheinander verdaut.

### 2.2.3. Dephosphorylierung und *blunting*

Linearisierte DNA Fragmente aus Restriktionsverdaus wurden mit Hilfe der FastAP Thermosensitive Alkaline Phosphatase (Fermentas) dephosphoryliert, um eine Religation der phosphorylierten Enden in Ligationen zu minimieren. In Falle der Ligation von inkompatiblen Nukleinsäureenden (beispielsweise durch zwei verschiedene Restriktionsenzyme mit nicht komplementären *sticky ends*) wurde der Fast DNA End Repair Kit (Fermentas) verwendet. Beim *blunting* durch T4 DNA-Polymerase und Klenow Fragment werden dabei unabhängig von der Art der Überhänge glatte Enden generiert, welche für nachfolgende Ligationen kompatibel sind. Selektive 5'-Phosphorylierung nicht-phosphorylierter PCR-Produkte (sowie 3'-Dephosphorylierung) wurde mit der T4 Polynukleotidkinase (T4 PNK) durchgeführt.

### 2.2.4. Ligationen

Klassische Ligationen zur kovalenten Bindung überlappender und glatter Fragmentenden wurden mit der T4 DNA Ligase (Fermentas) durchgeführt. Die molaren Verhältnisse aus Vektor:*Insert* lagen zwischen 1:3 und 1:5. Zusätzlich wurde die isothermale Assemblierungsmethode nach Gibson *et al.* (2009) für Ligation multipler DNA Fragmente verwendet (Gibson Assembly Master Mix, New England Biolabs). Durch die PCR-Primer wurden zwischen 15 und 21 Basen lange Sequenzen eingefügt, die jeweils homolog zu den Enden der zu ligierenden Fragmenten sind. Eine

Exonuklease generiert dabei 3'-überhängende ssDNA Enden, die jeweils komplementär zu einem der anderen Fragmentenden sind und hybridisieren können. Die hybridisierende DNA dient einer DNA Polymerase als Primer zum Auffüllen der einzelsträngigen Sequenzlücken, sodass die DNA Ligase schließlich die Lücken versiegeln kann. Einzelsträngige Überhänge (z.B. aus Restriktionsverdau) mussten dabei nicht weiter beachtet werden.

## 2.3. Transformation von Bakterien

### 2.3.1. Chemische Transformation von *E. coli*

Für die Präparation chemisch-kompetenter *E. coli*-Zellen wurde das Protokoll nach Inoue *et al.* (1990) mit einigen Abänderungen verwendet. Mehrere große Kolonien wurden von LB-Platte in 250 ml SOB-Medium (siehe Tabelle 8) inokuliert und in einem 1 Liter Schüttelkolben bei 19 °C schüttelnd kultiviert, bis die OD<sub>600nm</sub> etwa 0,5 betrug. Nach Inkubation der Zellen auf Eis und Pelletierung (4.000 × g, 4 °C) für jeweils 10 Minuten wurden sie in 80 mL eiskaltem TB-Puffer (Tabelle 8) resuspendiert.

**Tabelle 8: Lösungen zur Herstellung chemisch kompetenter Zellen.**

	Zusatz	Endkonzentration
<b>Super optimal broth (SOB):</b>		
	Hefextrakt	5,0 g/L
	Trypon	20 g/L
	NaCl	10 mM
	KCl	2,5 mM
	MgCl <sub>2</sub>	10 mM
	MgSO <sub>4</sub>	10 mM
	in ddH <sub>2</sub> O gelöst und autoklaviert	
<b>TB-Puffer:</b>		
	PIPES	10 mM
	CaCl <sub>2</sub>	15 mM
	KCl	250 mM
	In ddH <sub>2</sub> O gelöst, pH auf 6,7 eingestellt (KOH oder HCl), dann Zugabe	
	MnCl <sub>2</sub>	55 mM
	Lösung mit 0,45 µM Filter sterilisiert, Lagerung bei 4 °C für Monate	

Nach erneuten 10 Minuten Inkubation auf Eis und Pelletierung der Zellen wurden die Bakterien in 20 mL eiskaltem TB-Puffer und 1,4 mL DMSO resuspendiert, je 50  $\mu\text{L}$  pro Reaktionsgefäß aliquotiert und in Flüssigstickstoff weggefroren. Die kompetenten Zellen waren bei  $-70\text{ }^{\circ}\text{C}$  für mehrere Monate lagerbar.

Für die chemische Transformation wurden die kompetenten Zellen auf Eis aufgetaut, die Nukleinsäuren (Ligationsprodukte oder Plasmide) wurden mit der Suspension für 30 Minuten auf Eis inkubiert. Dem Hitzeschock bei  $42\text{ }^{\circ}\text{C}$  für 1 Minute folgte eine 3-minütige Inkubation auf Eis. Nach Zugabe von vorgewärmten SOC-Medium (SOB-Medium mit 20 mM steril-filtrierter Glucoselösung supplementiert) erfolgte eine Inkubation bei  $37\text{ }^{\circ}\text{C}$  für 1 Stunde, ehe die Zellen auf Selektionsmedien ausplattiert wurden.

### 2.3.2. Elektroporation von *E. coli*

Zur Erstellung elektrokompenter *E. coli*-Stämme wurde 1 L YENB-Medium (Tabelle 9) mit 5 mL ÜN Kultur inokuliert und für 3 bis 6 Stunden bei  $37\text{ }^{\circ}\text{C}$  geschüttelt. Die Zellen wurden bei einer optischen Dichte  $\text{OD}_{600\text{nm}}$  zwischen 0,5 bis 0,9 jeweils 10 Minuten auf Eis inkubiert und abzentrifugiert. Das Zellpellet wurde zweimal in 200 mL eiskaltem Wasser resuspendiert und gewaschen. Das Zellpellet wurde in 50 mL eiskalten 10 % (w/v) Glycerol resuspendiert und erneut abzentrifugiert. Die Zellen wurden schließlich in 2 mL 10 % Glycerol resuspendiert und 40  $\mu\text{L}$  pro Reaktionsgefäß aliquotiert und in Flüssigstickstoff weggefroren. Die Lagerung der elektrokompenten Zellen war dadurch für mehrere Monate bis Jahre bei  $-70\text{ }^{\circ}\text{C}$  möglich.

**Tabelle 9: Herstellung von YENB-Medium für elektrokompente Zellen.**

Zusatz	Endkonzentration
<b>YENB-Medium:</b>	
Bacto Yeast Extract	7,5 g/L
Bacto Nutrient Broth	8,0 g/L
in ddH <sub>2</sub> O gelöst und autoklaviert	

Die elektrokompenten Zellen wurden auf Eis aufgetaut und mit wenigen  $\mu\text{L}$  der aufgereinigten (entsalzten) Ligationsreaktion gemischt. Die Elektroporation wurde in einer Elektroporationsküvette

mit einem Elektrodenabstand von 1 mm (PeqLab Biotechnologie GmbH, Erlangen) mit einem Bio-Rad Gene Pulser II Electoporation System durchgeführt. Die verwendeten Elektroporations-Parameter waren  $2,5 \text{ kV mm}^{-1}$ ,  $25 \text{ }\mu\text{Farad}$  und  $200 \text{ Ohm}$  mit Zeitkonstanten um  $5 \text{ ms}$ . Nach erfolgtem Puls wurde  $1 \text{ mL}$  vorgewärmtes SOC-Medium zu den Zellen gegeben und die Suspension eine Stunde schüttelnd bei  $37 \text{ }^\circ\text{C}$  inkubiert, ehe der Transformationsansatz auf Selektionsmedien ausplattiert wurde.

### **2.3.3. Natürliche Transformation von *T. thermophilus* HB27**

Die Transformation von *T. thermophilus* erfolgte nach etablierten Protokollen (Angelov *et al.*, 2009). Eine  $5 \text{ mL}$  Vorkultur wurde ÜN bei  $70 \text{ }^\circ\text{C}$  inkubiert und zur Inokulation der Hauptkultur verwendet ( $5,0 \text{ \% v/v}$ ). Nach Erreichen der exponentiellen Wachstumsphase ( $\text{OD}_{600\text{nm}}$  bei  $0,7$ ) wurde die Kultur 1:1 mit frischem TB-Medium verdünnt und eine Stunde im  $15 \text{ mL}$  Falcon-Tube bei  $180 \text{ Upm}$  und  $70 \text{ }^\circ\text{C}$  inkubiert. Nach Zugabe von DNA (Nukleinsäuren in reinem  $\text{ddH}_2\text{O}$ , maximal  $35 \text{ }\mu\text{L}$  Gesamtvolumen) wurde der Transformationsansatz zwei weitere Stunden kultiviert und anschließend auf Selektionsmedien ausplattiert.

## **2.4. Strategien für genetische Manipulationen in *T. thermophilus* HB27**

### **2.4.1. Transposon-Mutagenese**

Für die Durchführung ungerichteter Gendeletionen in *Thermus* wurden *in vivo* und *in vitro* Transposonmutagenesen durchgeführt. Die *in vivo* Methode basierte auf der für *Bacillus* etablierten Transposonmutagenese mittels pMarB-Vektoren, welche eine hyperaktive *HimarI* Transposase der Hornfliege exprimieren können (Le Breton *et al.*, 2006). Die Insertionssequenz ‚TA‘ im Chromosom führt zur Verankerung des Tn $YLB-1$  Transposons. Ausgehend vom pMarB Vektor wurde das System für *Thermus* adaptiert. Als *in vitro*-Methode kam die EZ-Tn5™ Transposase und EZ-Tn5™ pMOD™3<R6K $\gamma$ ori/MCS> Transposon Construction Vector (Epicentre) zum Einsatz.

### **2.4.2. Gezielte Genunterbrechungen**

Wird ein nicht-replikativer Vektor (*Knockout*- oder *Suicide*-Plasmid) mit Selektionsmarker (z.B. Antibiotika-Resistenz) und einem zum genetischen Zielort identischen Sequenzabschnitt in den Wirt

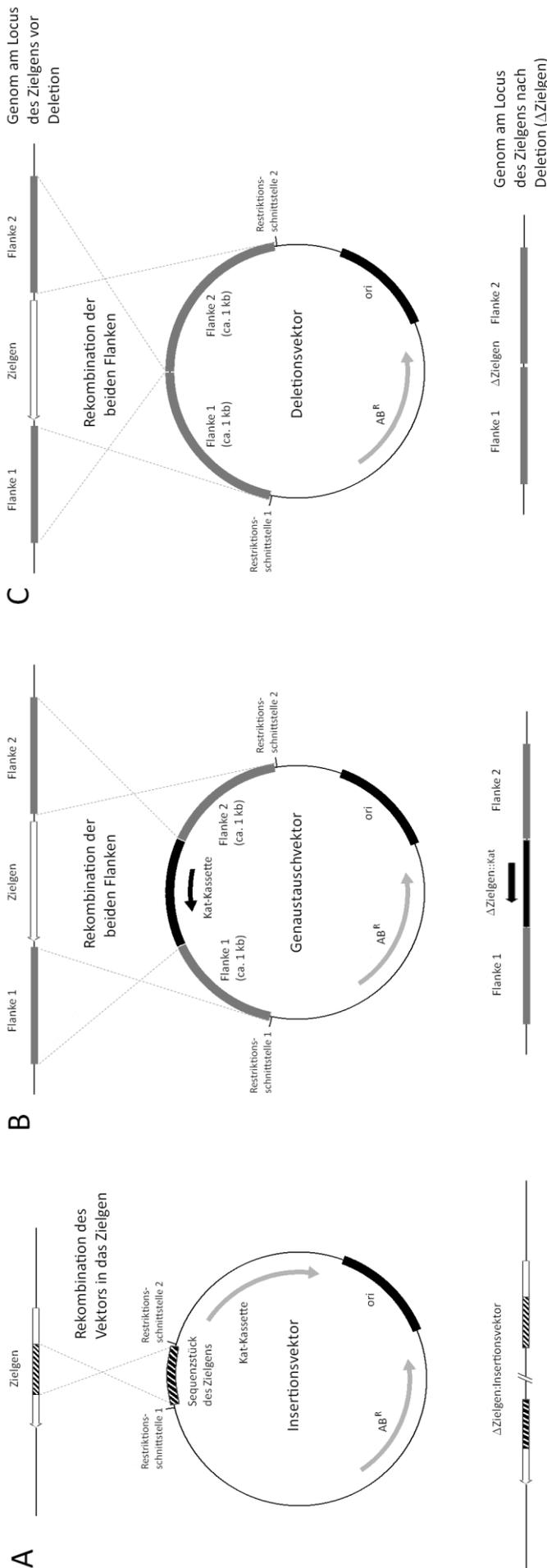
eingbracht, erfolgt über homologe Rekombination eine persistente Integration des Vektors unter Selektionsdruck (z.B. durch das Antibiotikum), siehe Abbildung 3A. Durch den Rekombinationsvorgang im Zielgen kann es dadurch zu einer Genunterbrechung kommen, beispielsweise durch Unterbrechung des Leserahmens durch den Vektor oder durch das Einbringen von gezielten Mutationen mit der integrierten DNA. Auf Basis dieser Strategie wurden mehrere potentielle Esterase-kodierende Kandidatengene in *Thermus* gezielt durch die Integration eines nicht-replikativen Vektors pBKat (A. Wiezer, Diplomarbeit, 2001) unterbrochen. Die Genabschnitte der Zielgene wurden in die multiple Klonierungsstelle (*multiple cloning site, MCS*) des Vektors kloniert (Zusammenfassung aller Konstrukte im Anhang, Tabelle 39). Nach Übertragung der *Knockout*-Plasmide in *Thermus* erfolgte die Selektion der Integrierten mittels Kanamycin-Selektion. Mindestens 10 Mutanten wurden in 5 mL TB Medium unter Selektionsdruck über Nacht bei 60 °C inkubiert und jeweils 5 – 10 µL der Kulturen auf TB Platten mit 1,0 % (v/v) Tributyrin aufgetropft. Die lipolytische Aktivität der Mutanten wurde nach mehrtägiger Inkubation untersucht, indem die Reduktion des Durchmessers des hydrolytischen Hofes um die Kolonien gegenüber einer Wildtyp-Kontrolle verglichen wurde. Zur Überprüfung der erfolgreichen Integration des Vektors in das Chromosom wurde für jede Mutante eine analytische PCR durchgeführt (Zusammenfassung im Anhang, Tabelle 40).

#### 2.4.3. Genaustauschmutanten und Deletionen

Positioniert man den Selektionsmarker zwischen die Flanken vor (*upstream*) und nach (*downstream*) dem Zielgen (*target gene*), erhält man durch homologe Rekombination beider Einzelflanken Genaustausch-Mutanten (Abbildung 3B). Mit Hilfe dieser Strategie wurden weitere Kandidatengene durch Kanamycin-Selektion vollständig durch die Antibiotika-Kassette ersetzt. Als Vektorgrundgerüst für die Deletion der beiden ORFs TT\_C0904 und TT\_C1787 diente pS18a (Blas-Galindo *et al.*, 2007), für die beiden ORFs des Operons TT\_C0340-1 wurde pCR2.1-XL-TOPO (Invitrogen, Carlsbad, USA) verwendet. Dabei wurden je 1 kbp lange Flanken der Zielsequenzen kloniert (Tabelle 43) und mit einer dazwischenliegenden Kanamycin-Kassette versehen. Die phänotypische Charakterisierung der Mutanten geschah wie oben beschrieben. Durch die Integration des Selektionsmarkers verbleibt dieser auch im Chromosom und kann nicht wieder für die Selektion weiterer Genmanipulationen verwendet werden. Dadurch reduziert sich die Nutzung eines Selektionsmarkers auf den Austausch eines einzelnen genomischen Locus, sofern der Marker nicht gezielt entfernt werden kann (beispielsweise durch Rekombinasen). Prinzipiell ließe sich aber für jeden weiteren Selektionsmarker jeweils ein weiteres Gen deletieren. Für die Konstruktion einer Mehrfachmutante in *Thermus* ist diese Strategie allerdings nicht praktikabel, da es bis heute nur eine begrenzte Anzahl etablierter, thermostabiler Marker gibt (z.B. Resistenzmarker gegenüber Bleomycin, Hygromycin).

#### 2.4.4. Gendeletionen ohne verbleibenden Marker

Durch die Vorarbeiten von H. Li (Lehrstuhl für Mikrobiologie) konnte bereits gezeigt werden, dass bei der Übertragung eines Selektionsmarker-freien Deletionsvektors in *T. thermophilus* markerlose Deletionsmutanten der  $\beta$ -Glucosidase ( $\Delta bgl$ , TT\_P0042) mit einer Häufigkeit von etwa 1% auftreten (unveröffentlichte Daten). Obwohl die Grundlagen dieses Phänomens (insbesondere im Zusammenhang mit der Polyploidie von *Thermus*) momentan noch Gegenstand weiterer Untersuchungen ist, wurde das System als neuartiges Deletionssystem zur Generierung multipler Gendeletionsmutanten verwendet. Für diese Strategie wurden pUC18 (für ORFs TT\_C1787 und TT\_C0904) und pCR2.1-XL-TOPO Vektor (ORFs TT\_C0340 bis TT\_C0341) verwendet, in welche etwa 1 kb lange Sequenzflanken der jeweiligen Zielgene wie bei Punkt III 2.4.3 kloniert wurden (schematische Abbildung 3C). Nach vollständiger Sequenzierung der Konstrukte wurden bis zu 15  $\mu$ g Plasmid-DNA über natürliche Kompetenz in *Thermus* übertragen, und der Transformationsansatz auf TB Festmedien ausplattiert. Von allen gewachsenen Klonen wurden mindestens 200 bis 300 Kolonien mit einem Zahnstocher gepickt und in 96-Deepwell Blöcke mit 1,1 ml TB Flüssigmedium überimpft und für zwei Tage bei 70 °C schüttelnd inkubiert. Kombinatorische gDNA Präparationen von je 50  $\mu$ L pro Reihe (z.B. Reihe A bestehend aus Position A1, A2, A3,... A12) sowie pro Spalte (z.B. Reihe 12 bestehend aus A12, B12, C12, ... H12) des Deepwell-Blockes einzeln zusammengeführt (Pools). Aus den 20 resultierenden Mischungen wurde die genomische DNA isoliert (MasterPure Complete DNA and RNA Isolation Kit) und mittels PCR-Detektion auf die Präsenz einer Mutante analysiert (siehe Beispiel Abbildung 13). Durch die kombinatorische Herangehensweise konnte aus positiven Zeilen (z.B. Zeile D) und Spalten (z.B. Nummer 8) die exakte Position im Block bestimmt werden (in diesem Beispiel Position D8). Von den Verdünnungsausstrichen aus den positiven Wells wurden mehrere Kolonien erneut per PCR auf homozygote Mutanten untersucht. Mittels Southern Blot und Aktivitätsmessungen konnte das Ergebnis verifiziert werden.



**Abbildung 3: Grundlegende Strategien zur Erstellung von gezielten genetischen Manipulationen zur Deaktivierung, zum Austausch oder zur vollständigen Deletion der Zielgene in *T. thermophilus*.** Die chromosomale Situation am Ziellocus ist jeweils über dem Vektor abgebildet, die Situation nach Gendelektion darunter. Gezeigt sind die Klonierungsstellen über Restriktionsschnittstellen der multiplen Klonierungsstelle (MCS) der jeweiligen Vektoren sowie die für *E. coli* notwendigen Elemente zur Selektion (Antibiotika-Resistenz AB<sup>R</sup>) und Mobilisierung (*ori*). A: Auf Basis des Insertionsplasmids pBKat (A. Wiezer, Diplomarbeit) konnten Sequenzabschnitte mit Homologien zu den putativen Esterasen ermittelt werden. Nach Transformation des Wirts mit dem Plasmid konnte über Kanamycin-Selektion die Insertion des Plasmids in die kodierende Gensequenz vermittelt werden. Dabei waren zwei Möglichkeiten zur Integrationsrichtung des Plasmids möglich (ΔZielgen:Insertionsvektor). B: Die Klonierung von zwei etwa 1 kbp langen Flanken, welche homolog zum Zielgen waren, dienten zur Herstellung eines Genaustauschvektors. Dabei wurde zwischen die Flanken eine Kanamycin-Resistenzkassette als selektierbarer Marker kloniert. Über doppelte homologe Rekombination am Locus des Zielgens konnten unter Selektionsdruck Genaustauschmutanten erstellt werden, welche statt dem Zielgen lediglich die *kat*-Kassette trugen (ΔZielgen::*kat*). C: Zur Generierung eines markerlosen Deletionsvektors wurden die oberen und unteren Flanken der Zielgene (je etwa 1 kbp lang) ligiert und der erhaltene Vektor sequenziert, überprüft und *T. thermophilus* damit transformiert. Aufgrund noch nicht vollständig aufgeklärter Rekombinationsvorgänge waren damit bis 1,0 % der gewachsenen Kolonien nach der Transformation markerlose Deletionsmutanten (ΔZielgen).

## 2.5. Southern Blot-Analyse

Die Analyse spezifischer genetischer Manipulationen in *Thermus* wurde mit Hilfe von DNA-DNA-Hybridisierung (Southern Blot) durchgeführt (Prinzip nach Southern, 1975, mit Modifikationen). Je 5 bis 10 µg genomische DNA von modifizierten und nicht-modifizierten Stämmen wurde mit geeigneten Restriktionsenzymen verdaut, um unterscheidbare DNA-Fragmente für die Hybridisierungen und Detektion zu erhalten (Tabelle 10). Nach Verdau über Nacht wurden die verdauten Fragmente in einer gekühlten Gelkammer (HE33 Mini Submarine Electrophoresis Unit, Hoefer, Holliston, USA) bei 80 V getrennt. Der Transfer der DNA vom Gel auf eine Nylonmembran HyBond N (Amersham Pharmacia) erfolgte im Vakuumblotter (Model 785 Vacuum Blotter, BioRad) unter Verwendung einer passgenauen Maske zum Gel. Zum Kitten von Rissen im Gel und zum Abdichten auf der Maske wurde 2,0 % Agarose verwendet. Unter Vakuum (etwa 5 bis 8 Hg) wurden die Blotting-Lösungen nacheinander (siehe Zusammensetzung und Ablauf des Blottingvorgangs in Tabelle 10) und auf das Gel gegossen. Nach dem Transfervorgang wurde die Membran unter UV-Licht bei 254 nm für 1 Minute fixiert.

**Tabelle 10: Arbeitsschritte für Southern Blot-Prozedur.**

Prozedur/Arbeitsschritt	Verwendete Lösungen	Zusammensetzung	Dauer
Depurinierung	Depurinierungslösung	0,25 M HCl	10 min
Waschen mit Wasser (2x)			
Denaturierung (2x)	Denaturierungslösung	1,5 M NaCl, 0,5 M NaOH	15 min
Waschen mit Wasser (2x)			
Neutralisierung (2x)	Neutralisierungslösung	0,5 M Tris-HCl, 1,5 M NaCl, pH 7,5 (HCl)	15 min
Waschen mit Wasser			
DNA Transfer	Natriumchlorid-Natriumcitrat (SSC) Transferlösung (20x SSC pH 7,0)	3,0 M NaCl, 0,3 M Na <sub>3</sub> Citrat, pH 7,0 (HCl)	120 min

Mindestens 1 µg der DNA-Sonden wurde mit Biotin DecaLabel DNA Labeling Kit (Fermentas) nach Herstellerangaben markiert. Als Hybridisierungslösung (5 mL Gesamtvolumen für die Hybridisierung eines Blots) dienten 1,5 mL 20x SSC, 0,5 mL 50x Denhardt's Solution, 0,25 mL 10% (w/v) SDS,

2,5 mL 100% (v/v) deionisiertes Formamid und 0,25 mL ddH<sub>2</sub>O. Die Biotin-markierte Sonde wurde in siedendem Wasser für 5 Minuten denaturiert und auf Eis abgekühlt. Die Sonde wurde dann zur Hybridisierungslösung gegeben und über Nacht bei 42 °C mit dem Blot in einem Hybridisierungssofen inkubiert. Die Membran wurde jeweils zweimal mit 2x SSC, 0,1 % (w/v) SDS für 10 Minuten bei Raumtemperatur gewaschen, nochmals zweimalig mit 0,1x SSC, 0,1 % (w/v) SDS bei 65 °C für 20 Minuten inkubiert, und schließlich zur Detektion der Biotin-markierten Sonde getrocknet. Die Detektion wurde mit dem Biotin Chromogenic Detection Kit (Fermentas) und für Chemolumineszenz mit dem CDP-Star Reagent (New England Biolabs) nach Herstellerangaben durchgeführt.

### 3. Konstruktion und funktionelle Analyse von Genbanken

#### 3.1. Verwendete Genbanken und Umweltproben

Für die funktionelle Metagenomanalyse standen verschiedene Genbanken in *E. coli* von Kooperationspartnern zur Verfügung (Tabelle 11). Eigene *E. coli* Genbanken wurden aus chromosomaler DNA von *T. brockianus* (S. Blank, Technische Universität Hamburg-Harburg) sowie aus Umwelt-DNA aus heißen Quellen von den Azoren erstellt und einzeln oder in Pools in *T. thermophilus* übertragen (Tabelle 12).

**Tabelle 11: Liste der durchmusterten Metagenombanken in *E. coli*.**

Probenbezeichnung	Größe der Metagenombank (Art des Fosmids) <sup>1</sup>	Herkunft
Elbe-Sedimente (ES)	960 in Mikrotiterplatten (in pCC1FOS)	W. Streit (Universität Hamburg)
Elephanten-Dung (ED)	960 in Mikrotiterplatten (in pCC1FOS)	
<i>Teredo navalis</i> (TN) (Anreicherungen auf CMC)	960 in Mikrotiterplatten (in pCC1FOS)	
Anreicherungen von Tiefseeproben (IFREMER, Brest, Frankreich):		
Mesophile Bakterien (MB)	~ 200.000 gepoolte Klone (in pCC1FOS)	P. Golyshin
Thermophile Bakterien (TB)	~ 200.000 gepoolte Klone (in pCC1FOS)	(Bangor University, UK)
Hyperthermophile Archaeen (HA)	~ 20.000 gepoolte Klone (in pCC1FOS)	
Kompost Miete II, 4. Tag, 63,3 °C	1.920 in Mikrotiterplatten (in pCT3FK)	M. Mientus (Lehrstuhl für Mikrobiologie, TUM)

<sup>1</sup>: Fosmid pCC1FOS ist lediglich kompatibel für *E. coli* als Expressionswirt, das Shuttle-Fosmid pCT3FK (Angelov *et al.*, 2009) hingegen kann für beide Wirte *T. thermophilus* und *E. coli* genutzt werden.

**Tabelle 12: Verfügbares Probenmaterial, das für die Erstellung der Genbanken verwendet wurde.**

Proben	Weitere Informationen	Herkunft
<i>Thermus brockianus</i> T17 chromosomale DNA	Stamm ist zu Wachstum auf diversen pflanzlichen Polysacchariden befähigt	K. Sahn, G. Antranikian (Technische Universität Hamburg-Harburg)
Umweltproben von heißen Quellen der Azoren	Wasser, Sedimente und Biofilm Probenmaterial, Azoren (Portugal)	L. Gonçalves, H. Santos (Universität Lissabon, Portugal)

### 3.2. Erstellung von Fosmidbanken aus (meta)genomischer DNA

Die Klonierung großer (meta)genomischer DNA-Fragmente in Fosmide bietet die Möglichkeit, größere Sequenzabschnitte (etwa 40 kb) pro Klon funktionell untersuchen zu können. Voraussetzung für die Klonierung dieser sogenannten *large-insert* Genbanken ist, entsprechend lange und intakte DNA-Fragmente aus Mikroorganismen und Umweltproben zu erhalten. Für die Konstruktion der genomischen *T. brockianus* Genbank sowie der metagenomischen Banken aus heißen Quellen der Azoren, wurde das CopyControl™ Fosmid Library Production Kit von Epicentre (Madison, USA) verwendet. In Änderung zu den Herstellerangaben wurde anstelle des pCC1FOS das Fosmid pCT3FK (Angelov *et al.*, 2009) verwendet, welches neben den genetischen Elementen des pCC1FOS zusätzliche Selektions- und Mobilisierungselemente für eine Übertragung der Genbank in *T. thermophilus* besitzt ( $P_{slp}$  Promotor und thermostabile Kanamycin-Adenyltransferase des *E. coli*-*Thermus Shuttle*-Plasmids pMK18 und chromosomale Sequenzelemente des *pyr*-Locus in *Thermus*). Die Konstruktion der Fosmidbank umfasste drei grundlegende Schritte: Die Klonierung der endständig reparierten (*blunting*), hochmolekularen und (meta)genomischen DNA in den *PmlI* (*Eco72I*)-linearisierten pCT3FK-Vektor, das Verpacken des Ligationsansatzes in  $\lambda$ -Phagen und die Infektion der Phagenresistenten EPI300 *E. coli*-Zellen (alles nach Herstellerangaben). Nach diesen Prozeduren wurden die transduzierten Fosmidklone auf LB mit Chloramphenicol (12,5  $\mu\text{g}/\text{mL}$ ) ausplattiert und mindestens 10 Einzelkolonien mittels analytischem Restriktionsverdau mit *EcoRI* und *BamHI* auf ihre Diversität hin analysiert. Nach erfolgreicher Überprüfung wurden je nach Genbankgröße mehrere hundert bis tausend gewachsene Kolonien in 96-Well Mikrotiterplatten übertragen. Jede Kolonie entspricht damit einer spezifischen Position in der Mikrotiterplatte. Die mit 180  $\mu\text{L}$  LB-Medium und Chloramphenicol (12,5  $\mu\text{g}/\text{mL}$ ) über Nacht bei 37 °C inkubierten Mikrotiterplatten wurden schließlich zur Langzeitlagerung mit 20  $\mu\text{L}$  DMSO versetzt, mit einer sterilen Plastikklebefolie (tape pads 50, Qiagen, Hilden) verschlossen und bei – 80°C weggefroren.

### 3.3. Erstellung von Plasmid-Genbanken

#### 3.3.1. Subklonieren von Fosmid-DNA

Im Falle nicht ausreichender Sequenzabdeckung bei der Sequenzierung wurden einzelne Fosmide in *shotgun*-Genbanken subkloniert. Dabei wurde die Fosmid-DNA durch wenige Ultraschall-Pulse physikalisch fragmentiert, und gezielt 1,5 bis 3 kb große DNA-Stücke nach gelektrophoretischer Trennung aufgereinigt. Die DNA-Enden wurden mit dem End-It™ DNA End-Repair Kit (Epicentre) behandelt. Das Anfügen von 3'-dA-Überhängen (*A-tailing*) wurde mit der DreamTaq DNA Polymerase (Fermentas) laut Herstellerangaben ohne Primer und lediglich unter Zugabe von 0,2 mM dATP bei 72 °C für 30 Minuten durchgeführt. Die Klonierung der *Insert*-DNA in den pCR2.1-XL-TOPO Vektor (Invitrogen) wurde laut Herstellerangaben durchgeführt.

#### 3.3.2. 16S rDNA Sequenzen

Die Metagenomproben wurden mit den Oligonukleotiden 616VALT und 100K für bakterielle 16S ribosomale DNA-Sequenzen sowie dem Primerpaar AC165 und AC1601 spezifisch für archaeelle Sequenzen analysiert (Oligonukleotidsammlung, Lehrstuhl für Mikrobiologie, AG Ludwig, TU München, persönliche Mitteilung). Die PCR-Produkte wurden in den CloneJET PCR Cloning Kit (Fermentas) zur Konstruktion einer 16S-Genbank verwendet (laut Herstellerangaben). Einzelne Kolonien wurden von der Agarplatte in Mikrotiterplattenformat überführt und mit den flankierenden Primern pJet1-FP und pJet1-RP bei GATC Biotech (Konstanz) sequenziert.

### 3.4. Funktionelle Analyse und Aktivitätsnachweise der Genbanken

#### 3.4.1. Übertragung der Fosmidbanken in *T. thermophilus*

Zu Durchmusterungszwecken wurden für jede Mikrotiterplatte der in *E. coli* erstellten Fosmidbanken eins zu eins temporäre Genbanken in *T. thermophilus* generiert (Angelov *et al.*, 2009), wobei jeder einzelne Fosmidklon pro Platte einer Transformation des anderen Wirts entspricht. Dazu wurden pro Mikrotiterplatte (entsprechend 96 *E. coli*-Klone) zwei Deepwell-Blöcke mit je 1,2 mL 2x LB-Medium und Fosmid Autoinduktions-Lösung (AIS) pro Vertiefung inokuliert und für zwei Tage bei 37 °C in

Plattenhalterungen fixiert bei 375 Upm geschüttelt. Zur Zellernte wurden die Blöcke für 10 Minuten bei 4.000 Upm zentrifugiert und der Überstand durch ruckartiges Invertieren der Platten abgegossen. Die Fosmidpräparation geschah in Analogie bereits zum beschriebenen alkalischen Lyseprotokoll (siehe III.2.1.1.), mit folgenden Abweichungen für jeden Deepwell-Block: Die Zellpellets in den beiden Blöcken wurden in je 200 µL Resuspensionspuffer mittels Vortex-Kreisschüttler (IKA, Staufen) resuspendiert, dann mit je 200 µL Lysepuffer pro Well versetzt und durch kurzes Aufstoßen des Blocks durchmischt. Nach der Zugabe von weiteren 200 µL Neutralisationspuffer wurden beide Blöcke bei 4.000 Upm in einer Sorvall RC 3B Plus-Zentrifuge mit H-6000A/HBB-6 Rotor (Thermo Scientific) für 30 Minuten zentrifugiert. Von beiden Blöcken wurde dann je 400 µL des Überstand pro Vertiefung abpipettiert und in einem neuen Block mit 1/10 Volumen 3,0 M Natriumacetat (pH 5,5) und 0,7 Volumen Isopropanol versetzt. Nach vorsichtigem Schütteln des Blocks wurde die präzipitierte DNA für eine Stunde bei 4 °C und 4.000 Upm abzentrifugiert. Nach Abgießen des Überstandes wurde das Pellet in 70 % (v/v) Ethanol gewaschen und erneut kurz abzentrifugiert. Nach Trocknen der Pellets an der Luft wurde die DNA in 35 µL ddH<sub>2</sub>O resuspendiert. Zur Übertragung der Fosmide in *T. thermophilus* wurde schließlich eine 150 mL Kultur bei 70 °C bis zum OD<sub>600nm</sub>-Wert von 0,7 in die exponentielle Wachstumsphase gebracht und 1:1 mit frischem Medium verdünnt. Nach einer Stunde Inkubation wurde 1,0 mL der Suspension in jede Vertiefung des Deepwell-Blocks zur vorgelegten DNA zugegeben. Nach einer weiteren Stunde Inkubation unter Schütteln des Blocks wurden 100 µL jedes einzelnen Transformationsansatzes in einen frischen Block mit je 1,0 mL vorgelegtem TB-Medium mit 20 µg/mL Kanamycin überführt und bei 375 Upm für mehrere Tage bei 60 °C geschüttelt.

### 3.4.2. Direkte Detektion

Die direkte phänotypische Durchmusterung der Genbanken fand auf Substratplatten und in flüssigen Substratformulierungen im Mikrotiterplattenformat statt. Tabelle 13 fasst alle getesteten, kommerziell erwerblichen und selbst hergestellten Substrate und ihre jeweilige Detektionsmethode zusammen. Zur Herstellung selbst gefärbter Substrate wurde der Farbstoff nach beschriebenen Methoden mit Xylan und CMC kovalent verknüpft (Lee und Lee, 1997; Ten *et al.*, 2005). Dazu wurden je 2 g des Substrates in 30 mL destilliertem Wasser, 10 mL 2,0 M NaOH, 1,9 g Cibacron Brilliant Red 3B-1 und 1,2 mL 1,4-Butandiglycidylether gemischt, für 5 Minuten gerührt und bei Raumtemperatur für 48 Stunden getrocknet. Das daraus entstandene Gel wurde mit 100 mL ddH<sub>2</sub>O gemischt und mit einem Pürierstab zerkleinert. Die Partikel wurden mit kochend heißem Wasser mittels Whatman Papier mehrmals abfiltriert, bis das Filtrat farblos wurde. Das feuchte Produkt wurde in aufsteigenden Ethanolkonzentrationen (25, 50, 75, 100 %) und abschließend mit Aceton gewaschen und Luft-

getrocknet. Das fertige Substrat konnte schließlich als Zusatz für Agarmedien, Overlay oder als chromogenes Substrat in Mikrotiterplatten verwendet werden.

**Tabelle 13: Liste der Substrate für funktionelle Durchmusterungen.**

Verwendete Substrate	Endkonzentration	Herkunft	Detektionsmethode
AZO-Avicel	0,05 %	Megazyme	Overlay, Mikrotiterplatte
AZCL-Galactomannan	0,05 %	Megazyme	Overlay, Mikrotiterplatte
AZCL- $\beta$ -Glucan (Barley)	0,05 %	Megazyme	Overlay, Mikrotiterplatte
AZCL-Hydroxyethylcellulose	0,05 %	Megazyme	Overlay, Mikrotiterplatte
AZCL-Xylan	0,05 %	Megazyme	Overlay, Mikrotiterplatte
Gelabeltes Xylan (Cibacron red)	0,05 %	nach Ten <i>et al.</i> (2005)	Mikrotiterplatte
Carboxymethylcellulose (CMC) Natriumsalz (geringe Viskosität) <sup>2</sup>	0,1 %	Sigma-Aldrich	Agarplatte, Mikrotiterplatte
Gelabeltes CMC (Cibacron red)	0,05 %	nach Ten <i>et al.</i> (2005)	Mikrotiterplatte
Xylan (Hafer) <sup>2</sup>	0,1 %	Fluka	Agarplatte
Avicel PH-101 <sup>2</sup>	0,1 %	Fluka	Agarplatte
Stärke, löslich <sup>3</sup>	0,3 %	Sigma-Aldrich	Agarplatte
Magermilch (skimmed milk)	2,0 %	Difco	Agarplatte
Tributyryn	1,0 % (v/v)	TCI-Europe/Fluka	Agarplatte
Triolein (mit Rhodamin B, nach Kouker und Jaeger, 1987)	1,0 % (v/v)	Sigma-Aldrich	Agarplatte
Tributyryn in SH-Minimalmedium <sup>1</sup>	1,0 % (v/v)	TCI Europe/Fluka	Agarplatte
Olivenöl	1,0 % (v/v)	Sigma-Aldrich	Agarplatte

<sup>1</sup>: nur für Durchmusterungen in *T. thermophilus* möglich.

<sup>2</sup>: Nachweis des Polysaccharidabbaus mittels Kongorotfärbung (Färbelösung aus 0,1 % (w/v) Kongorot für mehrere Minuten einwirken lassen, Entfärbung durch mehrmaliges Waschen mit 1 M NaCl).

<sup>3</sup>: Nachweis des Stärkeabbaus mittels Lugol'scher Lösung (1 g Iod, 2 g Kaliumiodid in 300 mL ddH<sub>2</sub>O).

Beim Platten-*Screening* wurden die Fosmidbanken in *E. coli* direkt auf die LB-Festmedien (mit den jeweiligen Substraten und Fosmid Induktorlösung) gestempelt und bei 37 °C kultiviert. Zur effizienteren Lyse der Zellen wurden die Platten verdampfendem Toluol ausgesetzt. Als weitere Strategie wurden angewachsene Klone auf LB-Medien mit ca. 10 mL Substrat-Overlay-Agar überschichtet. In einem weiteren Durchmusterungsansatz wurden 200  $\mu$ L von ÜN kultivierten *E. coli* Fosmidbanken aus 96-Deepwell-Blöcken in 96-Well Mikrotiterplatten überführt und pelletiert (4.000  $\times$  g, 4 °C, 10 Minuten). Die Zellpellets wurden durch wiederholtes Schockgefrieren in

flüssigem Stickstoff und direkt anschließendem Auftauen bei Raumtemperatur lysiert. In *T. thermophilus* wurde die Freisetzung von Protein aus dem gewonnenen Rohextrakt mittels Bradford bestimmt, um die optimalen Lysebedingungen zu ermitteln ( $9,8 \pm 0,2$  mg Protein pro mL<sup>-1</sup> Kulturvolumen und pro Schockgefrier-Auftau-Zyklus). Für die Zelllyse in den Mikrotiterplatten wurden drei Zyklen durchgeführt, was in etwa 25 % des Gesamtproteingehalts gegenüber mit Ultraschall behandeltem Zellextrakt beträgt.

Die lysierten Kolonien und Zellpellets wurden für die funktionellen Metagenomanalysen über alle relevanten Temperaturspektren (psychrophil, mesophil und thermophil) bei 8 °C, RT (20 bis 25 °C), 37 °C sowie bei 60 und 70 °C für mindestens 2 Tage oder länger inkubiert. Aktive Klone waren durch Farbstofffreisetzungen im Agar oder der Überstände sowie durch Bildung von hydrolytischen Höfen um die Kolonien sichtbar.

Nach deren Identifizierung wurden die Fosmide der aktiven Kandidatenklone isoliert und neu übertragen, um falsch-positive Ergebnisse durch Kontaminationen ausschließen zu können. Konnte die beobachtete Aktivität durch das Fosmid wieder hergestellt werden, wurde der betreffende Klon einzeln gesichert. Ein analytischer Verdau der positiven Fosmide zeigte, ob ein Klon bereits einmal aufgetreten ist, bzw. Sequenzteile identisch zu anderen positiven Fosmidklonen sind (Ausschluss von Redundanzen und Kreuzkontaminationen der Genbank). Nur reproduzierbare, nicht-redundante und kontaminationsfreie Fosmidklone wurden sequenziert. Fosmide, die nicht oder unvollständig sequenziert worden sind, bzw. vollständig sequenzierte Fosmide ohne die bioinformatische Identifizierung von Kandidatengenen wurden entweder durch *shotgun*-Klonierung oder aber durch Transposon-Mutagenese (*loss-of-function*) analysiert.

### 3.4.3. Heterologe Komplementationsanalyse

Neben der direkten Detektion von aktiven Fosmid-Klonen konnte in *T. thermophilus* eine Methode für die Durchmusterung nach Esterasen angewendet werden, die auf heterologer Komplementation beruht. Bei diesem Prinzip handelt es sich um die Wiederherstellung eines genetischen Defekts einer Mutante, die unter bestimmten Bedingungen nicht wachstumsfähig ist. Durch die Präsenz der heterolog exprimierten metagenomischen DNA ist es dann möglich, diesen Wachstumsdefekt zu komplementieren. Dieses Prinzip wurde in einer Hochdurchsatzmethode mit Hilfe des Esterase-negativen *Thermus* Stammes BL03 (diese Arbeit) angewendet. Dabei wurden die Kolonien einer Agarplatte mit je 48 einzelnen Fosmidklonen (ein halbiertes 96-Well Block) mit 50 mM Phosphatpuffer (pH 7,5) abgewaschen und von dieser Suspension eine Fosmidpräparation mittels Isopropanol-Fällung durchgeführt. Von dem isolierten DNA-Gemisch wurden über 10 µg DNA in maximal 35 µL Gesamtvolumen in *T. thermophilus* BL03 übertragen (standardisierter

Transformationsansatz mit 1 mL Gesamtvolumen) und die Zahl der Transformanten auf TB mit Kanamycin (20 µg/mL) nach 2 Tagen Inkubation bei 60 °C bestimmt. Bei mehr als 300 Kolonien pro Transformation entspricht die stochastische Wahrscheinlichkeit dafür, dass alle der 48 transformierten Fosmidklone mindestens einmal als Kolonie repräsentiert sind, mehr als 99,8 % (Berechnung laut binomialer Verteilungsfunktion mit Wahrscheinlichkeit  $p = 0,0208$  und Koloniezahl  $n = 300$ ). Die rekombinanten Kolonien wurden von den TB-Platten abgewaschen und in Phosphat-Puffer resuspendiert, bei  $4.000 \times g$  für 10 Minuten abzentrifugiert und in frischem Puffer resuspendiert. Das Waschen der Zellen wurde mindestens zweimal wiederholt, um Bestandteile des TB-Vollmediums zu entfernen. Auf SH Minimalmedium mit 1,0 % (v/v) Tributyrin wurden dann 5 bis 20 µL des Transformationsansatzes punktförmig aufpipettiert und unter der Sterilbank getrocknet. Die Inkubation der Platten fand bei 60 °C für mindestens 2 bis 3 Tagen statt, bis Wachstum auf den Medien sichtbar war. Gewachsene Kolonien wurden auf frischen Minimalmedien ausgestrichen und erneut auf Wachstum überprüft. Diese Passagierungen wurden insgesamt mindestens dreimal wiederholt, um die Stabilität des Phänotyps durch das Fosmid zu überprüfen. Um den exakten Fosmidklon aus dem Gemisch der ursprünglich 48 Fosmide zu identifizieren, wurde schließlich eine Transformation im Block von *E. coli* in *Thermus* durchgeführt. Die Durchmusterung der wachstumsfähigen Transformanten war analog zur oben beschriebenen Methodik durchgeführt worden.

## 4. Sequenzierungen und bioinformatische Analyse

### 4.1. Sanger-Sequenzierung und Next-Generation Pyrosequencing (454-NGS)

Sequenzierungen wurden mittels Sanger-Kapillarsequenzierungstechnologie (Sanger *et al.*, 1977) bei GATC Biotech AG (Konstanz, Deutschland) durchgeführt.

Fosmide aktiver Klone (*ANG48*) aus *E. coli* wurden am Göttingen Genomics Laboratory (G<sub>2</sub>L) des Instituts für Mikrobiologie und Genetik der Georg-August-Universität Göttingen sequenziert. Ein Pool aus 48 Einzelfosmiden wurde mit einem 454 GS-FLX System (Roche 454 Life Science, Mannheim) sequenziert und ausgewertet (A. Thürmer, G<sub>2</sub>L Göttingen). Endsequenzierungen wurden für die Zuordnung der im Pool NGS-sequenzierten Fosmide durchgeführt (Sanger-Kapillarsequencer). Verwendete Sequenzierprimer waren die T7for, -rev und KOF (Angelov *et al.*, 2009).

Im Falle der positiven *Thermus*-Fosmidklone (*EST6*) wurden die entsprechenden *E. coli*-Klone bei LGC-Genomics (Berlin) sequenziert (mittels 454 GS-FLX Titanium) und ausgewertet (inklusive Genbank-Präparation und *Indexing* auf 1/16 Pikotiterplatte, PTP).

## 4.2. Prozessierung und Analyse der Sequenzen

Sanger-Sequenzdaten wurden mit dem STADEN Software-Paket assembliert (Staden *et al.*, 2000). NGS-Daten der 454-Pyrosequencer wurden mit dem *de novo* Assemblierungsprogramm Newbler bearbeitet (A. Thürmer, G<sub>2</sub>L Göttingen; LGC Genomics, Berlin). Die bioinformatische Auswertung der Sequenzdaten wurde mit einer Reihe an Programmen durchgeführt (Tabelle 14).

**Tabelle 14: Verwendete Programme zur bioinformatischen Auswertung der Sequenzdaten.**

Prozess	Programm/Server	Referenz
ORF-Vorhersage und Annotation <sup>1</sup>	SEED	Overbeek <i>et al.</i> (2005)
Phylogenetische Analyse <sup>1</sup>	MEGAN	Huson und Mitra (2012)
Manuelle Annotation	ARTEMIS	Rutherford <i>et al.</i> (2000)
Phylogenetische 16S rRNA-Klassifikation	Ribosomal Database Project (RDP-II)	Cole <i>et al.</i> (2003)
Proteindomänen	PFAM	Punta <i>et al.</i> (2012)
Ähnlichkeitssuche	<i>blastn</i> und <i>blastp</i>	Altschul <i>et al.</i> (1990)
Identifizierung von Glykosidhydrolasen <sup>1</sup>	HMMER3	Eddy (2011); modifiziert nach A. Angelov
Auffindung klassifizierter Kohlenhydrat-aktiver Enzyme	Carbohydrat-Active enZyme database (CAZy)	Lombard <i>et al.</i> (2013)
Signalpeptid-Vorhersage	SignalP 4.1	Petersen <i>et al.</i> (2011)
Vorhersage von Transmembrandomänen	TMHMM	Krogh <i>et al.</i> (2001)
Metabolische Stoffwechselwege und vergleichende Genomanalysen	MetaCyc/BioCyc	Caspi <i>et al.</i> (2012)
Rekonstruktion metabolischer Stoffwechselwege	KEGG DB	Kanehisa <i>et al.</i> (2014)
Identifizierung konservierter Sequenzmotive	PROSITE	Sigrist <i>et al.</i> (2013)
Fosmid-Visualisierung	DNAPlotter	Carver <i>et al.</i> (2009)
Multiple Genomanalyse	MAUVE	Darling <i>et al.</i> (2010)
Multiplere Sequenzalignement (MSA)	Clustal Omega	Sievers <i>et al.</i> (2011)
Phylogenetische Baumberechnung und -Visualisierung	MEGA 5.2	Tamura <i>et al.</i> , (2011)
Sequenzmanagement	Clone Manager Professional Version 9	Sci-Ed Software, Cary, USA

<sup>1</sup>: bioinformatische Analysen wurden von A. Angelov (Lehrstuhl für Mikrobiologie, TUM) durchgeführt.

## **5. Proteinbiochemische Analytik**

### **5.1. Allgemeine Methoden**

#### **5.1.1. Quantifizierung der Proteinkonzentration**

Proteinkonzentrationen wurden für Zellextrakte, aufkonzentrierte Zellüberstände sowie bei allen Schritten im Laufe einer Proteinaufreinigung bestimmt. Die nach Bradford (1976) benannte Methode wurde mit Bradford Reagent (Fermentas) nach Herstellerangaben ausgeführt. Die Messung wurde in Einweg-Plastikküvetten (Brand GmbH & Co. KG, Wertheim) mit einem Ultrospec 2100pro (Amersham Bioscience) Spektrophotometer bei 595 nm Wellenlänge durchgeführt. Die Proteinkonzentrationen wurden aus den gemessenen OD<sub>595nm</sub>-Absorptionswerten auf Basis einer Kalibriergerade mit BSA-Standards zwischen 0 und 25 µg/µL berechnet. Die Absorptionen von Kontrollproben ohne Protein wurden jeweils abgezogen.

#### **5.1.2. Aufkonzentrieren und Dialyse von Protein**

Niedrige Proteinkonzentrationen (z.B. aus extrazellulären Fraktionen wie Medienüberständen) konnten mit Hilfe von Ultrafiltrationssäulen (Vivaspin 15, molecular cutoff 5-10 kDa, Sartorius Stedim Biotech GmbH, Göttingen) auf kleinere Volumina eingengt werden. Zum Umpuffern und Entsalzen von Proteinlösungen wurden VISKING Dialyseschläuche (Serva, Heidelberg) verwendet, welche vor Verwendung 15 Minuten lang in Wasser gekocht wurden. Die Proteinprobe wurde im Schlauch durch Klammern versiegelt und über Nacht gegen 3-5 Liter Dialysepuffer unter leichtem Rühren bei 4 °C inkubiert.

#### **5.1.3. Formulierung der Proteinlösungen**

Proteine aus Zellextrakten, Zellüberständen und zwischen den einzelnen Aufreinigungsschritten wurden nur wenige Tage bei 8 °C gelagert. Aufgereinigte Proteine waren durch Zugabe von 20 % Glycerol für mehrere Wochen lagerbar, für längere Lagerung wurden sie bei -20 °C weggefroren. Instabile Proteine wurden zwischen allen Aufreinigungsschritten mit Protease Inhibitor Cocktail Tabletten (Roche) versetzt.

## 5.2. Polyacrylamid-Gelelektrophorese

Die Zusammensetzung von Proteingemischen aus Rohextrakten sowie die Analyse aufgereinigter Proteine wurde mit der SDS-Polyacrylamid-Gelelektrophorese (SDS-PAGE) bestimmt. Je nach Molekülmasse der Zielproteine fand die Auftrennung bei verschiedenen Acrylamid/Bisacrylamid-Polymerisierungsgraden der Trenngele statt. Die genaue Zusammensetzung der Gele, des verwendeten Lade- und Laufpuffers ist in Tabelle 15 ersichtlich. Die Vorrichtungen für das Gießen der Gele, Glasplatten, Kämmen, die Gelelektrophoresekammer und Zubehör (Mini-PROTEAN® 3 Cell Elektrophorese-System, BioRad) wurden nach Herstellerangaben verwendet. Nach Zugabe von Ammoniumperoxodisulfat (APS) und Tetramethylethyldiamin (TEMED) wurde das Trenngel sofort in die Glasplatten gegossen und mit etwa 2 cm Isopropanol überschichtet. Nach Aushärtung wurde die Flüssigkeit abgegossen, das Gel mit Sammelgel überschichtet und der Kamm ohne eingeschlossene Luftblasen platziert. Die Proteinproben wurden mit 4x Ladepuffer versetzt und für 5 Minuten bei 95 °C aufgekocht, kurz abzentrifugiert und in die vorgesehenen Taschen pipettiert. Bis zum Eindringen der Lauffront in das Trenngel wurde eine Stromstärke von 15 mA pro Gel angelegt und anschließend auf 30 mA pro Gel erhöht, bis das untere Ende des Gels erreicht wurde. Die Coomassie Färbung und Entfärbung der Gele (siehe Puffer in Tabelle 15) fand auf einem Wippschüttler statt.

## 5.3. Membrantransfer und Immunodetektion von Protein (Western Blot)

Elektrophoretisch-aufgetrennte Proteine aus SDS-Gelen wurden mittels einer *Semi-dry* Blotting Apparatur (2117 Multiphor II Elektrophoresis Unit, LKB, Bromma, Schweden) auf eine Polyvinyliden (PVDF)-Western Blot-Membran (Roche) transferiert. Dazu wurden je 5 in Transferpuffer (5,8 g/L Tris, 2,9 g/L Glycin, 0,37 g/L SDS und 20 % v/v Methanol) getränkte Whatman Filterpapiere (Typ 1) auf die Graphit-Elektroden der Apparatur gelegt, sodass die PVDF-Membran (in Methanol angefeuchtet) auf der Seite der Anode und das SDS-Gel auf der Seite der Kathode luftblasenfrei dazwischen aufliegen können. Mit etwa 0,8 mA pro cm<sup>2</sup> Kontaktfläche wurden die Proteine sowie der molekulare Marker in 1 bis 1,5 Stunden auf die Membran transferiert. Nach dem Blotting-Vorgang wurde die Membran mit PBS-T Puffer (*phosphate buffered saline*: 100 mM Phosphat-Puffer pH 7,4 mit 0,9 % (w/v) NaCl, 0,05 % (v/v) Tween 20) gewaschen. Unspezifische Bindestellen wurden mit 5 % (w/v) Milchpulver in PBS-T für 30 Minuten abgeblockt. Die Zielproteine mit poly-Histidin-*tag* (6xHis) wurden mittels 1:1000 verdünnten primärem Histidin-spezifischen Antikörper (pAb Rabbit, GenScript, Piscataway, USA) für 1 Stunde bei 37 °C oder bei 4 °C ÜN auf einem Wippschüttler inkubiert. Nach weiterem dreimaligen Waschen der Membran für je 15 Minuten in PBS-T wurde der

sekundäre Antikörper (Anti-Rabbit Immunglobulin IgG-Alkaline Phosphatase Konjugat, Boehringer Mannheim GmbH, Deutschland) in einer 1:5000 Verdünnung mit PBS-T und 5 % (w/v) Milchpulver zugegeben und 30 Minuten bei Raumtemperatur inkubiert, ehe der Blot erneut dreimalig mit Wasser oder PBS gewaschen wurde. Die Visualisierung der hybridisierten Antikörper geschah mit Zugabe von BCIP/NBT Färbelösung für wenige Minuten bei Raumtemperatur (ProtoBlot II AP System, Promega).

**Tabelle 15: Zusammensetzung der Lösungen für SDS-PAGE.**

	Reagens	eingesetzte Menge
Trenngelpuffer (pH 8,8):	Tris	182 g/L
	SDS	4,0 g/L
Sammelgelpuffer (pH 6,8):	Tris	61 g/L
	SDS	4,0 g/L
Trenngel (z.B. 10 %):	Trenngelpuffer	1,0 mL
	Acrylamid/Bisacrylamid	1,0 mL
	ddH <sub>2</sub> O	2,0 mL
	Ammoniumperoxodisulfat (10%)	30 µL
	Tetramethylethyldiamin	3,0 µL
Sammelgel (4 %):	Sammelgelpuffer	0,48 mL
	Acrylamid/Bisacrylamid	0,2 mL
	ddH <sub>2</sub> O	1,32 mL
	Ammoniumperoxodisulfat (10%)	15 µL
	Tetramethylethyldiamin	2,0 µL
4x Proteinprobenpuffer:	Glycerin	15 % (v/v)
	β-Mercaptoethanol	5,0 % (v/v)
	SDS	2,4 % (v/v)
	Bromphenolblau (2 %)	1,0 % (v/v)
	Tris-HCl	0,8 % (w/v)
10x Elektrophoresepuffer:	Tris	3,03 % (w/v)
	Glycin	14,4 % (v/v)
	SDS	1,0 % (v/v)
Coomassie Färbepuffer:	Coomassie-Brilliant Blue R 250	0,15 % (w/v)
	Methanol	45,5 % (v/v)
	Essigsäure	8,0 % (v/v)
Entfärbelösung:	Methanol	45 % (v/v)
	Essigsäure	10 % (v/v)

## 5.4. Proteinaufreinigung

### 5.4.1. Hitzefällung von Zellextrakten

Im Zuge der Expression von mutmaßlich thermostabilen Zielproteinen in *E. coli* konnten wirtseigene Proteine aus Zellextrakten durch Hitzeeinwirkung größtenteils denaturiert werden. Nach Inkubation der Rohextrakte bei 60 bis 80 °C über 15 Minuten wurde gefälltes Protein durch Zentrifugation (4 °C, 15 Minuten bei 21.000 × g) pelletiert und der Überstand für nachfolgende Aufreinigungszwecke weiterverwendet.

### 5.4.2. Aufreinigung mittels Fast Protein Liquid Chromatography (FPLC)

Für die Aufreinigung von Zielproteinen wurde das ÄKTA™explorer System (Amersham Pharmacia Biotech, Uppsala, Sweden) mit P-920 Pumpe, UPC-900 Monitor, INV-907 Verteiler und den Fraktionskollektor Frac-900 verwendet. Die Steuerung wurde durch die UNICORN Software bereitgestellt. Alle verwendeten Proben, Lösungen und Puffer wurden mit einem Sterilfilter (Minisart 0,2 µm Porengröße, Sartorius-Stedim, Göttingen) von störenden Partikeln und Schwebstoffen befreit. Um ein Eindringen von Luftblasen in das Säulenmaterial zu verhindern, wurden alle Anschlüsse durch ständigen Flüssigkeitskontakt („drop-to-drop“) verbunden. Die Säulen wurden in 20 % (v/v) Ethanol gelagert. Während des Umpufferns der Pumpen, Schläuche und Säulen zwischen salzhaltigen Lösungen und Ethanol wurde zuvor stets mit ddH<sub>2</sub>O gespült. Die verwendeten Säulen sowie gepacktes Material (GE Healthcare Life Sciences, Uppsala, Schweden) sowie die Aufreinigungs-Bedingungen sind in Tabelle 16 aufgelistet.

**Tabelle 16: Verwendete Säulen zur Aufreinigung von Protein Cel12E mittels FPLC (ÄKTA).**

	Anionenaustauscher	Gelfiltration
Säulentyp/gepacktes Material	HR 10/10 Source30Q	HiLoad XK 16/60 Superdex200
Puffer A	50 mM Tris-HCl pH 8,0	50 mM Tris-HCl pH 8,0, 150 mM NaCl
Puffer B	Puffer A + 1,0 M NaCl	
Gradient	10 CV	Isokratisch; 1,2 CV Fraktionierung
Flussrate	2,0 mL/min	1,0 mL/min
Fraktionsvolumen	4,0 mL	4,0 mL

### 5.4.3. Aufreinigung mittels immobilisierter Metallionen-Affinitätschromatographie

Proteine mit *His-tag* wurden mittels Metallionen-Affinitätschromatographie (IMAC) aufgereinigt. Zur Verwendung kamen das Protino Ni-IDA Protein-Aufreinigungskit (packed columns 2000, Macherey-Nagel) nach Herstellerangaben bzw. 1 mL HisTrap FF (GE Healthcare Life Sciences) Säulen mit einem LKB-Pharmacia FPLC System mit P-500 Pumpeneinheit, LCC 501-Plus Controller, V-7 Valve und Uvicord SII. Als Laufpuffer diente 50 mM 3-(N-Morpholino)propansulfonsäure (MOPS) pH 7,3 mit 20 mM Imidazol, 100 mM NaCl und 20 mM CaCl<sub>2</sub>, bei Flussraten von 1,0 mL/min. Der Elutionspuffer entsprach dem Laufpuffer mit 1,0 M Imidazol.

## 5.5. Biochemische Charakterisierungen

### 5.5.1. $\beta$ -Glucosidase Test

Zur Bestimmung der Aktivität der *Thermus*  $\beta$ -Glucosidase (Bgl, kodiert durch ORF TT\_P0042) wurde ein modifizierter Reporterassay nach Ohta *et al.* (2006) verwendet. Ein standardisierter Test umfasste das Ernten von 1,2 mL Zellsuspension und Resuspension in 1,2 mL 0,1 M Phosphatpuffer (60,1 mM Na<sub>2</sub>HPO<sub>4</sub>, 45,8 mM NaH<sub>2</sub>PO<sub>4</sub>). Davon wurden 0,2 mL für die Bestimmung der optischen Dichte der Kultur verwendet (OD<sub>600nm</sub>), die restliche Suspension wurde mit 1,0 % (v/v) Toluol versetzt und kräftig durchmischt. 0,5 mL der permeabilisierten Zellen wurden zügig mit weiteren 0,5 mL Phosphatpuffer gemischt und mit 8,3  $\mu$ M *para*-Nitrophenyl- $\beta$ -D-Glucopyranosid (*p*NP-Glu) Endkonzentration versetzt (50  $\mu$ L einer 50 mg/mL Substratlösung pro Reaktion). Die Inkubationsdauer bei 80 °C wurde bis zur sichtbaren Gelbfärbung durch freigesetztes *p*-Nitrophenol durchgeführt, ehe der Ansatz mit 0,2 mL Na<sub>2</sub>CO<sub>3</sub> abgestoppt wurde. Nach Abzentrifugieren bei 21.000  $\times$  g für 5 Minuten wurde der Überstand im Spektrophotometer bezüglich der Absorption bei 420 nm gemessen. Zur Absorptionskorrektur durch Zellbestandteile wurde die Lichtstreuung bei 550 nm protokolliert. Die Enzymaktivität der  $\beta$ -Glucosidase (Bgl) wird als Miller Units ausgedrückt und wurde mit folgender Formel berechnet (Gleichung 1):

$$\text{Bgl-Aktivität (Miller Units)} = \frac{1000 \times [\text{OD}_{420\text{nm}} - (1,75 \times \text{OD}_{550\text{nm}})]}{(t \times V \times \text{OD}_{600\text{nm}})}$$

Die Inkubationszeit (t) ist in Minuten angegeben, das Reaktionsvolumen (V) entspricht in einem Standardassay immer 1,0 mL.

### 5.5.2. Bestimmung reduzierender Zuckerenden mit DNSA-Test

Der Abbau von  $\beta$ -glycosidisch verknüpften Polysacchariden und Heteropolysacchariden sowie deren modifizierte Derivate (siehe die Zusammenstellung aller getesteten Substrate für Abbau und Bindestudien in Tabelle 17) durch enzymatische Hydrolyse wurde mittels DNSA-Test kolorimetrisch bestimmt (modifiziert nach Miller, 1959). In einem Standardansatz wurden dabei 250  $\mu$ L Substratlösung in 50 mM Reaktionspuffer mit ddH<sub>2</sub>O auf das Endvolumen aufgefüllt. Nach einer 5-minütigen Vorinkubation der Proben bei der gewünschten Temperatur wurden aufgereinigten Enzympräparationen dazugegeben (auf insgesamt 500  $\mu$ L Gesamtvolumen des Ansatzes). Je nach eingesetzter Enzymmenge (Rohextraktmenge) und verwendetem Substrat wurden die Proben unterschiedlich lange in einem Ölbad-Schüttler (Infors HT Aquatron, Bottmingen, Schweiz) inkubiert.

**Tabelle 17: Polysaccharid-Substrate für biochemische Charakterisierung von Cell2E.**

Substrat	Hersteller	eingesetzte End-Konzentration % (w/v) <sup>1</sup>
Avicel PH-101 <sup>2,4</sup>	Sigma Aldrich, St. Louis, USA	0,5 % - 2,0 %
$\beta$ -glucan (Haferspelz) <sup>2</sup>	Megazyme, Wicklow, Irland	0,5 % - 2,0 %
Carboxymethyl-cellulose (CMC) (niedrige Viskosität) <sup>3,4</sup>	Sigma Aldrich, St. Louis, USA	0,35 % - 5,6 %
Hydroxyethyl-cellulose (HEC)	Fluka, Buchs, Schweiz	1,0 - 2,0 %
Cellulose (mit Phosphorsäure aufgequollen, PASC) <sup>2,4</sup>	Zverlov V., Lehrstuhl für Mikrobiologie, TUM	0,5 – 2,0 %
Pachyman ( <i>Poria cocos</i> ) <sup>2</sup>	Megazyme, Wicklow, Irland	0,5 % - 1,0 %
Chitin <sup>2</sup>	Sigma Aldrich, St. Louis, USA	0,5 % - 2,0 %
Chitosan <sup>2</sup>	Sigma Aldrich, St. Louis, USA	0,5 % - 1,0 %
Laminarin ( <i>Laminaria digitata</i> ) <sup>2</sup>	Sigma Aldrich, St. Louis, USA	0,5 % - 1,0 %
Natrium-Alginat	Sigma Aldrich, St. Louis, USA	1,0 – 2,0 %
Stärke (Kartoffel)	Sigma Aldrich, St. Louis, USA	1,0 %
Amylopektin (Kartoffel)	Sigma Aldrich, St. Louis, USA	2,0 %
Xylan (Haferspelz, Holz von Lärche, Birke) <sup>2</sup>	Fluka, Buchs, Schweiz	0,5 % - 1,0 %
Cellopentaose <sup>4</sup>	Sigma Aldrich, St. Louis, USA	0,7 %

<sup>1</sup>: Die eingesetzten Mengen variierten je nach Experiment: 0,5 % Endkonzentration für Bindestudien, 1,0 bis 2,0 % für enzymatische Hydrolyse (DNSA), 0,35 % bis 5,6 % für Enzymkinetiken.

<sup>2</sup>: Für Bindestudien verwendete unlösliche Substrate (je 0,5 %).

<sup>3</sup>: Lösliche Carboxymethylcellulose wurde für enzymkinetische Messungen verwendet (0,35 bis 5,6 % CMC).

<sup>4</sup>: Für Dünnschichtchromatographie (TLC) verwendete Substrate (je 0,7 % Endkonzentration).

Die Reaktion wurde durch Zugabe von 750  $\mu\text{L}$  DNSA-Reagenslösung (Zusammensetzung siehe Tabelle 18) abgestoppt. Nach 15-minütigem Aufkochen bei ca. 96 °C wurden die Ansätze auf Eis abgekühlt und restliches Substrat und Präzipitat abzentrifugiert (21.000  $\times$  g, 1 Minute, 4 °C). Durch die Braunfärbung der Ansätze waren die Absorptionen bei 515-575 nm Wellenlänge im Spektrophotometer messbar. Die Quantifizierung wurde anhand der Kalibrierkurven aus Glucose/Maltose-Standards zwischen 0,1 und 1,0  $\mu\text{mol}$  pro DNSA Reaktion durchgeführt. Die Absorptionen von Kontrollreaktionen ohne Substrat und ohne Enzymzugaben wurden jeweils abgezogen. Die enzymatische Aktivität von einem Unit (U) wurde definiert als die Bildung von einem  $\mu\text{mol}$  reduzierender Enden pro Minute Inkubationsdauer.

**Tabelle 18: Zusammensetzung des DNSA-Reagens.**

	<b>Reagens</b>	<b>eingesetzte Menge</b>
DNSA-Reagens	3,5-Dinitrosalicylsäure	1,0 % (w/v)
	Phenol	0,2 % (v/v)
	Na <sub>2</sub> SO <sub>3</sub>	0,05 % (w/v)
	K-Na-Tartrat	20 % (w/v)
	NaOH (Plätzchenform)	1,0 % (w/v)

Die hydrolytischen Aktivitäten wurden über einen breiten pH-Bereich getestet (je 50 mM Endkonzentration der Reaktionspuffer: pH 2-3: Glycin-HCl-Puffer; pH 4-6: Natriumacetat-Puffer; pH 5,0: MES-Puffer; pH 5,5-8: Phosphat-Puffer; pH 7,5-8,8: Tris-HCl-Puffer). Getestete Temperaturspektren erstreckten sich von 37 °C bis zum Siedepunkt des Wassers bei etwa 97,5 °C. Der Einfluss verschiedener Detergenzien (Natriumdodecylsulfat, Dithiothreitol, Triton X-100), Salze (Natrium-, Kalium-, Magnesium-, Calciumchlorid) und Metallionen (Kupfer, Kobalt, Nickel, Lithium, Mangan, zwei- und dreiwertige Eisenionen) wurde für biochemische Charakterisierungen analysiert.

### 5.5.3. Qualitative Analyse der Hydrolyseprodukte durch Dünnschichtchromatographie

Niedermolekulare Mono- und Oligosaccharide wurden mit Hilfe der Dünnschichtchromatographie-Methode (engl., *thin layer chromatography*, TLC) untersucht. Die Zweiphasen-Verteilungschromatographie (hydrophobe, mobile Phase und hydrophile, stationäre Phase) erlaubt

eine Auftrennung der Moleküle anhand ihrer unterschiedlichen Wanderungseigenschaften zwischen den Phasen. Als stationäre Phase dienten mit Kieselgel beschichtete Aluminiumfolien (Typ Silicagel 60 F<sub>254</sub>, Fa. Merck, Darmstadt) und als mobile Phase diente ein Gemisch aus 60 % (v/v) 1-Propanol, 10 % (v/v) Ethylacetat und 30 % (v/v) ddH<sub>2</sub>O. Je 2 µL der Überstände aus Reaktionsansätzen mit Substrat und Enzym (wie in III.5.5.2 beschrieben) sowie Zuckerstandards (Mischung aus Monosaccharid bis Heptasaccharid, 0,1 % Endkonzentration) wurden auf die Kieselgel-Platte pipettiert (im Falle zu geringer Produktmengen wurden die Volumina mittels Vakuumkonzentrator entsprechend eingengt). Die chromatographische Auftrennung wurde zweimal durchgeführt, zwischen allen Schritten wurde die Platte jeweils getrocknet. Das frisch angesetzte Entwicklungsreagens, bestehend aus 1,0 % (v/v) Anilin und 1,0 % (w/v) Diphenylamin gelöst in Aceton (Lagerung im Dunkeln bei 4 °C) und 1:10 Volumen *ortho*-Phosphorsäure (85 %), wurde zur Färbung gleichmäßig auf die Plattenfläche gesprüht (5,5 mL pro Platte). Die Platte wurde für 10 bis 15 Minuten bei 140 °C entwickelt.

#### **5.5.4. Zymogramm zur Detektion enzymatischen Abbaus von Polysacchariden**

Zur Feststellung der Aktivität von aufgereinigter Glycosidhydrolase (Cel12E) mittels Zymogramm-Nachweis wurde 0,15 % (w/v) lösliche Carboxymethyl-Cellulose als Substrat in einem herkömmlichem SDS-Polyacrylamidgel gleichmäßig eingegossen und die Auftrennung von denaturiertem Cel12E gemäß dem SDS-PAGE-Verfahren in doppelter Ausführung durchgeführt. Nach der elektrophoretischen Auftrennung wurde ein Gelteil in Coomassie gefärbt, der andere Teil für enzymatische Aktivitätsnachweise in McIlvaine-Puffer (0,2 M Na<sub>2</sub>HPO<sub>4</sub>-Lösung, Einstellen von pH 5,5 mit 0,1 M Citratlösung) weiterverwendet. Das SDS-Gel wurde zunächst mit Wasser und schließlich in Puffer mit 25 % (v/v) Isopropanol gewaschen, und eine halbe Stunde in McIlvaine-Puffer renaturiert. Die frische Pufferzugabe und Inkubation über 2 Tage bei 60 °C erlaubte die enzymatische Hydrolyse des eingegossenen Substrats durch Cel12E. Nach Waschen des Gels in Wasser, folgte die Färbung in 0,1 % (w/v) Kongorot-Lösung und anschließend Entfärbung mit 1,0 M NaCl-Lösung, bis hydrolytische Aufklarungszonen sichtbar wurden.

#### **5.5.5. Bindestudien von Proteinen an unlösliche Polysaccharide**

Zum Nachweis der Bindungsfähigkeit von Cel12E wurden Rohextrakte aus 35 mL Expressionskultur gewonnen und ihr Proteingehalt quantifiziert. Je Ansatz wurden 300 µg des Rohextraktes mit 0,5 % (w/v) unlöslichen Polysacchariden (siehe Tabelle 17) in Bindepuffer (50 mM MES-Puffer pH 5,0 sowie 200 mM NaCl) in 200 µL Gesamtvolumen gemischt und ÜN bei 6 °C unter ständigem

Schütteln in einem Thermomixer Comfort (Eppendorf) inkubiert. Durch Abzentrifugieren bei 4 °C und  $21.000 \times g$  für 30 Minuten konnte gebundenes Protein mit den unlöslichen Substraten pelletiert werden. Der Überstand wurde entfernt und das Pellet in ausreichendem Volumen mit Bindepuffer alleine versetzt und gut resuspendiert. Dieser Waschschrift wurde viermal wiederholt, um unspezifisch gebundene Proteine abzuwaschen. Das gewaschene Pellet wurde in 30  $\mu\text{L}$  Wasser resuspendiert, mit 10  $\mu\text{L}$  4x Proteinladepuffer versetzt und zur elektrophoretischen Trennung im SDS-Gel geladen. Als Kontrolle diente der identische Ansatz ohne Inkubation mit Substraten.

### 5.5.6. Bestimmung lipolytischer Aktivitäten

Esterase/Lipase-Aktivitäten von aufgereinigten Enzympräparationen und Rohextrakten wurden mittels *p*-Nitrophenol-Tests bestimmt (modifiziert nach Fuciños *et al.*, 2005a, Leis *et al.*, 2014). Durch hydrolytische Spaltung der Esterbindung in *para*-Nitrophenyl-Fettsäuren entsteht gelbliches *p*-Nitrophenol (*p*NP), welches bei 400 nm Wellenlänge photometrisch quantifiziert werden kann. Die lipolytischen Enzymaktivitäten aus Zellextrakten und aufgereinigten Proteinen wurden hierfür mit einer Reihe an unterschiedlich langen Fettsäuren wie *p*NP-propionat (*p*NP-C<sub>3</sub>), -butyrat (*p*NP-C<sub>4</sub>), -valerat (*p*NP-C<sub>5</sub>), -caproat (*p*NP-C<sub>6</sub>), -caprylat (*p*NP-C<sub>8</sub>), -caprat (*p*NP-C<sub>10</sub>), -laurat (*p*NP-C<sub>12</sub>), -myristat (*p*NP-C<sub>14</sub>) und -palmitat (*p*NP-C<sub>16</sub>) durchgeführt (alle Substrate stammen von Sigma-Aldrich, St. Louis, USA). Stammlösungen mit 125 mM der Substrate waren jeweils in DMSO löslich und über mehrere Monate bei -20 °C stabil. Der Standardtest wurde dabei in 1,0 mL Reaktionen mit 0,25 bis 2,5 mM *para*-Nitrophenyl-Substraten (Endkonzentration) und 50 mM Tris-HCl (pH 7,5) in einem Thermomixer comfort (Eppendorf, Hamburg, Deutschland) durchgeführt. Für die Charakterisierung der Esterasen wurde ein thermostatisierbares Wasserbad (Thermo Haake, Karlsruhe, Deutschland) verwendet. Die Inkubationsdauer richtete sich nach Menge der zugegebenen lipolytischen Enzyme und Substratkonzentrationen. Der Reaktionsansatz konnte durch die Zugabe von 250  $\mu\text{L}$  2,0 M Natriumcarbonat gestoppt werden. Nach 15-minütiger Inkubation auf Eis wurde der Ansatz bei 4 °C und  $21.000 \times g$  für 10 Minuten abzentrifugiert. Unter den genannten Assaybedingungen konnte eine Kalibriergerade zwischen 1,25 bis 125  $\mu\text{M}$  *p*NP erstellt und der resultierende Extinktionskoeffizient ( $\epsilon = 16.700 \text{ M}^{-1} \times \text{cm}^{-1}$ ) bestimmt werden. Die spezifische Aktivität ist definiert als 1,0  $\mu\text{mol}$  *para*-Nitrophenol, das pro Minute und pro mg eingesetztem Protein freigesetzt wurde. Bei der Quantifizierung der Aktivität von *Thermus*-Zellextrakten wurde aufgrund der hohen Hintergrundaktivitäten (durch Autohydrolyse) lediglich 0,125 mM *p*NP-C<sub>4</sub> in der Reaktion verwendet. Für extrazelluläre Aktivitäten wurden zusätzlich zu den *p*NP-Substraten die wesentlich stabileren und sensitiveren Indigo-Farbstoffderivate 5-Brom-4-Chlor-3-Indolyl-butyrate (BCI-C<sub>4</sub>) und -caprylat (BCI-C<sub>8</sub>) mit 1,6  $\mu\text{M}$  Endkonzentration verwendet (Biosynth, Thal, Schweiz). Ihre Absorption war bei 615 nm photometrisch bestimmbar. Die Aktivitätsparameter der Extrakte, extrazellulären Fraktionen und aufgereinigten Enzyme wurden bei Temperaturen zwischen 60 °C und

97 °C sowie bei pH-Werten zwischen 6-7 (Phosphat-Puffer) und pH 7,2-8,8 (Tris-HCl Puffer) bestimmt. Der Einfluss einiger Salze (1 bis 10 mM Zink, Magnesium, Natrium, Kalium) und Inhibitoren (EDTA, Phenylmethylsulfonylfluorid) wurde ebenfalls untersucht.

## 6. Datendokumentation, -verarbeitung und statistische Auswertungen

Zur Analyse der experimentellen Daten kamen verschiedene Programme und Methoden zum Einsatz. Die Datenverarbeitung und Visualisierung fand mit Hilfe von Microsoft Excel 2007 statt. Mittelwerte mit Fehlerbalken wurden als einfache Standardabweichungen ( $1 \times SD$ ) dargestellt. Zur Ermittlung von nicht-linearen Regressionskurven und statistischen Auswertungen wurde SigmaPlot Version 11.0 (Systat Software Inc.) verwendet (Analyse der Daten auf Normalverteilung mittels Shapiro-Wilk-Test, statistischer Hypothesentest normalverteilter Daten mittels Student's *t*-Test). Die Vorhersage von kombinatorischen Wahrscheinlichkeiten mittels Binomialverteilungen wurde mit dem Java-Applet der Texas A&M Universität (Department of Statistics, <http://www.stat.tamu.edu/~west/applets/binomialdemo.html>) durchgeführt. Die Dokumentation von Bild-basierten Koordinaten (z.B. Pixel-basierte Abstände zu Flächenberechnungen) wurden mittels Grafikprogrammen wie Photoshop CS2 (Adobe, San José, USA) ermittelt. Vektor-basierte Grafiken wurden mittels Inkscape ([www.inkscape.org/](http://www.inkscape.org/)) erstellt. *Heatmap*-Repräsentationen wurden mittels OpenOffice Makro erstellt (Quellcode im Anhang).

Die Berechnung der lipolytischen Aktivität auf Substratplatten mit Tributyrin geschah durch die Bestimmung der Kreisringfläche ( $A$ ) der hydrolytischer Höfe (mit  $D$  = Durchmesser des Hofes) um die gewachsenen Kolonien ( $d$  = Durchmesser der Kolonie) (Gleichung 2):

$$A(\text{hydrolytischer Hof}) = \frac{[D(\text{Hof})^2 - d(\text{Kolonie})^2]}{4} \times \pi$$

Die Zahl der notwendigen Klone zur repräsentativen Abdeckung eines Genoms/Fosmids durch eine Genbank wurde mit folgender Formel bestimmt (Gleichung 3) (Sambrook *et al.*, 1989):

$$N = \frac{\ln(1 - P)}{\ln(1 - f)}$$

Dabei steht  $N$  für die Anzahl an notwendigen Einzelklonen, die für eine erwünschte Wahrscheinlichkeit  $P$  notwendig sind.  $f$  gibt den Anteil genomischer oder Fosmid-Sequenzen wieder, der durch einen Klon repräsentiert wird.

## IV. Ergebnisse

### 1. Verbesserung des Wirt-Vektor-Systems in *Thermus*

Die Nutzung des neuen extrem-thermophilen Expressionswirts *T. thermophilus* HB27 umfasste neben dem Transfer von Genbanken aus *E. coli* und die vergleichende funktionelle Analyse weitere Verbesserungen des Wirt-Vektor-Systems. Die Kenntnis darüber, welche Hintergrundaktivitäten auf den getesteten Substraten zu erwarten sind und wie sich Genbanken effizient im Hochdurchsatzformat übertragen lassen waren dabei ebenso essentiell wie die Etablierung eines neuen Systems zur Expression heterologer Proteine, sowie die Einführung eines gezielten oder ungezielten Systems für genetische Manipulationen in dem thermophilen Wirt.

#### 1.1. Hintergrundaktivitäten in *T. thermophilus* HB27

Die Hintergrundaktivitäten in *T. thermophilus* HB27 spielen für die Nutzung des thermophilen Bakteriums als Expressionswirt für funktionelle Metagenomanalysen eine wesentliche Rolle. Der Erfolg oder Misserfolg der Durchmusterungen hängt im Wesentlichen von der Frage ab, ob die Hintergrundaktivität schwach genug ist, um aktive Klone von nicht aktiven auf den verwendeten Indikatormedien und chromogenen Substraten unterscheiden zu können. Da *T. thermophilus* dafür bekannt ist, lipolytische und proteolytische Aktivitäten auszubilden (Henne *et al.*, 2004; Fuciños *et al.*, 2005a,b), und anderen *Thermus*-Stämmen wie *Thermus caldophilus* GK24 (Kim *et al.*, 2006) und *T. Brockianus* (S. Blank, TU Hamburg-Harburg, persönliche Mitteilung) ebenso (hemi-)cellulolytische Aktivitäten zugeschrieben werden konnten, wurden sämtliche getesteten Substrate (Tabelle 13) auf ihren Abbau hin untersucht. Tabelle 19 zeigt ausgewählte Substrate, die vergleichend in *T. thermophilus* und *E. coli* auf Hintergrundaktivitäten getestet wurden.

Zur Untersuchung der lipolytischen Aktivitäten kam *Staphylococcus carnosus* TM300 mit dem Vektor pLipPS (Liebl und Götz, 1986) zum Einsatz, der eine Lipase sekretiert und als Positivkontrolle auf allen getesteten Substraten (Tributyrin, Triolein und Olivenöl) diene. Dazu wurde der Stamm in LB-Medium bei 37 °C mit 10 µg/mL Chloramphenicol kultiviert und wenige µL der gewachsenen Suspensionen für 3 Tage auf den Substratplatten bei 37 °C inkubiert. *E. coli* zeigte auf allen getesteten Substraten bei 37 °C selbst nach mehrtägiger Inkubation keine Aktivität. In *T. thermophilus* HB27 konnten hydrolytische Höfe auf Tributyrin bei 60 und 70 °C festgestellt werden, die Aktivität auf Olivenöl beschränkt sich auf den Abbau von Fettkügelchen um die Kolonien auf dem Festmedium, nicht auf die Bildung eines klaren Hofes.

**Tabelle 19: Liste der beobachteten Hintergrundaktivitäten in *E. coli* und *T. thermophilus* HB27.**

Hintergrund-Aktivität	Substrat	Verwendeter Stamm	Abbau <sup>1</sup>
Lipasen/Esterasen	Tributyryn 1,0 % (w/v)	<i>E. coli</i> EPI300	-
		<i>T. thermophilus</i> HB27	+
		<i>Staphylococcus carnosus</i> TM300 + pLipPS	++
	Triolein 1,0 % (w/v) (± Rhodamin B)	<i>E. coli</i> EPI300	-
		<i>T. thermophilus</i> HB27	-
		<i>Staphylococcus carnosus</i> TM300 + pLipPS	++
	Olivenöl 1,0 % (w/v) (± Rhodamin B)	<i>E. coli</i> EPI300	-
		<i>T. thermophilus</i> HB27	+/-
		<i>Staphylococcus carnosus</i> TM300 + pLipPS	++
Proteasen	Magermilch 2,0 % (w/v)	<i>E. coli</i> EPI300	-
		<i>T. thermophilus</i> HB27	+/-

<sup>1</sup>: Erklärung der Abkürzungen: nicht aktiv (-), kaum feststellbarer Abbau (+/-), konsistente (+) und hohe (++) Aktivität.

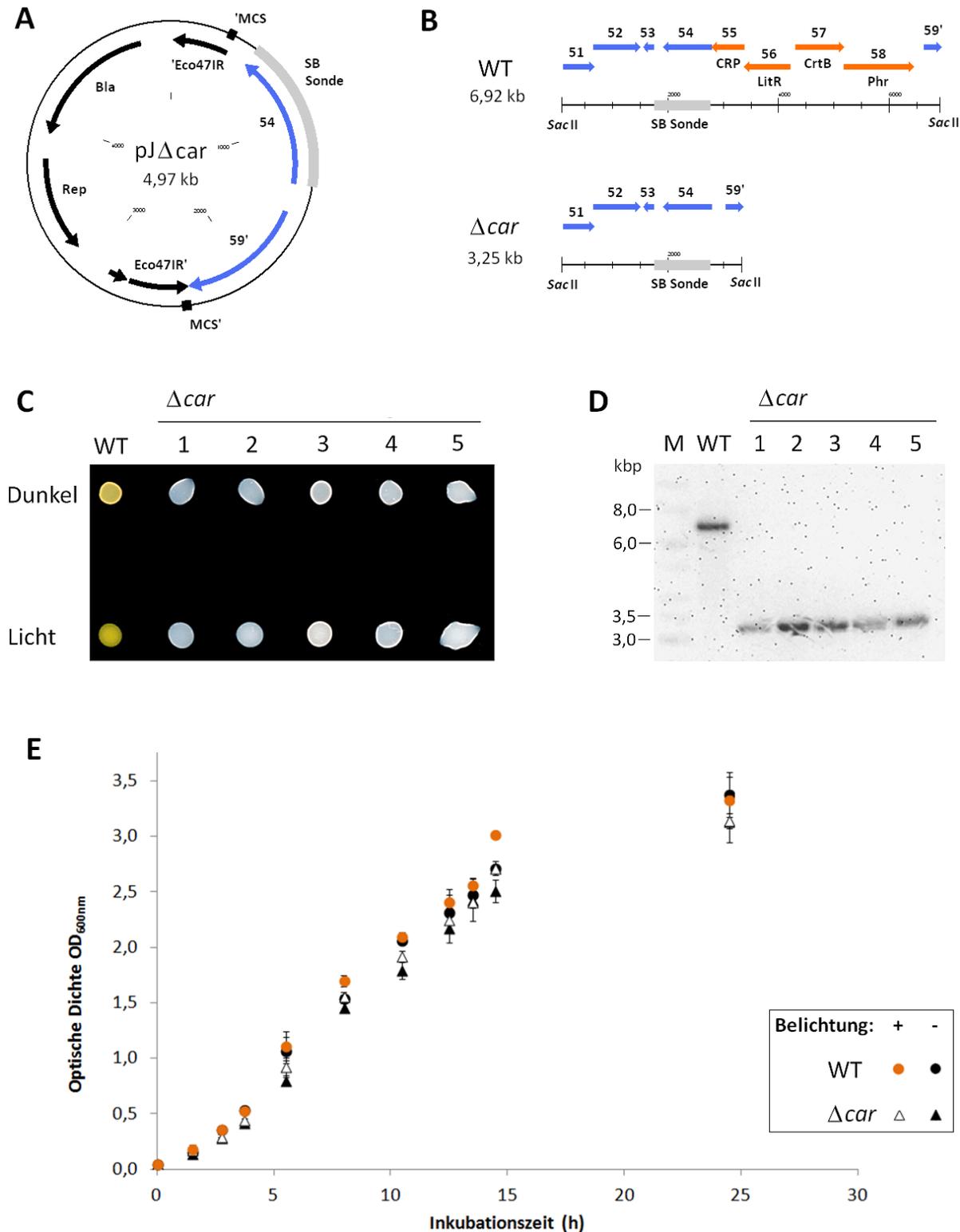
Auf Magermilchplatten konnte in *T. thermophilus* ebenfalls eine geringe Aktivität nachgewiesen werden, die sich auf eine Aufklärung des Mediums unter den gewachsenen Kolonien beschränkt und damit ebenfalls keine Hofbildung um die Zellen nachzuweisen war. Der Einsatz von Farbstoffmarkierten Substraten wie den verwendeten AZCL-Substraten (Megazyme) war in *E. coli* ohne Probleme möglich. In dem thermophilen Bakterium hingegen konnte der Abbau des nicht löslichen AZCL-Substrats für Xylan, Hydroxyethylcellulose und  $\beta$ -Glucan bei 60 °C beobachtet werden. Die Farbstoffbildung im Überstand war dabei willkürlich und korrespondierte mit der Menge an eingesetztem Substrat, nicht aber mit der Menge der Zellen oder Zellextrakte. Der Einsatz dieser Substrate war für funktionelle Durchmusterungen der Metagenombanken in *T. thermophilus* daher nicht möglich.

## 1.2. Entwicklung eines licht-induzierbaren Expressionsystems auf Basis des Car-Clusters

Das Carotinoid-Cluster (Car-Cluster) von *Thermus* und dessen lichtinduzierte Regulation wurde bereits im Detail studiert (Ortiz-Guerrero *et al.*, 2011; Takano *et al.*, 2011). Dies bietet die Möglichkeit, Proteinexpression selektiv über die Belichtung zu induzieren. Aufbauend auf dem Car-Cluster sollte ein Expressionsvektor generiert werden, der auf Basis des pMK18-Vektors die  $\beta$ -Glucosidase (Bgl) exprimieren sollte.

### 1.2.1. Deletion des Car-Clusters ( $\Delta car$ ) in *Thermus*

Das auf dem Megaplasmid pTT27 vorliegende Car-Cluster zur lichtinduzierten Carotinoid-Biosynthese sollte zunächst aus *T. thermophilus* Stamm  $\Delta bgl$  entfernt werden, um das identische Cluster später über einen Expressionsvektor wieder einzuführen (W. Ott, Praktikumsreport, 2012). Zur Erstellung des Deletionsvektors wurde ein 5,67 kb großes Stück des Car-Clusters mit den Primern RT-1 und *car\_rev* amplifiziert und in den CloneJet-Vektor ligiert (resultierender Vektor pJ-car). Eine SLIM Hybridisierungsmethode (Chiu *et al.*, 2004) mit den Primern  $\Delta car_K_{for}$ ,  $\Delta car_K_{rev}$ ,  $\Delta car_L_{for}$  und  $\Delta car_L_{rev}$  ermöglichte die selektive Amplifikation des pJet-Vektors mit den jeweiligen 1,0 kb langen Flanken des Car-Clusters zur Generierung des resultierenden *Knockout*-Vektors pJ $\Delta car$  (Abbildung 4A). Durch die Integration und Rekombinationsvorgänge des Deletionsvektors pJ $\Delta car$  im Carotinoid-Cluster von *T. thermophilus* sollte das Cluster von ORFs TT\_P0055 bis TT\_P0058 vollständig entfernt werden (Abbildung 4B). Etwa 7,5  $\mu$ g des Vektors wurden in *T. thermophilus*  $\Delta bgl$  übertragen und Kolonien waren bereits nach einem Tag bei 70 °C auf TB Festmedien auf der 10<sup>-5</sup>-Verdünnung sichtbar. Um den Kontrast weißer Kolonien (Carotinoid-defiziente Mutanten) unter den sonst gelblich bis orange-farbenen Kolonien besser unterscheiden zu können, wurde die Platte mit etwa 2 bis 3 kLux Belichtungsstärke bei 70 °C für 5 Stunden belichtet. Es konnten zehn weiße Kolonien auf dieser Platte identifiziert werden. Von diesen wurden fünf Kandidatenkolonien auf die Stabilität des Phänotyps hin positiv getestet, indem gewachsene Kolonien in wenigen  $\mu$ L Wasser resuspendiert wurden und in Form kleiner Tröpfchen auf einen Parafilm pipettiert wurden (Abbildung 4C). Die Zellen zeigten sowohl im Dunklen als auch unter Lichtexposition im Inkubator bei 70 °C weißliche Kolonien während Wildtyp-Zellen durch vermehrte Carotinoid-Bildung im Licht eine intensive orange Färbung angenommen haben. Der Genotyp wurde mittels Detektions-PCR und Southern Blot-Analyse bestätigt (Abbildung 4D). Die genomische DNA wurde hierfür mit dem Restriktionsenzym *SacII* verdaut. Die 1,0 kb lange DNA-Sonde (Abbildung 4B, grauer Balken) wurde aus einer Flanke des Car-Clusters mittels *car\_for* und  $\Delta car_K_{rev}$  amplifiziert und hybridisierte im intakten Car-Clusters des Wildtyps mit einem 6,9 kb großem Fragment, die  $\Delta car$ -Mutante hingegen zeigte ein 3,3 kb großes Fragment an. Ausgehend von der  $\Delta car$ -Mutante konnte der Einfluss des fehlenden Car-Clusters auf das Wachstumsverhalten bei 60 °C im Vergleich zu unmodifizierten Zellen in Abhängigkeit der Belichtung untersucht werden (Abbildung 4E). Die Wachstumskurven zeigen dabei einen identischen Verlauf beider Kulturen im unbelichteten und belichteten Zustand. Die Untersuchungen höherer Temperaturen bei 70 °C und 80 °C konnten ebenfalls keinen Unterschied zwischen der Mutante und Wildtyp-Stamm aufzeigen (Daten nicht gezeigt).

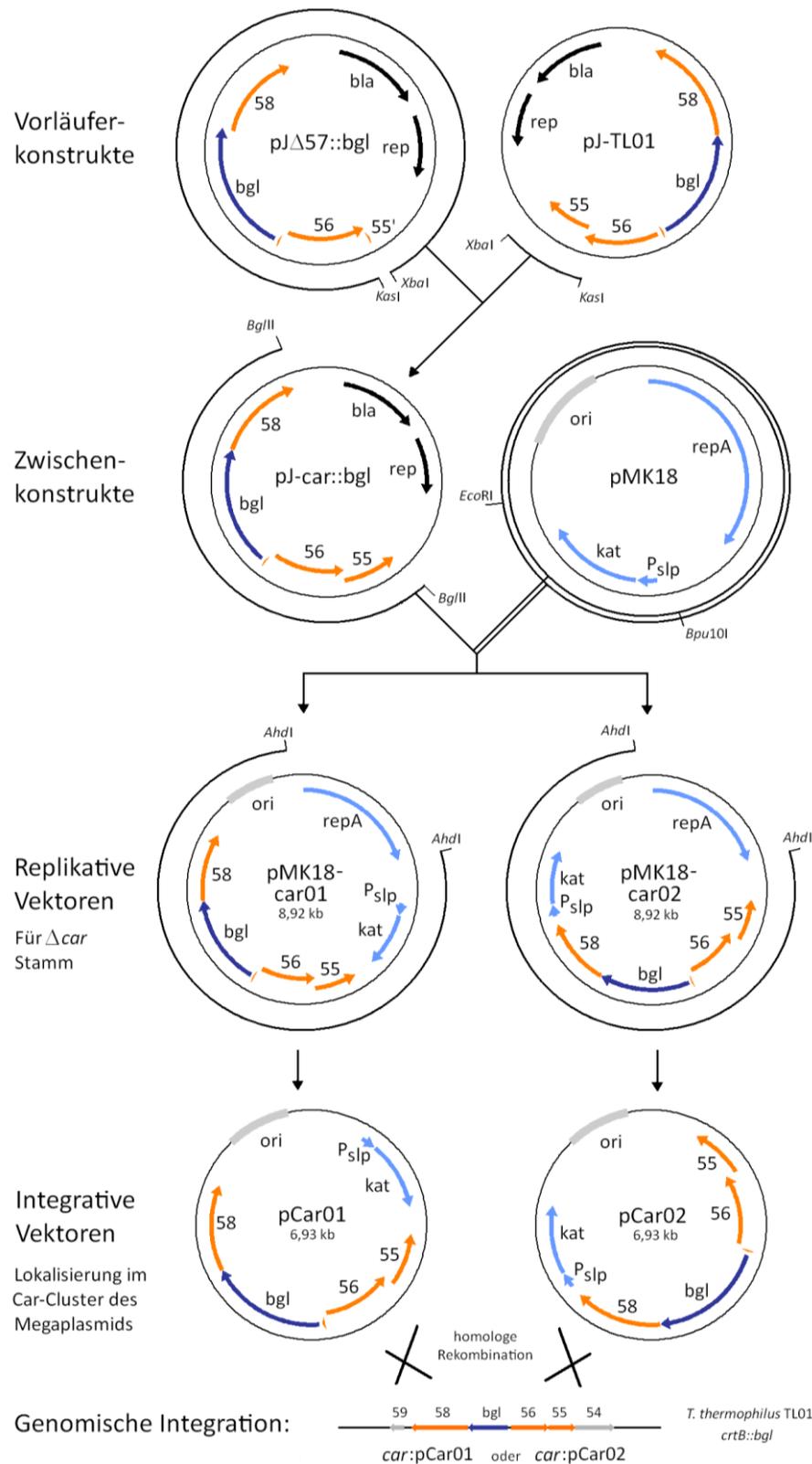


**Abbildung 4: Charakterisierung der markerlosen Deletionsmutante  $\Delta car$  in *T. thermophilus*  $\Delta bgl$ .**  
**A:** Plasmidkarte des konstruierten Deletionsvektors pJ $\Delta car$ , die Gennummern (TT\_P) wurden abgekürzt.  
**B:** Genetische Situation des WT gegenüber der geplanten Deletion des Car-Clusters in *T. thermophilus*.  
**C:** Phänotypische Untersuchung belichteter  $\Delta car$ -Mutanten (1-5) im Vergleich zum Wildtyp (WT).  
**D:** Southern Blot-Analyse der  $\Delta car$ -Mutanten (1-5), Detektion mittels CDP-Star Reagens und Chemolumineszenz-Detektion.  
**E:** Wachstumskurve von *T. thermophilus* HB27 WT (Punkte) und  $\Delta car$ -Mutante (Dreiecke) bei 60 °C ohne Belichtung im Dunkeln (schwarze Punkte und Dreiecke) und unter Lichtexposition (orangefarbene Punkte für WT, weiße Dreiecke für Mutante). Gezeigt sind die jeweiligen Mittelwerte von drei biologischen Replikaten, Fehlerbalken als einfache Standardabweichung ( $\pm 1 \times SD$ ). Nähere Erläuterungen befinden sich im Text. Abb. C-D nach W. Ott, Praktikumsreport (2012).

### 1.2.2. Konstruktion von Expressionsvektoren auf Basis des Carotinoid-Clusters

Für die Nutzung des neuen Expressionssystems zur homologen Expression von  $\beta$ -Glucosidase (Bgl) als Reporter wurden zwei Strategien eingeschlagen. Eine Möglichkeit ist es, das Zielgen in einem eigenständigen Replikon als Vektor zu exprimieren. Dazu sollte das pMK18-*Shuttle*-Plasmid mit den Regulationselementen des Car-Clusters versehen werden. Die Expression kann dann in dem generierten  $\Delta car$ -Stamm stattfinden. Als zweite Möglichkeit sollte der Vektor als *Suicide*-Vektor in ein bestehendes Car-Cluster integrierbar sein. Dabei konnte auf den bestehenden Stamm *T. thermophilus* TL01 (A. Lichev, Bachelorthesis, 2011) zurückgegriffen werden, welcher anstelle des *crtB*-Gens (ORF TT\_P0057) bereits eine  $\beta$ -Glucosidase (Bgl) als Reportergen integriert hat (Genotyp  $\Delta crtB::bgl$ ). Dieser Stamm dient als Reporterstamm, um den Einfluss der Lichtinduktion im Miller-Assay quantifizierbar zu machen (die kontrollierte Belichtung führt zur Aktivierung des  $P_{crtB}$ -Promotors und damit zur homologen Expression der  $\beta$ -Glucosidase). Als integrativer Vektor diente ein modifiziertes pMK18 *Shuttle*-Plasmid, welches keine *repA*-Sequenz zur Replikation in *Thermus* trägt und unter Kanamycin-Selektion im Genom lokalisiert werden kann. Die konstruierten Plasmide sind in Abbildung 5 schematisch wiedergegeben.

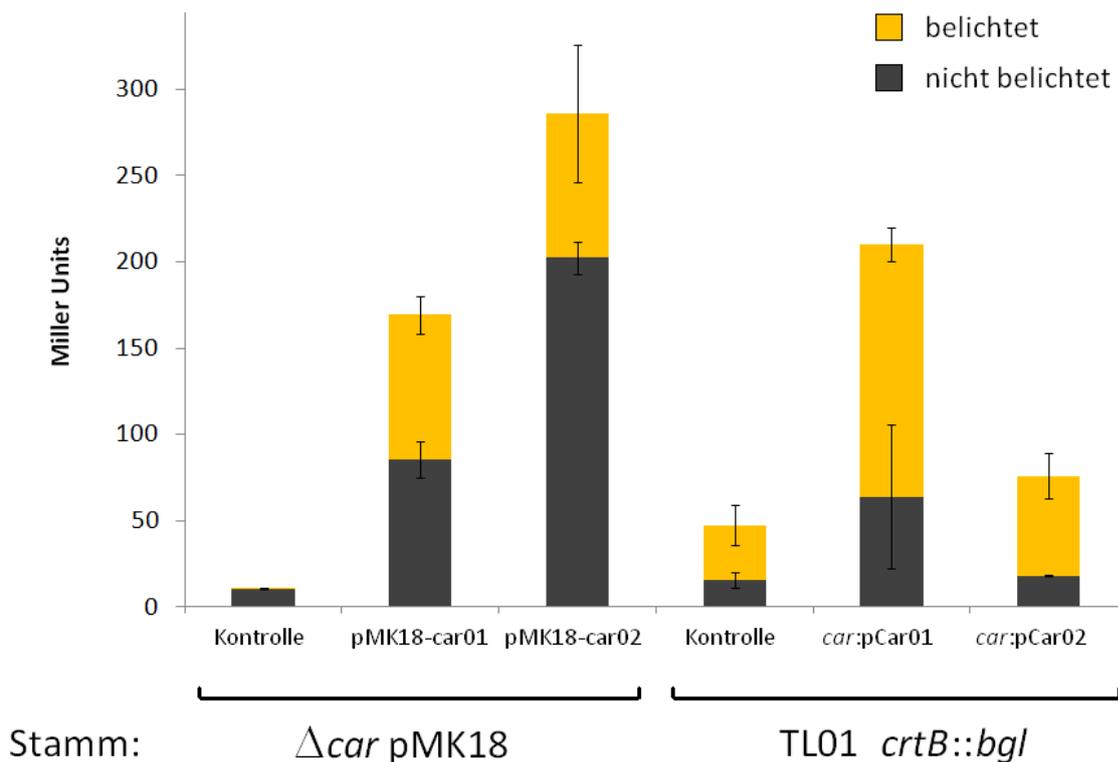
Das Car-Cluster aus dem *T. thermophilus* Reporterstamm TL01 wurde mit dem Primerpaar *car\_BlpI\_for* und *car\_BlpI\_rev* amplifiziert, und das 4,20 kb große PCR-Produkt wurde in einen CloneJet-Vektor (Fermentas) kloniert (resultierender Vektor: pJ-TL01). Da die Amplifikation des vollständigen Car-Clusters nicht ohne nachteilige Mutationen möglich war (Verschiebung des Leserahmens, Änderung der Aminosäuresequenz durch Basenveränderungen) (eigene Beobachtungen, Daten nicht gezeigt), wurde der vollständig sequenzierte Vektor pJ $\Delta$ 57::bgl (A. Lichev, Bachelorthesis, 2011) als Ausgangsbasis für die Klonierungen der Expressionvektoren weiterverwendet. Ein Fragment mit ORF TT\_P0055 (684 bp) wurde aus dem vorher generierten Vektor pJ-TL01 mit *KasI* (*SspDI*) und *XbaI* ausgeschnitten und über die identischen Restriktionsstellen in das 6,12 kb große Rückgrat von pJ $\Delta$ 57::bgl inseriert. Der resultierende Vektor pJ-car::bgl umfasste 6,81 kb Größe und diente als Ausgangsvektor für die Klonierung des modifizierten Car-Clusters in pMK18. Das Cluster wurde mit *BglII* als 3,88 kb großes Fragment ausgeschnitten und die Enden geglättet (*blunting*). Der pMK18-Vektor wurde einmal mit *EcoRI* und einmal mit *Bpu10I* linearisiert, die Enden ebenfalls geglättet. Das Car-Cluster konnte damit in zwei Varianten in den pMK18-Vektor integriert werden, dabei einmal nach der *kat*-Kassette mit *slp*-Promotor in der MCS (resultierender Vektor über *EcoRI* war pMK18-car01) und einmal davor (über *Bpu10I* entstand pMK18-car02). Die resultierenden Vektoren hatten eine Größe von 8,92 kb. Nach Verdau der Vektoren mit *AhdI* konnte die *repA*-Sequenz entfernt werden. Die resultierenden Vektoren wurden pCar01 und pCar02 genannt und hatten eine Größe von 6,93 kb.



**Abbildung 5: Flussdiagramm zur Konstruktion der *Thermus*-Expressionsvektoren basierend auf dem Car-Cluster in *T. thermophilus* HB27. Als Reportergen wurde die  $\beta$ -Glucosidase (Bgl) anstelle der Phytoen-Synthase (*crtB*) verwendet ( $\Delta 57::bgl$  aus Stamm TL01, A. Lichev, Bachelorthesis 2011). Das vollständige Car-Cluster ( $pJ-car::bgl$ ) wurde aus den Vektoren  $pJ-TL01$  und  $pJA57::bgl$  in  $pMK18$  als neuartiger Expressionsvektor kloniert, welcher für den  $\Delta car$ -Stamm verwendet werden konnten ( $pMK18-car01$  und  $pMK18-car02$ ). Integrative Vektorvarianten waren  $pCar01$  und  $pCar02$ , welche im Stamm TL01 über homologe Rekombination in das modifizierte Car-Cluster (*crtB::bgl*) integriert werden können ( $car:pCar01$  oder  $car:pCar02$ ). Die Gennummern (TT\_P) wurden abgekürzt.**

### 1.2.3. Vergleich der Expressionsvektoren im Reporterassay

Da in allen vorhandenen Expressionsvektoren die *Thermus*-eigene  $\beta$ -Glucosidase im Car-Cluster mitkloniert war, konnten Aktivitäten vor und nach Induktion im Miller-Assay quantifiziert werden. ÜN Vorkulturen mit 5 mL TB-Medium wurden mit Aluminiumfolie abgedeckt und bei 60 °C inkubiert. Die 35 mL Hauptkulturen wurden etwa 10 %ig inokuliert und in 100 mL Erlenmeyerkölbchen bei 180 Upm und 60 °C im Dunkeln geschüttelt. Ab einer optischen Dichte ( $OD_{600nm}$ ) von circa 0,5 (nach etwa 5 Stunden Inkubation) wurde die Suspension auf je 5 mL Kulturen aufgeteilt. Eine diente zur Bestimmung der Basalaktivitäten ohne Lichtinduktion und wurde daher jeweils mit Aluminium-Folie abgedeckt. Der andere Teil wurde im Schüttelinkubator mit einer photosynthetischen Belichtungseinheit (Innova 42, New Brunswick) bei etwa 2-3 kLux Beleuchtungsstärke zwischen 5 und 6 Stunden und über Nacht inkubiert. Die Zellsuspensionen wurden geerntet und im Ultraschall aufgeschlossen. Zur Bestimmung der  $\beta$ -Glucosidase-Aktivität wurde ein Teil der Zellen für den Miller-Assay mit *para*-Nitrophenyl- $\beta$ -D-glucopyranosid (*pNP*-Glu) bei 80 °C gemessen (Abbildung 6).



**Abbildung 6:** Aktivitätsdaten der Expressionsplasmide in *T. thermophilus*. Im  $\Delta car$ -Stamm kamen replikative Vektoren zum Einsatz (pMK18-basiert), im Stamm TL01 integrative Plasmide (pCar). Aus 5 mL Kulturen wurden die Hintergrundaktivitäten ohne Beleuchtung (schwarze Balken) und Aktivitäten nach Induktion durch Beleuchtung (orangefarbene Balken) bei 60 °C und 180 Upm nach 5 bis 6 Stunden Inkubation bestimmt. Während die uninduzierten Kulturen im Dunkeln belassen wurden (Reaktionsgefäße eingepackt in Aluminiumfolie), waren die induzierten Kulturen mit etwa 2-3 kLux beleuchtet worden. Gezeigt sind die Mittelwerte von je zwei biologischen Replikaten sowie deren Fehlerbalken (als einfache Standardabweichung). Als Kontrollen dienen der pMK18 Leervektor im  $\Delta car$ -Stamm (links) und der unmodifizierte Stamm TL01 (rechts).

Die Induktion durch Beleuchtung zeigte immer höhere  $\beta$ -Glucosidase-Aktivitäten als die uninduzierten Zellen im Dunkeln. Im Carotinoid-defizienten Stamm *T. thermophilus*  $\Delta car$  (Genotyp  $\Delta TT\_P0055$  bis  $\Delta TT\_P0058 \Delta bgl$ ) zeigte die Kontrolle mit pMK18-Leervektor keine Aktivitätszunahme und eine gleichbleibende Basalaktivität ohne Induktion. In den beiden Expressionsvektoren pMK18-car01 und pMK18-car02 waren die Aktivitätszunahmen durch Belichtung in etwa gleich stark, jedoch unterschieden sich die Hintergrundaktivitäten ohne Induktion deutlich. Im TL01-Reporterstamm konnte ebenfalls eine Aktivitätszunahme nach Induktion festgestellt werden. Durch die selektive Integration des *Suicide*-Plasmids pCar01 konnte die höchste gemessene Zunahme der  $\beta$ -Glucosidase-Aktivität durch Licht-Induktion beobachtet werden (146,2 Miller Units zwischen unbelichteten und belichteten Zustand). Der zweite integrative Vektor mit der Lokalisierung des Car-Clusters *upstream* des *slp*-Promotors zeigte keine wesentliche Veränderung gegenüber dem Verhalten des TL01-Reporterstammes. Die Orientierung des Car-Clusters in pCar02 spielte dabei keine Rolle (Daten nicht gezeigt).

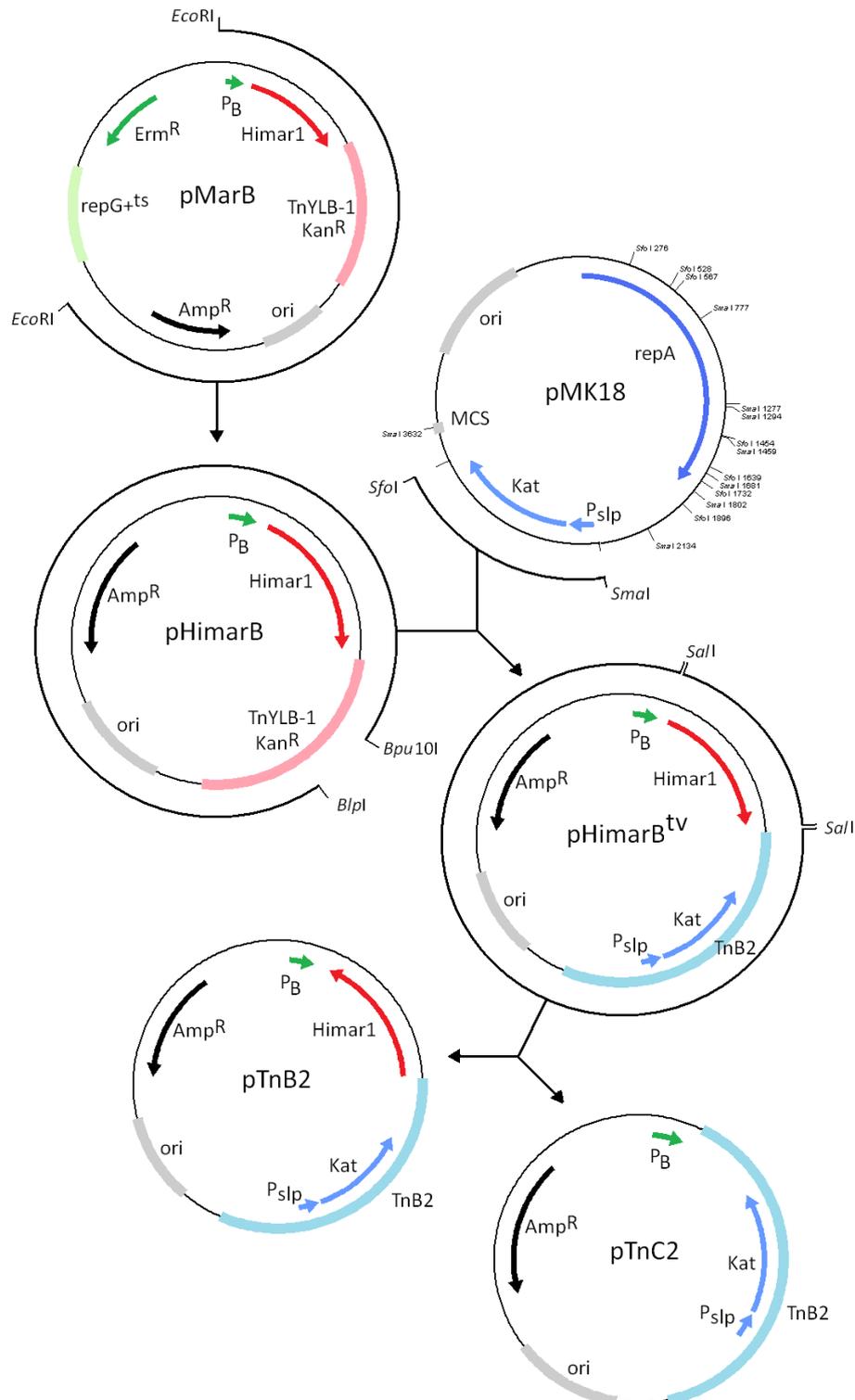
### 1.3. Erstellung einer Transposonmutagenese-Bank in *T. thermophilus* HB27

Die Erstellung von Transposon-Mutagenesebanken eröffnet das intensive Studium eines Organismus und ermöglicht die selektive Identifizierung einzelner Gene und kodierter Proteine, die für bestimmte Eigenschaften des Bakteriums verantwortlich sein können (*loss-of-function* Prinzip). In dieser Arbeit wurden mit *in vitro* und *in vivo*-Mutagenesen zwei verschiedene Strategien gewählt, um in *Thermus* zufällige Transposon-Mutagenesebanken herzustellen.

#### 1.3.1. *In vivo* Transposonmutagenese

Das in *Bacillus* etablierte *in vivo* Transposon-Mutagenesesystem TnYLB-1 in pMarA und pMarB (Le Breton *et al.*, 2006) trägt die Transposase *Himar1* der Hornfliege. Da die Transposase hyperaktiv ist und die Insertionssequenz ,TA<sup>c</sup> bevorzugt, wurde ein Transposon-Mutagenesevektor auf Basis des pMarB Vektors für *Thermus* konstruiert (Abbildung 7). Dazu wurden zunächst die *Bacillus*-Elemente für Selektion (Erythromycin-Resistenz) und Mobilisierung (temperatursensitives *repG*) des pMarB über *EcoRI* entfernt und der Vektor religiert (pHimarB). Um die Kanamycin-Resistenzkassette aus dem Transposon TnYLB-1 zu ersetzen, wurde der Vektor mit *BlpI* und *Bpu10I* geschnitten, und ein 1.018 Bp großes Fragment mit *slp*-Promotor ( $P_{slp}$ ) und thermostabiler Kanamycin-Adenyltransferase (*kat*) aus pMK18 (mit *SmaI* und *SfoI* geschnitten und die Enden mit Klenow-Fragment geglättet) eingefügt. Der entstandene Vektor pHimarB<sup>tv</sup> wurde anschließend mit *SalI* verdaut, um einerseits die Transposase zu invertieren (pTnB2) und andererseits ganz auszuschneiden (pTnC2). Der entstandene

Vektor pTnC2 diente als Negativkontrolle für die Transposonmutagenese, um unspezifische Integration ohne die Präsenz der *Himar1* Transposase im Vergleich zu protokollieren.



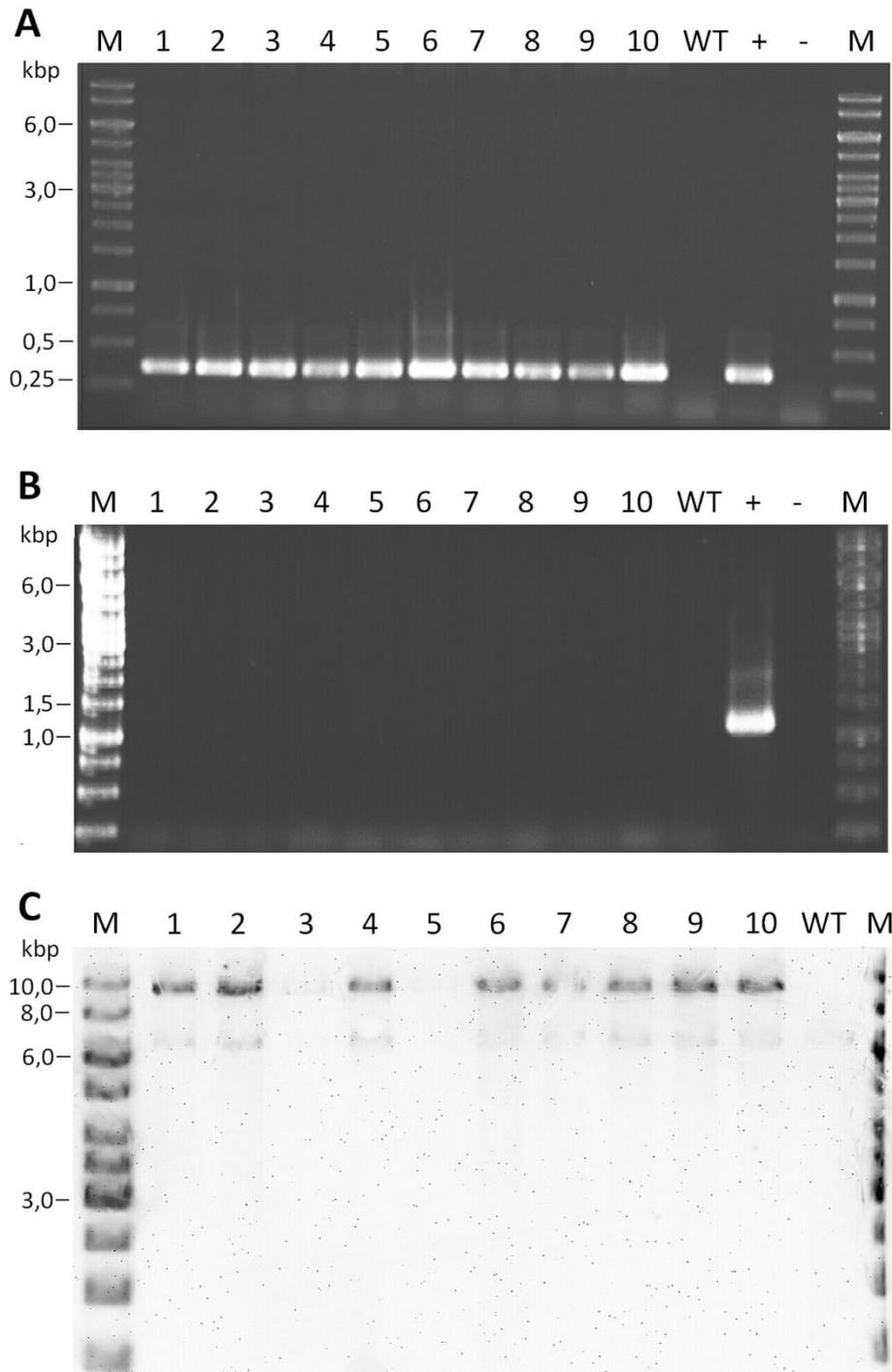
**Abbildung 7:** Schematische Darstellung der Konstruktion von pTnB2 und pTnC2 für *in vivo* Transposon-Mutagenese in *T. thermophilus* HB27. Gene sind als Pfeile dargestellt, die dickeren Markierungen stellen Replikationselemente (*ori*) und Transposonbereiche dar. Die Farbkodierung repräsentiert die genetischen Elemente für den jeweiligen bakteriellen Wirt (grün: *Bacillus*; schwarz/grau: *E. coli*; Blau: *Thermus*). Die *Himar1* Transposase und Transposon TnYLB-1 sind rot gekennzeichnet.

Die Inversion der Transposase über *SalI* ermöglichte die Ausrichtung in gleicher Transkriptionsrichtung hinter dem konstitutiven *slp*-Promotor und der *kat*-Kassette im Transposon. Die flankierende Transposase-Erkennungssequenz des Transposons (*inverted terminal repeat* (ITR)-Sequenz 5'-TAACAGGTTGGCTGATAAGTCCCCGGTCT-3') blieb dabei vollständig erhalten. Nach Übertragung von je 1 µg aller generierten Vektoren wurden die Transformationsmische für die Transposase-Aktivität für 15 Minuten bei 37 °C inkubiert. Lediglich für pTnB2 konnten mehrere Dutzend *Thermus*-Klone auf TB Kanamycin (20 µg/mL) nach 2 Tagen bei 60 °C identifiziert werden (Tabelle 20).

**Tabelle 20: Zusammenfassung der *in vivo* Transposonmutagenese in *T. thermophilus*.**

Vektor	Beschreibung	Zahl Transformanten KBE pro µg Vektor
pMarB	Ausgangsvektor mit Tn <i>YLB-1</i> Transposon, <i>Himar1</i> unter <i>Bacillus</i> Promotor P <sub>B</sub>	0
pHimarB	pMarB ohne <i>Bacillus</i> -Elemente	0
pHimarB <sup>tv</sup>	pHimarB mit thermostabiler Kanamycin-Resistenz (Kat) in Tn <i>YLB-1</i> (TnB2)	0
pTnC2	pHimarB <sup>tv</sup> ohne <i>Himar1</i> Transposase	0
pTnB2	<i>Himar1</i> und Kat unter dem Einfluss des P <sub><i>slp</i></sub> Promotors (TnB2)	240

Zur Überprüfung des Insertionsorts des Transposons durch *Himar1* wurden zehn Kolonien zufällig ausgewählt und genomische DNA isoliert. Mittels PCR-Nachweis konnte das Transposon spezifisch amplifiziert werden, während die Sequenz der Transposase selbst nicht mehr nachweisbar war (Abbildung 8A und B). Zur Überprüfung der erfolgreichen und zufälligen Insertion des Transposons in das Genom von *T. thermophilus* HB27 wurde eine Southern Blot Analyse der zehn Kolonien durchgeführt (Abbildung 8C). Die DNA-Sonde stammte aus dem pTnC2-Vektor, wobei das 1,268 kb große TnB2-Transposon-Fragment mit der Kanamycin-Kassette mittels *BglII* ausgeschnitten und mit Biotin markiert wurde. Die genomischen DNA-Proben wurden mit *BamHI* verdaut. Der Nachweis der Hybridisierung zeigte anstelle verschieden großer Fragmente nur identische Signale bei etwa 9,5 kb Größe, die keinem bestimmten Locus im Genom zugeordnet werden konnten. Die Ursache für diese möglichen Sequenzpräferenzen wurde nicht weiter untersucht, sodass anstelle der *in vivo* Mutagenese eine andere Mutagenese-Strategie gewählt wurde.



**Abbildung 8:** Analyse der genomischen DNA zehner verschiedener Transposonmutanten (1-10) in *T. thermophilus* HB27 aus einer *in vivo* Mutagenese mit pTnB2. Kontrollen waren jeweils unmodifizierte Wildtyp-DNA (WT), pTnB2-Vektor (+) und ein Ansatz ohne Zugabe von Matrizen-DNA (-). **A:** Detektions-PCR der Transposon-Sequenz mittels pMK18\_NdeI\_for und pMK18\_NdeI\_rev Primer, welches bei erfolgter Insertion des Transposons TnB2 ein positives Signal bei 351 Bp erwarten ließ. **B:** Die Ausschluss-PCR für die Präsenz von *Himar1* Transposase erlaubte den Nachweis der Insertion des kompletten pTnB2-Vektors in das Chromosom oder Megaplasmid von *T. thermophilus*. Die Primer *Himar1\_for* und *Himar1\_rev* gaben ein 1.135 Bp großes Produkt bei der Vektorkontrolle, nicht aber aus der genomischen DNA der Klone. **C:** Southern Blot Analyse der 10 Klone im Vergleich zum Wildtyp. Die Sonde wurde als 0,35 kbp großes *BglIII/PdmI*-Verdauprodukt aus Vektor pTnB2 hergestellt. Als Marker wurde der GeneRuler 1 kb DNA Ladder (Fermentas) verwendet.

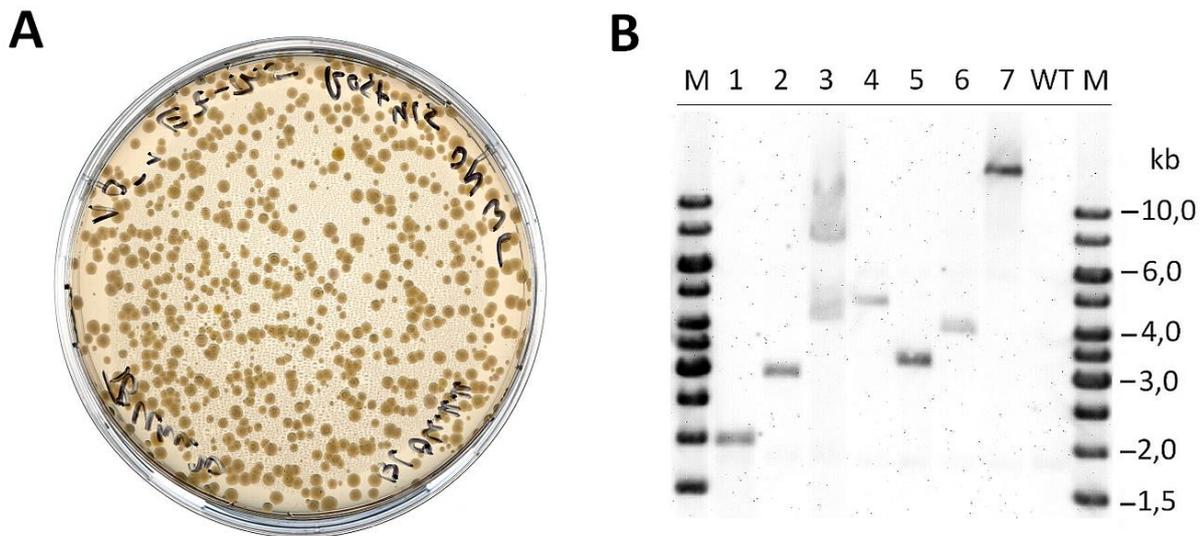
### 1.3.2. *In vitro* Transposonmutagenese

Die Erstellung von Transposon-Mutanten in genomischer DNA wurde mit EZ-Tn5™ Transposase und EZ-Tn5™ pMOD™3<R6K $\gamma$ ori/MCS> Transposon Construction Vector (Epicentre) durchgeführt. Statt pMOD™3 wurde pMOD3-kat (Angelov *et al.*, 2013) zur Gewinnung des Transposons verwendet. Der Verdau des Plasmids mit *BoxI* (*PshAI*) resultierte in einem 1,6 kb langen Transposon-Fragment. Für die *in vitro* Transposon-Reaktion wurden 1  $\mu$ g Transposon und 20  $\mu$ g genomische DNA von *T. thermophilus* HB27 verwendet und nach Herstellerangaben ausgeführt. Die Transposon-mutagenisierte genomische DNA (tgDNA) wurde zusätzlichen endständigen Modifikationen unterzogen (B. Leis, nicht veröffentlichte Daten). Nach erneuter Aufreinigung und Konzentrierung der Probe auf 35  $\mu$ L (Sureclean, Bionline) wurden verschiedene Volumina (1, 5 und 10  $\mu$ L) der modifizierten tgDNA zu einem Standard-Transformationsansatz in *T. thermophilus* HB27 gegeben. Die Transformanten wurden über 2-3 Tage auf TB Kanamycin (20  $\mu$ g/mL Endkonzentration) bei 60 °C selektiert. Die normierten Koloniezahlen pro  $\mu$ L eingesetztem tgDNA-Volumen sind in Tabelle 21 zusammengefasst. Zur Maximierung der Zahl der Transposon-Mutanten wurde die restliche DNA aus der Transposon-Reaktion mit jeweils 1  $\mu$ L tgDNA-Volumen erneut übertragen.

**Tabelle 21: Zusammenfassung der *in vitro* Transposonmutagenese in *T. thermophilus*.**

$\mu$ L der <i>in vitro</i> Transposon-Reaktion bei einer Transformation	Kolonien pro Transformation (KBE)	Maximale erreichbare Klonzahl pro 1 $\mu$ L Transposon-Reaktion
0 $\mu$ L	0	-
1 $\mu$ L	380	380
5 $\mu$ L	1.470	294
10 $\mu$ L	2.050	205

Unter den genannten Bedingungen konnten damit insgesamt etwa 7.900 Transposon-Mutanten erhalten werden. Bereits nach Sichtung der Kolonien konnten morphologische Unterschiede der einzelnen Klone festgestellt werden. Diese umfassten farbliche Änderungen (rötlich bis tief orange, weiß) sowie Unterschiede in der Koloniegröße und -Form (Abbildung 9A). Die Kontrolle weniger Mutanten (n = 7) erfolgte mittels Southern Blot Analyse und zeigte einfache sowie mehrfach erfolgte Transposon-Insertionen an unterschiedlichen Loci im Genom (Abbildung 9B).



**Abbildung 9:** Analyse von Mutanten in *T. thermophilus* nach *in vitro* Transposonmutagenese. **A:** Phänotypische Charakterisierung der angewachsenen Kolonien auf TB Kanamycin (20 µg/mL). **B:** Southern Blot Analyse von sieben verschiedenen Transposonmutanten mit DNA-Sonde (1,2 kb großes Fragment, das mittels *BoxI* aus pMOD3-Kat ausgeschnitten wurde). Die genomische DNA wurde mit *Bam*HI verdaut, die Detektion erfolgte mittels Chemolumineszenz-Aufnahme.

Die Transposon-Mutantenbank wurde nach dem Ausplattieren der Zellen direkt von den Agarplatten abgewaschen und als Mischung der Klone (pool) für eine spätere Weiterverwendung bei -80 °C gelagert. Zur Deaktivierung der lipolytischen Hintergrundaktivität in *T. thermophilus* HB27 wurde die zufällige Transposon-Mutagenesebank nach Phänotypen durchmustert, welche keine oder eine deutlich verminderte Hofbildung auf TB-Substratplatten mit Tributyrin-Overlay zeigten. Es konnte in der Bank jedoch keine Mutante identifiziert werden, die einen deutlich verringerten hydrolytischen Hof verglichen mit Wildtyp-Kolonien zeigte. Zur funktionellen Charakterisierung der Mutantenbank wurden in einer anderen Arbeit Mutanten identifiziert, die auf für *Thermus* sonst toxischen 5-Brom-4-Chlor-3-Indolyl-Substraten wachsen können (Angelov *et al.*, 2013).

#### 1.4. Deaktivierung der lipolytischen Hintergrundaktivität in *T. thermophilus* und Herstellung einer Multideletionsmutante BL03 als Expressionswirt für Genbanken

##### 1.4.1. Deaktivierung ausgewählter Gene zur Verminderung der *T. thermophilus*-eigenen Esterase-Aktivität

Mittels einer *hmmscan* Proteomanalyse aller Pfam-annotierten Proteinfamilien und Domänen in *T. thermophilus* (A. Angelov, unveröffentlichte Ergebnisse), konnten mehrere Kandidatengene identifiziert werden, welche lipolytische Aktivitäten erwarten ließen (Tabelle 22).

**Tabelle 22: *In silico* Proteom-Datensatz vorhergesagter lipolytischer Proteine in *T. thermophilus* HB27.**

Pfam-Eintrag	Beschreibung	Annotation, zusätzliche Informationen	Protein Accession-Nummer	ORF-Nummer (TT_C)
Abhydrolase_6	$\alpha/\beta$ -Hydrolase	Homologie zu Hydrolase / Esterase EstTs1 aus <i>Thermus scotoeductus</i> SA-01 (du Plessis <i>et al.</i> , 2010)	YP_004315.1	0340
Abhydrolase_6	$\alpha/\beta$ -Hydrolase	Liegt im Locus mit TT_C0340	YP_004316.1	0341
Abhydrolase_6	$\alpha/\beta$ -Hydrolase	Carboxylesterase	YP_005310.1	1341
COesterase	Carboxylesterase	Carboxylesterase, Signalpeptid vorhergesagt*	YP_005756.1	1787
Esterase	Putative Esterase	Ferric-Enterobactin Esterase-ähnliches Protein	YP_004722.1	0749
PAE	Pectinacetylerase	Extrazelluläre Esterase*, bereits charakterisiert (Fuciños <i>et al.</i> , 2011)	YP_004875.1	0904
Patatin	Patatin-ähnliche Phospholipase	Hypothetisches Protein	YP_004770.1	0797
zwei PLDc_2	Phospholipase D-ähnliche Domänen	Phospholipase	YP_004797.1	0824
zwei PLDc_2	Phospholipase D-ähnliche Domänen	Hypothetisches Protein, Signalpeptid vorhergesagt*	YP_005054.1	1085
4HBT	Thioesterase Superfamilie	Acyl-CoA Hydrolase	YP_005421.1	1452

\* Signalpeptidvorhersage mittels SignalP Server (<http://www.cbs.dtu.dk/services/SignalP/>)

Unter den identifizierten Genkandidaten wurden neben bereits charakterisierten oder beschriebenen Esterasen auch eine Reihe bisher noch nicht näher untersuchter Proteine aufgefunden. Darunter waren  $\alpha/\beta$ -Hydrolasen, Phospholipasen, eine Thioesterase sowie eine putative Esterase zu finden. Die in der Tabelle genannten Kandidatengene wurden mittels gezielter Genunterbrechungen durch das *Suicide*-Plasmid pBKat (A. Wiezer, Diplomarbeit, 2001) im chromosomalen Locus in *T. thermophilus* HB27 unterbrochen. Dabei wurde stets versucht, mögliche Domänen, aktive Zentren oder den N-terminalen Teil des Zielproteins zu unterbrechen bzw. die genetische Integrität zu stören. Dazu wurden Gensequenzen der ORFs mit Größen zwischen etwa 500 bis 700 bp per PCR amplifiziert und über die jeweiligen Schnittstellen der Primer (siehe Liste der Oligonukleotide in Tabelle 42 im Anhang) in die MCS des Plasmids pBKat ligiert. Nach Übertragung der fertig gestellten Deletionsplasmide in *T. thermophilus* HB27 wurden Mutanten auf TB-Agarplatten mit 20  $\mu\text{g}/\text{mL}$  Kanamycin 2 Tage lang bei 60 °C selektiert. Die Analyse der Insertionsrichtung des Plasmids (parallele und antiparallele Insertionen in das Chromosom von *T. thermophilus*) durch homologe Rekombination am Genlocus erfolgte mittels PCR-Nachweis (Zusammenfassung der Klonierungsergebnisse und Genunterbrechungen in *T. thermophilus* siehe Anhang, Tabelle 39 bis Tabelle 40).

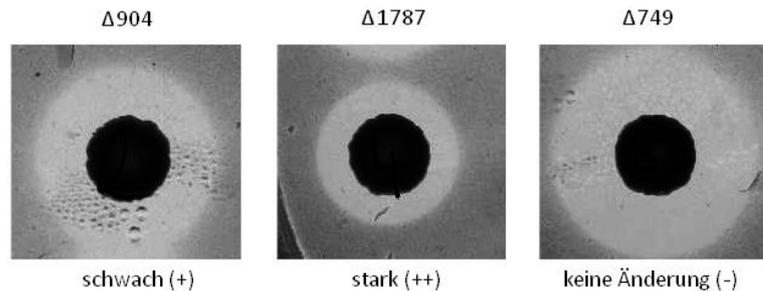
Das gemeinsame Operon der beiden Gensequenzen TT\_C0340 und TT\_C0341 konnte nicht mittels Insertionsmutanten inaktiviert werden, sodass beide ORFs vollständig durch eine Kanamycin-Kassette ersetzt wurden. Hierfür wurden die beiden 1 kbp großen Flanken der Zielgene in den Vektor pCR-TOPO (Invitrogen) kloniert (Primer  $\Delta 340_{\text{for}}/\Delta 341_{\text{rev}}$ ). Zwei *Bg*III-Restriktionsstellen wurden an den äußeren Flanken der kodierenden Sequenzen von TT\_C0340-1 mittels Change-IT Multiple Site Directed Mutagenesis Kit (Affymetrix, Santa Clara, USA) und den Primerpaaren  $\Delta 340_{\text{mut\_for}}$  und  $\Delta 341_{\text{mut\_for}}$  eingefügt. Durch Verdau und Religation des Vektors mit *Bg*III konnten die Zielgene basengenau entfernt werden (resultierender Vektor: pCR- $\Delta 340$ -1). Zwischen die beiden 1 kbp langen Flanken wurde schließlich eine Kanamycin-Kassette aus dem pBKat-Vektor über das 1,22 kb lange *Bam*HI-Fragment in die *Bg*III-Stelle ligiert (resultierendes Deletionsplasmid: pCR- $\Delta 340$ -1::kat).

Unter allen der zehn verschiedenen Gendisruptions-Mutanten zeigten zwei Kandidatengene eine verminderte Hofbildung auf TB-Substratplatten mit 1,0 % Tributyrin (w/v) nach mehrtägiger Inkubation bei 60 °C (2-4 Tage) (Abbildung 10A): Den größten Einfluss hatte dabei die Deaktivierung der Carboxylesterase des ORFs TT\_C1787, gefolgt von der Pectinacetylerase (TT\_C0904). Für weitere Untersuchungen wurden ÜN-Kulturen der beiden Mutanten geerntet, das Zellpellet mittels Ultraschall aufgeschlossen und der Rohextrakt gewonnen. Ebenso wurde die sekretierte Proteinfraction des Mediumüberstandes gesichert. Aus den gewonnenen Proben wurde die Proteinkonzentration mittels Bradford-Test bestimmt und die lipolytischen Aktivitäten der jeweiligen Zell-gebundenen, intra- und extrazellulären Fraktionen mit *p*NP-Substraten getestet (Abbildung 10B). Alle anderen Genunterbrechungen konnten zwar auf Nukleinsäureebene nachgewiesen werden (siehe Anhang, Tabelle 40), hatten aber keinerlei Einfluss auf die Hofbildung zur Folge und wurden daher

nicht weiter untersucht (Daten nicht gezeigt). Die Unterbrechung von ORF TT\_C0904 hatte vor allem eine Abnahme der lipolytischen Aktivität im extrazellulären Medium zur Folge. Dabei war der höchste Aktivitätsverlust bei mittellangen Fettsäureester-Verbindungen wie *p*NP-Caprat (C<sub>10</sub>) zu beobachten. Die intrazelluläre bzw. Zell-gebundene Fraktion war über alle getesteten Substrate hinweg etwa 20 – 30 % weniger aktiv. Im Falle der Deletion von ORF TT\_C1787 zeigte sich eine uniforme Aktivitätsabnahme aller Fraktionen, wobei die des Zell-gebundenen Teils stärker ausgeprägt war.

**A**

Rückbildung der Hydrolysehöfe auf TB-Medium mit 1 % (v/v) Tributyrin:

**B**

Mutante	Fraktion	verbleibende Aktivität (%) auf <i>para</i> -Nitrophenyl-Substraten							
		<i>p</i> NP-C <sub>4</sub>	<i>p</i> NP-C <sub>5</sub>	<i>p</i> NP-C <sub>6</sub>	<i>p</i> NP-C <sub>8</sub>	<i>p</i> NP-C <sub>10</sub>	<i>p</i> NP-C <sub>12</sub>	<i>p</i> NP-C <sub>14</sub>	<i>p</i> NP-C <sub>16</sub>
Δ904	intrazellulär / Zell-gebunden	79,84	77,11	80,45	72,17	72,38	68,47	68,8	85,73
	extrazellulär	80,42	71,75	54,24	56,43	37,86	59,98	64,78	78,18
Δ1787	intrazellulär / Zell-gebunden	67,21	68,59	64,54	59,19	61,57	57,36	54,4	67,8
	extrazellulär	74,4	77,0	67,03	72,31	72,48	74,72	73,99	91,2

**Abbildung 10:** Charakterisierung einzelner Mutanten mit verringerter lipolytischer Aktivität. **A:** Ausprägung der lipolytischen Aktivitäten der beiden HB27 Gendisruptions-Mutanten Δ904, Δ1787 und einer Mutante Δ749 (letzte diente als Negativ-Kontrolle, d.h. es handelt sich um eine Mutante ohne Änderung des Phänotyps gegenüber *T. thermophilus* HB27 Wildtyp). 10 μL von ÜN Kulturen wurden auf TB-Agarplatte mit 1,0 % (v/v) Tributyrin punktiert und für mehrere Tage bei 60 °C inkubiert. **B:** *p*NP-Assay der Gendisruptions-Mutanten Δ904 und Δ1787 in *T. thermophilus*. Cytoplasmatische (bzw. Zell-gebundene) und sekretierte Protein-Fractionen wurden mit dem Wildtyp verglichen (mit entsprechend 100 % relativer Aktivität gegenüber den Mutanten). Getestete *p*NP-Substrate waren verschieden lange *p*NP-Fettsäureester (Butyrat bis Palmitat). Die spezifischen Aktivitätsdaten pro mg Protein wurden bei 60 °C als Mittelwerte dargestellt (Dreifachbestimmung aus biologischen Replikaten). Die *Heatmap*-Repräsentation reicht von 0 % verbleibender, relativer Aktivität (weiß) bis 100 % (rot). Abbildung nach Leis *et al.*, 2014.

Für die zwei Kandidaten-ORFs TT\_C0904 und TT\_C1787 wurden jeweils Kanamycin-Genaustauschmutanten in *T. thermophilus* erzeugt. Dazu wurden ca. 1 kbp lange Flanken der Zielgene

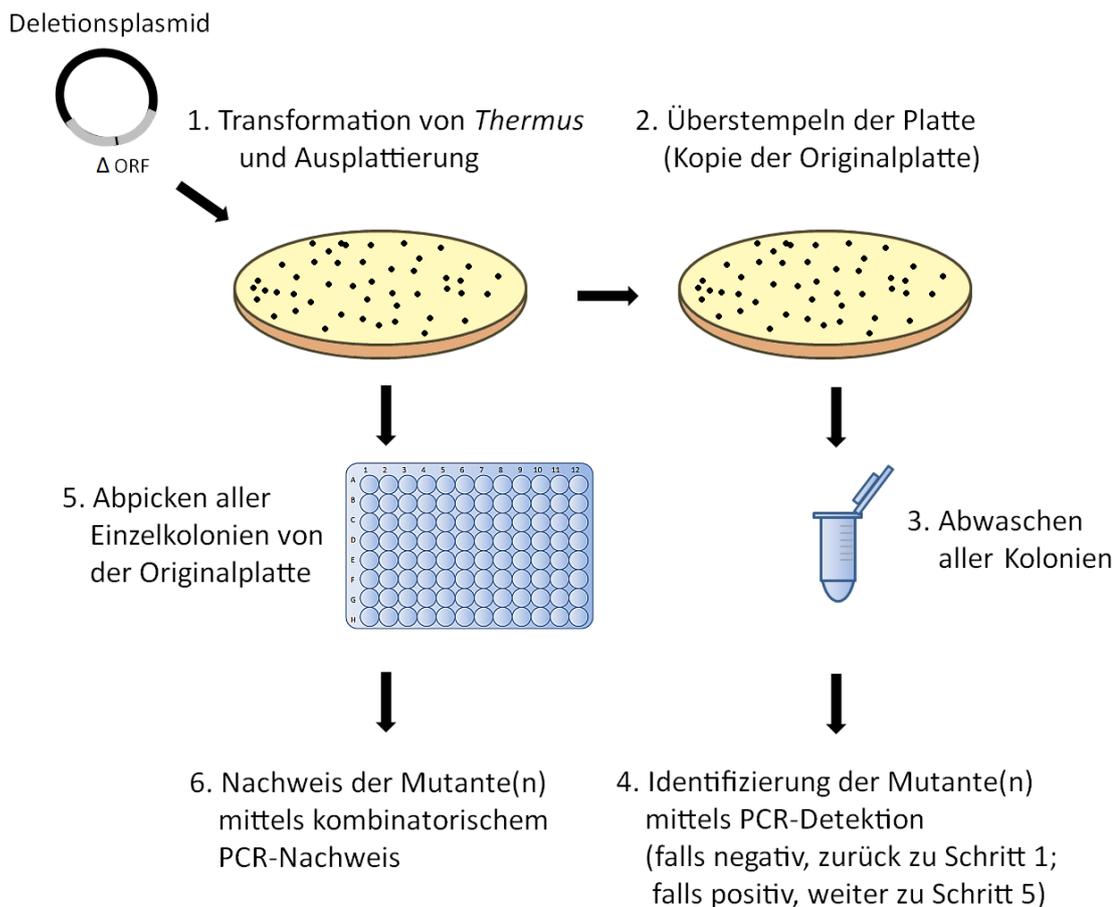
amplifiziert und mittels SOE-PCR fusioniert (*splicing by overlap extension* PCR, Primerpaare  $\Delta 904\_SalI\_for$ ,  $\Delta 904\_SOE\_rev$ ,  $\Delta 904\_SOE\_for$ ,  $\Delta 904\_HindIII\_rev$  sowie  $\Delta 1787\_SalI\_for$ ,  $\Delta 1787\_SOE\_rev$ ,  $\Delta 1787\_SOE\_for$ ,  $\Delta 1787\_HindIII\_rev$ ) und in den pS18a *Knockout*-Vektor (Blas-Galindo *et al.*, 2007) ligiert (die erhaltenen Vektoren wurden pS18a- $\Delta 904$  und pS18a- $\Delta 1787$  genannt). Zwischen den Flanken war über die SOE-Oligonukleotide eine *BclI*-Restriktionsschnittstelle integriert worden, welche den optionalen Einbau einer Kanamycin-Kassette aus pBKat über das 1,22 kb lange *BamHI*-Fragment ermöglichte (da *BclI* methylierungssensitiv ist, wurde dieser Klonierungsschritt in *E. coli* JM109 durchgeführt, welcher nicht über *dam*- oder *dcm*-Methylierungen verfügt). Die resultierenden *Knockout*-Plasmide pS18a- $\Delta 904::kat$  und pS18a- $\Delta 1787::kat$  wurden in *T. thermophilus* HB27 Wildtyp übertragen und auf TB-Medium mit 20  $\mu\text{g}/\text{mL}$  Kanamycin selektiert. Der Effekt der Genunterbrechungen auf die lipolytische Aktivität konnte auch für die jeweiligen Kanamycin-Austauschmutanten auf Substratplatten bestätigt werden (Daten nicht gezeigt).

#### 1.4.2. Markerlose, multiple Gendelektionen in *Thermus*

Mit dem Ziel mehrere Esterasen in einem Stamm gleichzeitig zu deaktivieren, wurde es unabdingbar, markerlose Deletionen am *Thermus* Wildtyp-Stamm durchzuführen. Das etablierte Deletionssystem auf Basis des *rpsL1*-Allels (Blas-Galindo *et al.*, 2007) wurde mit den Vektoren pS18a- $\Delta 904$  und pS18a- $\Delta 1787$  nach Angaben der Autoren verwendet. Dabei wurden die *Knockout*-Vektoren mit den entsprechenden Genflanken in *Thermus* übertragen und auf TB Medium mit 100  $\mu\text{g}/\text{mL}$  Streptomycin ausplattiert. Durch Überprüfen der Streptomycin-Abhängigkeit der Transformanten konnte sichergestellt werden, dass der Vektor homolog ins Chromosom integriert wurde. Nachdem 50 positive, kleine Kandidatenkolonien wieder auf TB Medium ohne Selektionsmarker ausgestrichen wurden, konnten Klone selektiert werden, deren *Knockout*-Plasmid durch Rückrekombination aus dem Chromosom nicht mehr nachweisbar war. Laut Blas-Galindo *et al.* (2007) beträgt der Anteil der Kolonien mit der spezifischen Deletion des Zielgens erwartungsgemäß etwa 50 %. Wie von den Autoren beschrieben können zunächst viele Streptomycin-abhängige Transformanten selektiert werden. Zur Rückrekombination werden alle gewachsenen Kolonien auf TB Medium ohne Selektion ausgestrichen und auf Deletion beider Zielgene überprüft. Trotz zweimaliger Durchführung der Prozedur konnte unter dutzenden „Rückrekombinanten“ jedoch keine Gendelektion per PCR nachgewiesen werden, sodass diese Deletionsstrategie nicht weiter verfolgt wurde.

Eine andere, direkte und neuartige *Knockout*-Strategie zur Erstellung markerloser Mutanten ergab sich aus der Verwendung eines Gens für die  $\beta$ -Glucosidase (*Bgl*, TT\_P0042) in *T. thermophilus* HB27 (A. Angelov und H. Li, unveröffentlichte Ergebnisse). Durch Übertragung eines nicht replikationsfähigen Deletionsplasmids (*Suicide*-Vektor) mit flankierenden Sequenzen von ORF TT\_P0042 ohne zusätzlichen Selektionsmarker konnten auf Anhieb Kolonien mit markerlosen

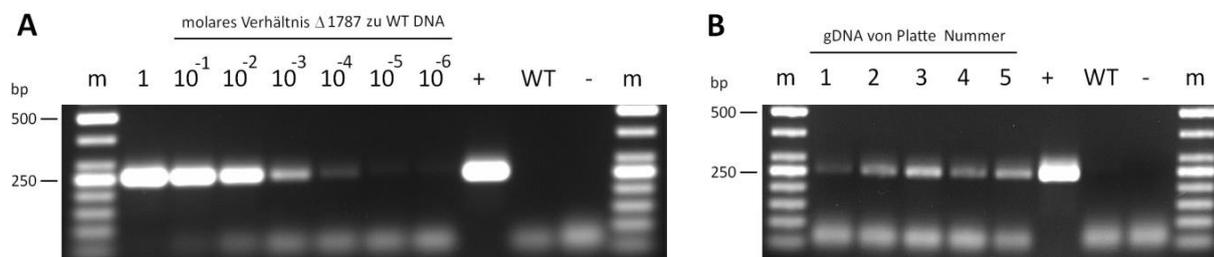
Gendeletionen auf TB Medium mit 5-Brom-4-Chlor-3-Indoxyl- $\beta$ -D-Glucopyranosid (X-Glu bzw. BCI-Glu) beobachtet werden. Auf dem Indikatormedium traten weiße Kolonien auf, die mittels PCR und Southern Blot näher charakterisiert werden konnten. Die Untersuchungen ergaben, dass die Mutanten mit dem Genotyp  $\Delta$ TT\_P0042 mit einer Häufigkeit von etwa 1,0 % aller gewachsenen Kolonien auftraten (Angelov *et al.*, 2013). Obwohl der Mechanismus dieses Phänomens derzeit noch untersucht wird, konnte das Prinzip für die Erstellung multipler Gendeletionen in *Thermus* genutzt werden. Die wesentlichen Schritte zur experimentellen Durchführung sind in Abbildung 11 visualisiert.



**Abbildung 11: Zusammenfassung der wesentlichen Arbeitsschritte zur Erstellung von Markerlosen Gendeletionen in *T. thermophilus*.** Nach Konstruktion des Deletionsplasmids mit je etwa 1 kbp langen Flanken zur Zielgenesequenz ( $\Delta$ ORF) werden 10 bis 15  $\mu$ g des *Suicide*-Vektors in *T. thermophilus* übertragen und anschließend Transformanten auf TB-Festmedium ohne Selektion kultiviert (1). Von der Originalplatte wird eine Plattenkopie überstempelt (2). Nachdem Kolonien gewachsen sind, wird die Platte abgewaschen und die genomische DNA aller Kolonien auf der Platte isoliert (3). Schließlich kann mittels PCR-Detektion die Präsenz einer oder mehrerer Mutanten unter den überwiegend unmodifizierten Zellen (Wildtyp) nachgewiesen werden (4). Die Detektion eines Mutantensignals im Bakteriengemisch lässt auf die Präsenz von mindestens einer Mutante schließen, welche jedoch noch aufgefunden und vereinzelt werden muss. Von der Originalplatte werden daher die Einzelkolonien in 96 Wellplatten überführt und kultiviert (5). Nach der Isolierung der gDNA erfolgt die PCR-Detektion zur Bestimmung der einzelnen Mutante(n) im Block (6).

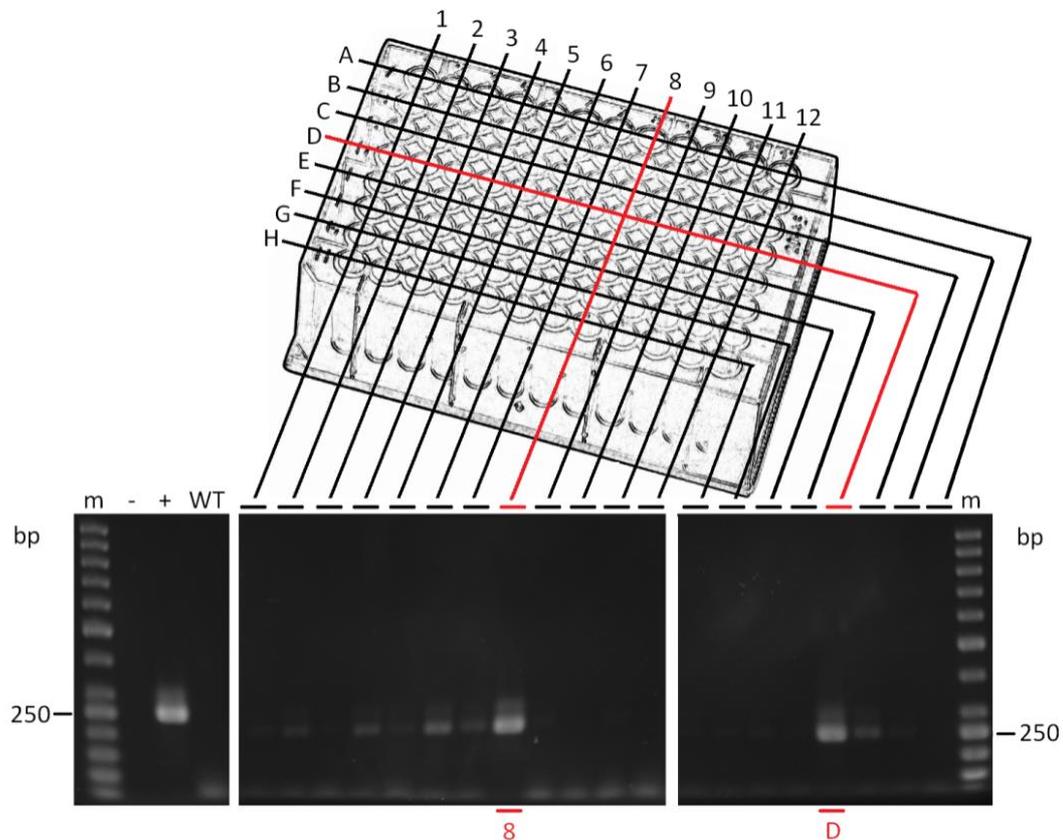
Für die Deletion der Gene TT\_C0904 und TT\_C1787 wurden die bestehenden pS18a-Vektoren (pS18a- $\Delta$ 0904 und pS18a- $\Delta$ 1787, welche jeweils 1 kbp lange Flanken der Gensequenzen tragen) mit *NdeI* verdaut und religiert, um den 745 bp langen Sequenzbereich mit dem *rpsLI*-Allel vollständig zu entfernen. Die resultierenden Vektoren wurden pUC18- $\Delta$ 0904 und pUC18- $\Delta$ 1787 genannt. Für die Gene TT\_C0340 und TT\_C0341 wurde das Plasmid pCR- $\Delta$ 0340-1 verwendet, welches ein Zwischenkonstrukt zur Herstellung von pCR- $\Delta$ 0340-1::kat war (Konstruktion und Klonierungsdaten sind in Tabelle 43 im Anhang zusammengefasst).

Die drei *Knockout*-Vektoren wurden zur sequenziellen Deletion der Esterasen eingesetzt. Insgesamt wurden 15  $\mu$ g des Gendeletionsvektors pUC18- $\Delta$ 1787 für eine Transformationsreaktion mit dem Deletionsstamm *T. thermophilus* HB27  $\Delta$ *bgl* (H. Li, unveröffentlicht) verwendet. Nachdem Verdünnungsreihen auf TB Festmedien ausplattiert wurden, konnten nach 1 Tag bei 70 °C in den Verdünnungsstufen  $10^{-7}$  bis  $10^{-8}$  mit 100 bis 1.000 Kolonien gerechnet werden. Platten mit etwa 100 Einzelkolonien wurden als Kopien auf eine neue TB-Agarplatte überstempelt und inkubiert. Die dort gewachsenen Kolonien wurden mit 50 mM Tris-HCl pH 7,5 abgewaschen und genomische DNA des Gemisches aus Wildtyp und möglichen Mutanten isoliert. Von fünf dieser Plattenkopien wurde eine Detektions-PCR auf die Präsenz einer *Knockout*-Mutante im Gemisch durchgeführt. Die Sensitivität der PCR-Reaktion musste dafür mindestens eine Mutante in hunderten Kolonien nachweisen können. Daher wurde eine Titrationsreihe aus dem *Knockout*-Plasmid pUC18- $\Delta$ 1787 im Gemisch mit Wildtyp genomischer DNA durchgeführt, um die Sensitivität der PCR-Reaktion anzupassen (Abbildung 12A). Nach Anpassung der optimalen PCR-Bedingungen (*Annealing*-Temperatur und Menge der eingesetzten Matrizen-DNA) konnte noch ein Molekül mit der Gendeletion  $\Delta$ 1787 unter 10.000 Molekülen der Wildtyp-Situation identifiziert werden. Die erreichte Sensitivität der PCR machte es außerdem möglich, unter jeder der fünf abgewaschenen Agarplatten ein *Knockout*-Signal zu identifizieren und deren Häufigkeit im Gemisch semi-quantitativ abzuschätzen (Abbildung 12B).



**Abbildung 12: Visualisierung der Detektions-PCR zum Nachweis der Präsenz der Gendeletionsmutante  $\Delta$ TT\_C1787 in *T. thermophilus*.** A: Die molaren Verhältnisse von *Knockout*-Plasmid pUC18- $\Delta$ 1787 zu genomischer WT DNA wurden im Bereich von 1:1 bis zu 1:10<sup>6</sup> getestet. B: Die Detektion unter 100 Kolonien wurde für fünf abgewaschene Agarplatten (1-5) getestet. Die optimalen PCR-Bedingungen waren bei 72 °C *Annealing*-Temperatur (2-Step PCR) mit dem Primerpaar  $\Delta$ 1787\_for und  $\Delta$ 1787\_rev unter Verwendung von 20 ng Matrizen-DNA. Kontrollen waren 50 pg pUC18- $\Delta$ 1787 (+), Wildtyp genomische DNA (WT) und eine Kontrolle ohne Matrizen-DNA (-), als DNA-Größenstandard wurde der GeneRuler 50 bp (m) verwendet.

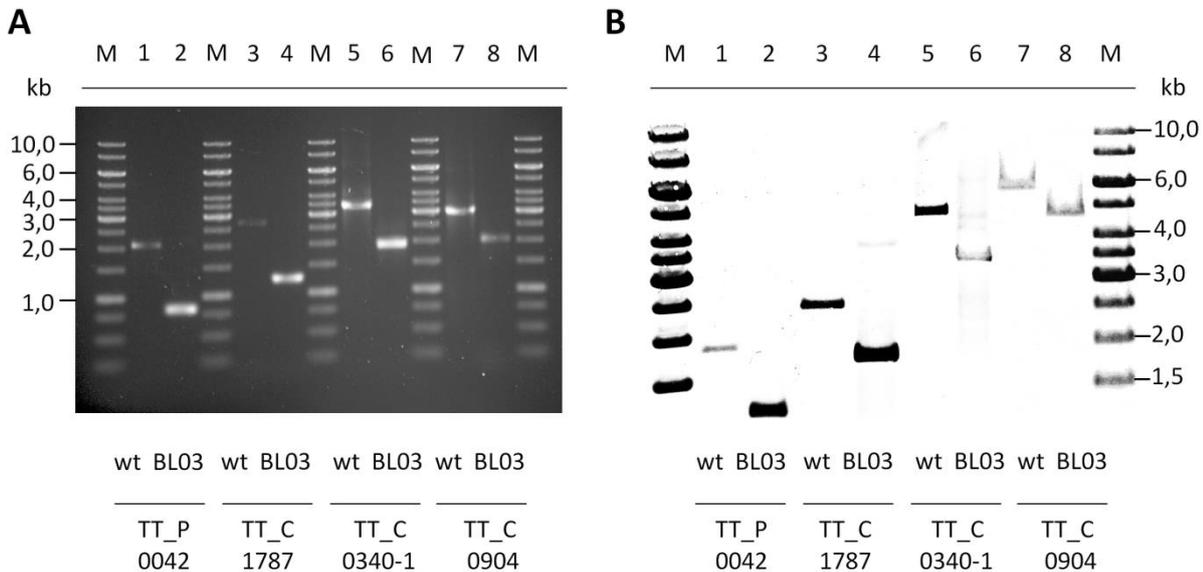
Von der Sicherungskopie der Agarplatte Nummer 2 wurden schließlich 96 Kolonien einzeln in einen 96-Well Block abgepickt und bei 60 °C für einen Tag inkubiert. Zur Isolation der genomischen DNA wurden die einzelnen Klone nach Reihen (A-H) und Spalten (1-12) zusammengeführt und die Detektions-PCR kombinatorisch für jede Reihe und Spalte des 96-Well Blocks ausgeführt (Abbildung 13).



**Abbildung 13:** Schematische Darstellung zur Identifizierung der Gendelektionsmutanten im 96-Well Block mittels kombinatorischem PCR-Nachweis. Die Isolation der DNA aus den zusammengeführten Spalten (1-12) und Reihen (A-H) führte zu 20 DNA-Proben, die mittels PCR die schnelle Identifizierung der Mutante in Mitten des Blocks erlaubte. Hier im Beispiel ist die Detektion der  $\Delta TT\_C1787$ -Mutante gezeigt, welche in Reihe D und Spalte 8 als 255 bp großes PCR-Produkt aufgefunden werden konnte (rote Markierungen). Die PCR-Bedingungen waren optimiert worden (72 °C Annealing, 20 ng Matrizen-DNA).

Von den analysierten Proben konnten mehrere Reihen und Spalten identifiziert werden, welche ein spezifisches Signal für die Gendelektion ergaben. Nachdem die entsprechenden Kandidaten des 96-Well Blocks erneut einzeln getestet wurden, konnte lediglich bei Position D8 die *Knockout*-Bande bestätigt werden. Von diesem Well wurde schließlich ein Verdünnungsausstrich gemacht, um mögliche Kontaminationen mit WT-Zellen auszuschließen. Die Echtheit der Deletion wurde dann mittels Southern Blot-Analyse bestätigt. Der entstandene Stamm mit der Doppelmutation  $\Delta bgl \Delta 1787$  wurde BL01 genannt. Analog zu der oben beschriebenen Vorgehensweise wurden in dem Stamm BL01 die beiden ORFs TT\_C0340-1 deletiert (Genotyp  $\Delta bgl \Delta 1787 \Delta 0340-1$ ). Der entstandene Stamm wurde BL02 genannt.

Auf Basis dieser Dreifachmutante wurde schließlich der ORF TT\_C0904 deletiert, der entstandene Stamm BL03 hatte den Genotyp  $\Delta bgl \Delta 1787 \Delta 0340-1 \Delta 0904$  (Abbildung 14).

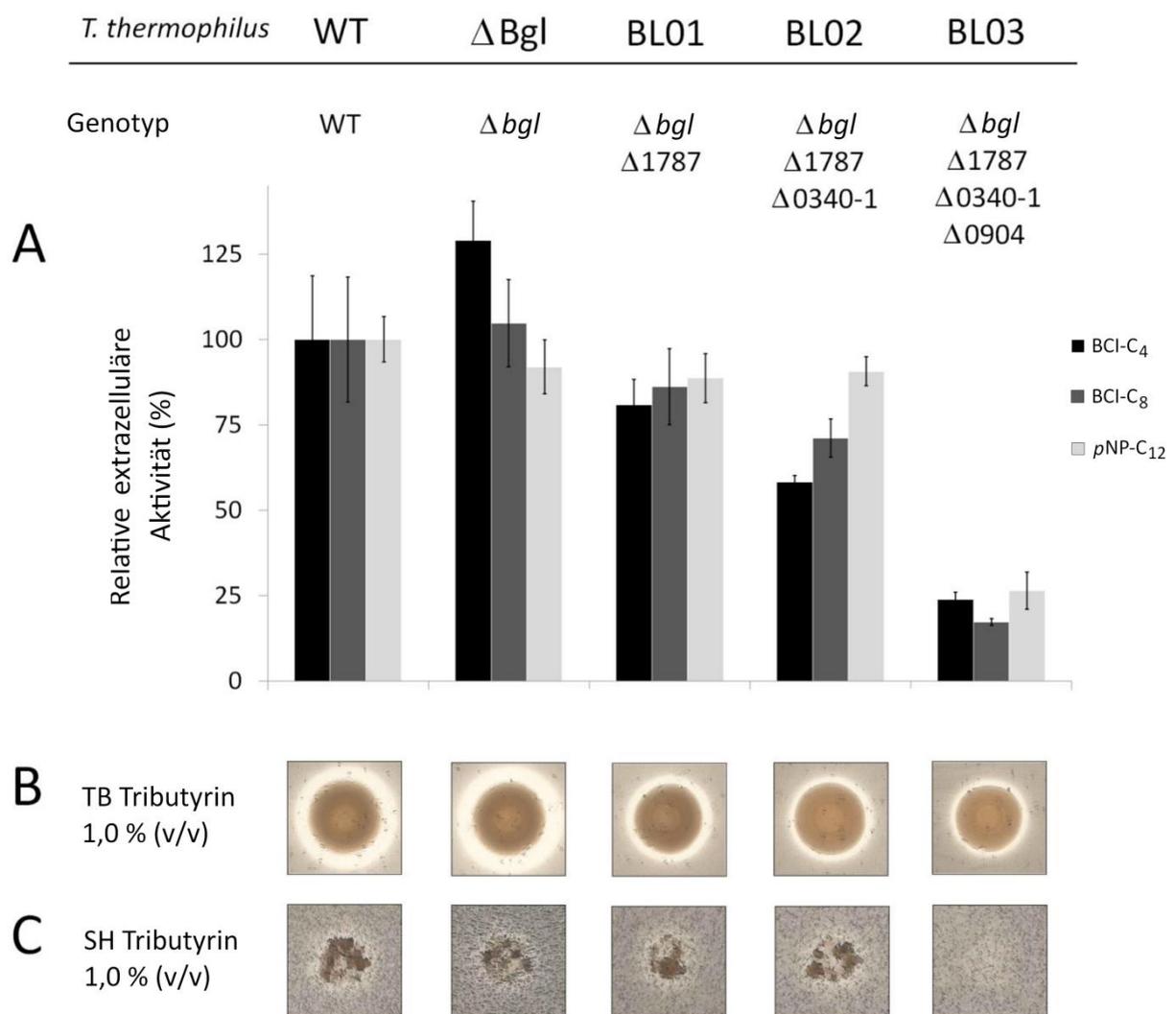


**Abbildung 14: Bestätigung des Genotyps der Vierfachmutante *T. thermophilus* BL03 ( $\Delta bgl \Delta 1787 \Delta 0340-1 \Delta 0904$ ) mittels PCR-Nachweis (A) und Southern Blot Analyse (B).** Für die Hybridisierung der Sonde für die  $\beta$ -Glucosidase (Bgl, TT\_P0042) wurde genomische DNA mit *Bam*HI verdaut, sodass für den Wildtyp (wt, Spur 1) ein 1,88 kb großes Fragment, für die Mutante BL03 (Spur 2) ein 1,21 kb-Fragment sichtbar waren. Für die Detektion der Loci TT\_C1787 und TT\_C0340-1 wurde die chromosomale DNA mit *Sac*II verdaut, sodass die spezifischen Sonden für TT\_C1787 im WT ein 2,55 kb-Fragment (Spur 3) und für BL03 Fragment von 1,80 kb (Spur 4) bzw. 4,80 kb (Spur 5) und 3,33 kb (Spur 6) für TT\_C0340-1 zeigten. Für die Detektion von TT\_C0904 zeigte der WT (Spur 7) ein 5,47 kb langes, die Vierfachmutante BL03 ein 4,48 kb großes Fragment (Spur 8). Als Größenstandard diente der GeneRuler 1 kb Marker (M). Die genaue Herstellung der Sonden ist im Tabelle 43 (Anhang) zusammengefasst (nach Leis *et al.*, 2014).

### 1.4.3. Phänotypische Charakterisierungen der Multideletionsmutante BL03

Der Einfluss der einzelnen Gendeletionen auf die lipolytische Aktivität von *T. thermophilus* wurde qualitativ und quantitativ untersucht. Dabei konnte die Abnahme der sekretierten Esterase-Aktivitäten auf dem Substratmedium als auch mit chromogenen Substraten für alle Einzelmutanten gezeigt werden (Abbildung 15). Mit steigender Zahl an Gendeletionen nahm auch graduell die Esteraseaktivität für 5-Brom-4-Chlor-3-Indoxyl-butyrate (BCI-C<sub>4</sub>) ab, während die Aktivität für länger-kettige Fettsäureester wie BCI-Caprylat und *p*NP-Laurat erst nach Deaktivierung des ORFs TT\_C0904 signifikant abnahm. Insgesamt konnte nach Deletion aller drei Esterasen die lipolytische Aktivität der getesteten Substrate auf etwa ein Viertel der Ausgangsaktivität gesenkt werden. Im Fall der TB Substratplatten mit 1,0 % (v/v) Tributyrin zeigte sich ein ähnliches Bild, mit graduell abnehmender Hofbildung gegenüber dem Wildtyp. Dennoch konnte eine vollständige Deaktivierung der Tributyrase-Aktivität durch multiple Esterasendeletionen nicht nachgewiesen werden. Die Vierfachmutante BL03 war bei längeren Inkubationen auf TB-Substratmedien ebenfalls zur Hofbildung befähigt, obgleich zu durchschnittlich etwa 30 % weniger (Abbildung 15B). Obwohl der

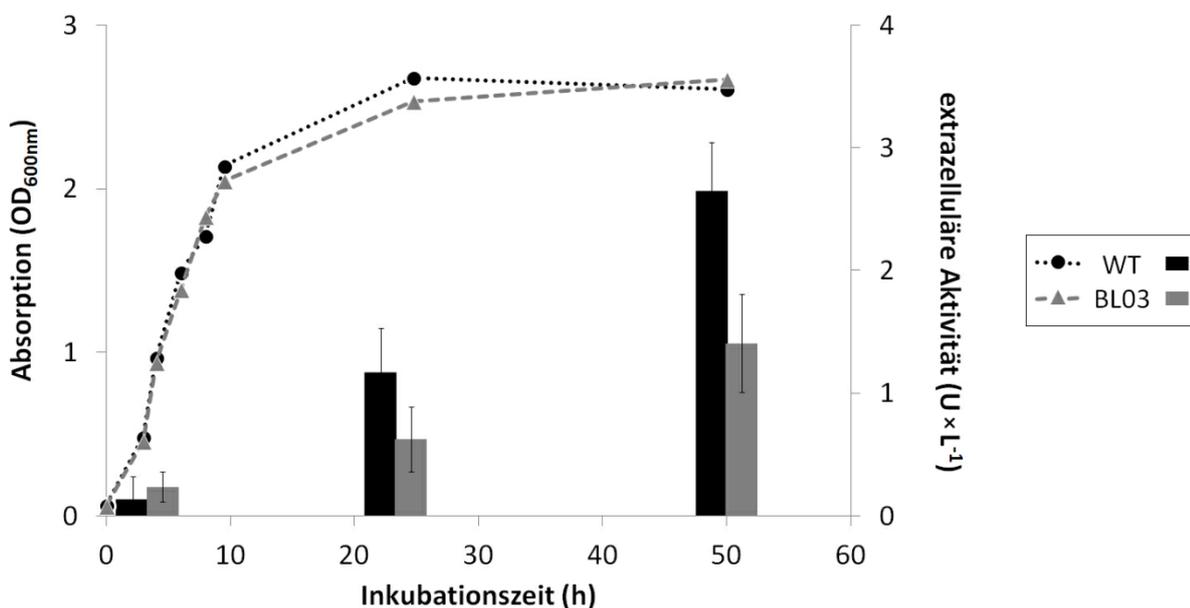
Unterschied der Hydrolysehöfe zwischen dem Wildtyp-Stamm und der multiplen Mutante BL03 nur sehr gering ausfiel, wurde das Wachstum der multiplen Mutante BL03 auf SH-Minimalmedium mit 1,0 % (v/v) Tributyrin ohne weitere definierte Kohlenstoffquelle untersucht. Nach mindestens dreimaligen Waschen einer 5 mL TB ÜN Kultur in 50 mM Phosphatpuffer pH 7,5 wurden verschiedene Volumina zwischen 5, 10 und 20 µL der Zellsuspension auf SH-Agarmedien gespottet. Im Gegensatz zum komplexen, undefinierten TB-Vollmedium vermochte die Mutante BL03 nach 2 bis 3 Tagen nicht auf den Minimalmedien zu wachsen, während der HB27 Wildtyp sichtbares Wachstum zeigte (Abbildung 15C).



**Abbildung 15:** Zusammenfassung der lipolytischen Aktivitäten in *T. thermophilus* HB27 Wildtyp und den jeweiligen Mutanten  $\Delta$ bgl, BL01, BL02 und BL03. Gezeigt wird die Reduktion der sekretierten Esterase-Aktivitäten aus dem Medienüberstand gegenüber *para*-Nitrophenol-Laurat (pNP-C<sub>12</sub>), 5-Brom-4-Chlor-3-Indolyl-butyrat (BCI-C<sub>4</sub>) und -caprylat (BCI-C<sub>8</sub>) im direkten Vergleich (A). Die enzymatischen Reaktionen fanden in 50 mM Tris-HCl pH 7,5 bei 70 °C statt. Die Balkendiagramme zeigen die Mittelwerte von je drei biologischen und technischen Replikaten ( $\pm 1 \times$  Standardabweichung). Die Reduktion der Hofbildung, visualisiert auf TB-Substratplatten mit 1,0 % (v/v) Tributyrin als gespottete Suspensionen nach drei Tagen Inkubation bei 60 °C (B) sowie das Wachstumsverhalten der Mutanten gegenüber dem Wildtyp auf SH-Minimalmedium mit 1,0 % Tributyrin (v/v) bei 70 °C (C). Abbildung nach Leis *et al.* (2014).

Als Kontrolle zu diesem Experiment wurde *T. thermophilus* HB27 auf SH-Minimalmedien unter anderem mit einer nicht verwertbaren Kohlenstoffquelle (z.B. D-Xylose) auf Wachstum hin analysiert. Im Gegensatz zum TB-Vollmedium ( $1,5$  bis  $2,5 \times 10^8$  KBE pro mL Kulturvolumen) traten auf dem SH-Minimalmedium Kolonien mit einer geringen Häufigkeit von  $1,2$  und  $2,0 \times 10^{-6}$  auf.

Um einen wachstumshemmenden Effekt der multiplen Deletionen auszuschließen, welche für die verminderte Aktivität gegenüber dem Wildtyp-Stamm verantwortlich gemacht werden könnte, wurde eine Wachstumskurve mit gleichzeitiger Messung der lipolytischen Aktivität zwischen WT und BL03 untersucht (Abbildung 16). Zur Inokulation der 35 mL TB-Medium wurden vergleichbare Zellmengen verwendet, mit entsprechenden  $OD_{600nm}$ -Werten zwischen  $0,067 \pm 0,004$  (WT) und  $0,062 \pm 0,002$  (BL03) zu Beginn der Inkubation. Der Verlauf der Wachstumskurve bei  $70^\circ\text{C}$  zeigte keine Unterschiede zwischen den untersuchten Stämmen in nährstoffreichen TB-Medium, dennoch akkumulierte der Wildtypstamm über den Verlauf der Wachstumskurve etwa doppelt soviel extrazelluläre Aktivität als die Mutante BL03 (im Durchschnitt  $61,3$  mU pro Liter und Stunde gegenüber  $32,6$   $\text{mU} \times \text{L}^{-1} \times \text{h}^{-1}$ ).



**Abbildung 16:** Untersuchung des Wachstumsverhaltens bei  $70^\circ\text{C}$  und der Ausprägung der lipolytischen Aktivität über die Wachstumsphase. Die Wachstumskurven zwischen *T. thermophilus* WT (schwarze Punkte, gepunktete Linie) und der Vierfachmutante BL03 (graue Dreiecke, gestrichelte Linie). Der Verlauf der sekretierten Esteraseaktivität im Kulturmedium wurde zum Beginn (3 Stunden) und am Ende der exponentiellen Phase (24 Stunden) sowie in der späten stationären Phase (48 Stunden) protokolliert, mit Wildtyp (schwarze Balken) gegenüber BL03 (graue Balken). Die enzymatischen Tests wurden bei  $65^\circ\text{C}$  mit *p*NP-Laurat bestimmt. Gezeigt sind die Mittelwerte der optischen Dichten aus Duplikatmessung von zwei Kulturen ( $n=2$ ) sowie die Aktivitätsdaten als Balkendiagramme ( $n=3$ ) und die jeweiligen Fehlerbalken (als  $\pm 1 \times$  Standardabweichung). Abbildung nach Leis *et al.* (2014).

## 2. Identifizierung und Charakterisierung von biotechnologisch relevanten Enzymen aus diversen metagenomischen Fosmidbanken in *E. coli*

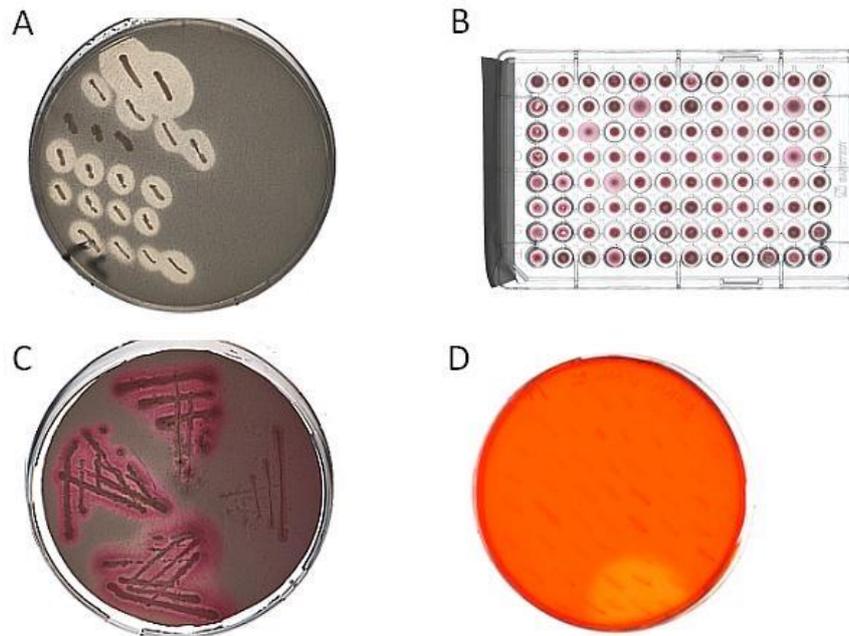
### 2.1. Durchmusterung psychrophiler, mesophiler und thermophiler Metagenombanken

Verschiedene metagenomischen Fosmidbanken, die zum Zeitpunkt der experimentellen Arbeiten verfügbar waren, wurden in *E. coli* EPI300 als Standardwirt auf Aktivitäten hin untersucht (Tabelle 11). Als psychrophile und mesophile Banken standen 2.880 Fosmidklone von Elefantendung (ED), Elbe Sedimente (ES) und Anreicherungskulturen von Mikroorganismen des Schiffsbohrwurms *Teredo navalis* (TN) im Mikrotiterplattenformat (je  $10 \times 96$  Einzelklone pro Bank) bereit (Kooperationspartner W. Streit, Universität Hamburg). Weitere Metagenombanken stammten aus der Tiefsee (Schwarze Raucher), welche als Mischungen (Pools) abgewaschener Kolonien vorhanden waren (P. Golyshin, Bangor University, England). Die Proben wurden auf 20.000 bis 200.000 Einzelklone geschätzt, welche in drei Banken aus mesophilen Bakterien (MB), thermophilen Bakterien (TB) und hyperthermophilen Archaeen (HA) unterteilt wurden.

Für die Durchmusterung der Banken wurden die Einzelklone im Mikrotiterplattenformat überstempelt, bei den Fosmidmischungen (Pools) wurden Verdünnungen ausplattiert und vereinzelte Kolonien untersucht. Überstempelte Kopien der Substratplatten stellten sicher, dass aktive Klone wieder aufgefunden und auf neue Medien passagiert werden konnten. Die getesteten Substrate (vgl. Tabelle 13) waren Substratagarplatten (Tributyryn, Triolein, CMC, Xylane und Stärke) und Mikrotiterplatten mit gefärbten Substraten (HEC, CMC und Xylane). Die Inkubationsbedingungen waren 8 °C für eine Woche, 15 – 20 °C für eine Woche, 30 – 37 °C für mehrere Tage sowie 60 bis 70 °C für zwei bis drei Tage. Aktive Klone konnten mittels Hofbildung oder farbliche Aufklarungen um die Kolonien sichtbar gemacht werden. Lipolytisch aktive Fosmidklone bildeten durchsichtige Höfe auf den sonst trüben Emulsionen von Tributyrin während die Triolein-Hydrolyse durch Zugabe von 0,001 % Rhodamin B (w/v) unter UV-Licht als fluoreszierender Hof um die Kolonien sichtbar war. Der Nachweis des Abbaus von cellulolytischen und hemicellulolytischen Substraten erfolgte mit Kongorot, der Stärkeabbau konnte mittels Lugol'scher Lösung sichtbar gemacht werden. Bei den gefärbten Substraten (HEC, CMC und Xylane) in den Mikrotiterplatten geschah der Nachweis über die Freisetzung des sonst unlöslichen Farbstoffes in den Überstand. Zusammenfassend konnten aus diesen funktionellen Durchmusterungen in *E. coli* für alle getesteten Substrate (außer auf Hydroxyethylcellulose) und über alle Temperaturbereiche aktive Klone aufgefunden werden (Tabelle 23). Eine Zusammenstellung ausgewählter aktiver Klone ist in Abbildung 17 beispielhaft gezeigt.

**Tabelle 23: Zusammenfassung aller aktiver *E. coli* Fosmidklone aus funktionellen Durchmusterungen.**

<b>Aktivität</b>	<b>Substrat (Konzentration)</b>	<b>Temperatur</b>	<b>Anzahl aktiver Klone (Bank, Platte, Position)</b>
Esterase	Substratplatte mit Tributyrin (1% v/v)	8 °C	<b>Mesophile Bakterien (MB): 4</b> MB-est-1 bis 4 <b>Thermophile Bakterien (TB): 2</b> TB-est-1 bis 2
		15 – 20 °C	<b>Elefantendung (ED): 5</b> ED-22-(G1, H2) ED-23-(D12, E12) ED-29-(H12) <b>Elbe Sediment (ES): 2</b> ES-7-(D11, E10)
		37 °C	<b>Mesophile Bakterien (MB): 6</b> MB-est-5 bis 10 <b>Thermophile Bakterien (TB): 8</b> TB-est-3 bis 10
		60 °C	<b>Thermophile Bakterien (TB): 5</b> TB-est-11 bis 15
Lipase	Substratplatte mit Triolein (1% v/v) und 0,001 % (w/v) Rhodamin B	15 – 20 °C	<b>Elbe Sediment (ES): 2</b> ES-7-(D11, E10)
		37 °C	<b>Thermophile Bakterien (TB): 4</b> TB-lip-1 bis 4
Cellulase	Mikrotiterplatte mit Cibacron Brilliant Red 3B-A gefärbter Hydroxyethylcellulose (HEC)	15 – 65 °C	keine Aktivitäten
	Substratplatte mit 0,1% (w/v) Carboxymethylcellulose (CMC)	70 °C	<b>Hyperthermophile Archaeen (HA): 1</b> HA-cmc-1
Xylanase	Mikrotiterplatte mit Cibacron Brilliant Red 3B-A gefärbtem Haferspelz-Xylan	30 °C	<b>Elefantendung (ED): 8</b> ED-21-(B6) ED-22-(C2, C12) ED-25-(A3) ED-26-(C9) ED-27-(B10, B11) ED-30-(A8) <b>Elbe Sediment (ES): 6</b> ES-6-(B5, B11, C3, D11, E4, H4) <b>Teredo navalis (TN): 1</b> TN-9-(B5)
		65 °C	<b>Teredo navalis (TN): 1</b> TN-2-(A8)
		37 – 60 °C	<b>Thermophile Bakterien (TB): 6</b> TB-xyl-1 bis 6
Amylase	Lösliche Stärke 0,3% (w/v)	60 °C	<b>Thermophile Bakterien (TB): 1</b> TB01-(A10) Als Zweitaktivität in MB-est-3,4 und MB-est-6 aufgetreten

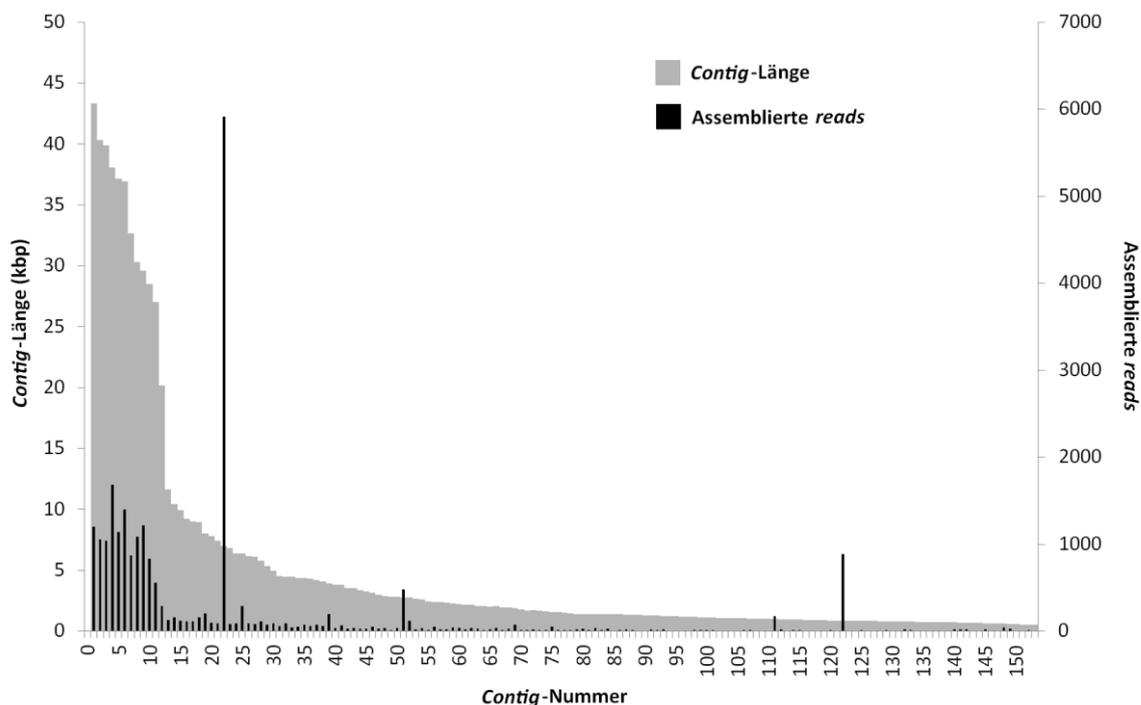


**Abbildung 17:** Ausgewählte Fosmidklone mit Aktivitäten auf unterschiedlichen Substraten. A: Hofbildung durch Abbau von Tributyrin (vereinzelte Fosmidbank-Klone aus der Tiefsee, mesophile Bakterien). B: Mikrotiterplatte mit Fosmidbank der Elbe Sedimentproben mit gefärbten Xylan. Aktive Klone waren ES-6 mit Positionen B5, B11, C3, D11, E4 und H4. C: Substratplatte mit Triolein und Rhodamin B (thermophile Bakterien). D: Vereinzelte Fosmidklone und nachgewiesene Aktivität des Klons HA-cmc-1 bei 70 °C auf Carboxymethylcellulose (CMC) nach Kongorot-Detektion.

Zusammenfassend konnten aus tausenden durchmusterten Fosmidklonen aus allen Banken mehrere dutzend ( $n = 60$ ) aktive Klone isoliert werden. Von den gemischten Fosmidklonen aus der Tiefsee-Bank wurden zehntausende vereinzelt Kolonien auf Substratplatten überprüft, wobei 45 aktive Klone identifiziert und gesichert werden konnten. Nach Analyse des Restriktionsmusters zwischen den isolierten Fosmiden konnten acht Redundanzen festgestellt werden, welche auf jeweils identische metagenomische Fragmente zurückgeführt werden konnten (die Zahl und Länge der Fragmente nach Restriktionsverdau mit *EcoRI* und *BamHI* stimmten größtenteils oder vollständig überein, Daten nicht gezeigt). Die Zahl der nicht-redundanten und anhand des Verdauemusters als unterschiedlich eingestuft Fosmidklone reduzierte sich damit auf 37. Die meisten davon (29) waren lipolytisch aktive Klone, gefolgt von Aktivitäten auf Xylan (6) sowie auf Stärke (1). Lediglich in einem Fall konnte aus der Bank der hyperthermophilen Archaeen (HA) ein positiver *E. coli* Klon identifiziert werden. Die beobachtete Aktivität konnte nur bei erhöhten Temperaturen zwischen 60 und 70 °C nach zwei Tagen Inkubation auf LB-Substratplatten mit Carboxymethylcellulose (0,1 % w/v) nachgewiesen werden. In den Genbanken aus diversen mesophilen Metagenomen (Elbesediment, Elephantendung und Schiffsbohrwurm *Teredo navalis*, W. Streit, TU Hamburg), wurden sieben Tributyrin-aktive Klone und 16 hemicellulolytische Klone entdeckt. Darunter befand sich ein Klon (TN-2-A8), der bei 65 °C Xylanabbau zeigte.

## 2.2. Sequenzierung und *in silico* Analyse der aktiven Fosmide

48 ausgewählte Fosmidklone wurden in einer Mikrotiterplatte angeordnet und im Göttingen Genomics Laboratory (G<sub>2</sub>L, Georg-August Universität Göttingen) mittels 454 Pyrosequenzierung parallel sequenziert und assembliert (Projektname *ANG48*, A. Thürmer). Die Zusammenfassung der Assemblierungsergebnisse und daraus resultierenden *contig*-Sequenzen ist in Abbildung 18 dargestellt. Es konnten 153 *contigs* mit Sequenzgrößen über 500 bp erhalten werden, wobei 24.640 einzelne *reads* zu 746,8 kbp Sequenzinformation assembliert wurden. Um die einzelnen Fosmidklone den dazugehörigen *contigs* zuzuordnen (*mapping*), wurden die Fosmide vom Ende her (mit Primer FP-FOS und RP-FOS für pCC1FOS sowie T7.for und KOF-Primer für pCT3FK-Fosmide) ansequenziert und auf überlappende Sequenzbereiche der *contigs* hin analysiert. Jedoch konnten nur neun Fosmide von beiden Seitenenden ansequenziert werden und gelten damit als vollständig sequenziert (*contigs* 1 bis 9). Sechs weitere *contigs* wiesen assemblierte Sequenzen über 10 kbp auf, wovon jedoch nur ein Teil den korrespondierenden Fosmidklonen zugeordnet werden konnte (*contig* 10 und 12). Ferner waren zwei Sequenzfragmente unter 10 kbp Länge (*contig* 17 und 25) ebenfalls bestimmten Fosmidklonen zuordenbar.



**Abbildung 18:** Statistik der 454-Sequenzierungsergebnisse der Fosmid-Bank *ANG48*. Für jedes *contig* über 500 bp Sequenz (*contig* 1 bis 153) ist jeweils die Zahl der assemblierbaren *reads* (rechte Achse, schwarze dünne Balken) sowie der daraus resultierenden assemblierten Fosmidsequenz (in kbp, linke Achse, graue Balken) dargestellt. *contigs* 22 und 122 zeigten eine signifikant erhöhte Sequenzabdeckung, die jeweils auf Sequenzteile des Fosmidrückgrates (pCC1FOS, pCT3FK) zurückzuführen waren.

Die Sequenzierung der metagenomischen Nukleinsäuren lieferte insgesamt acht vollständige Fosmide, die zur bioinformatischen Auswertung bereitstanden (Zusammenfassung der *in silico*-Analyse in Tabelle 24). Von diesen konnten für drei Fosmide (HA-cmc-1, TB-est-7 sowie ES-6-B5) eindeutig zuordenbare Genprodukte über Ähnlichkeitssuchen identifiziert werden, welche für die beobachtete Aktivität im Fosmidklon verantwortlich gemacht werden können. Die anderen fünf Fosmidsequenzen ergaben keine eindeutigen oder bioinformatisch vorhersagbaren Kandidatengene (TB-xyl-3, TB-xyl-4, TN-9-B5, TB01-A10 und ES-7-E10). Unvollständig sequenzierte Fosmide ergaben in vier Fällen ebenfalls potentielle ORFs und Genprodukte mit vorhergesagter Funktion. Die Suche nach vollständigen oder partiell sequenzierten ORFs in den restlichen *contig*-Sequenzen führte zur Identifizierung von sechs Glycosidhydrolasen (meist vorhergesagte Xylanasen) und einem womöglich lipolytischen Enzym mit Ähnlichkeit zu  $\alpha/\beta$ -Hydrolasen, welches als hypothetisches Protein annotiert ist (Domäne unbekannter Funktion DUF 3089). Für diese Sequenzen war jedoch keine bioinformatische Zuordnung (*mapping*) zu den korrespondierenden Fosmidklonen möglich. Tabelle 24 gibt die Ergebnisse aller identifizierbaren Kandidatengene und Fosmidzuordnungen wieder.

Als einziger lipolytischer Klon konnte Fosmid TB-est-7 vollständig sequenziert werden. Das Fosmid umfasst 29,69 kbp Sequenzinformation bei einem GC-Gehalt von 59,8 %. Unter den insgesamt 33 vorhergesagten ORFs konnte ORF Nummer 32 als Patatin-ähnliche Phospholipase identifiziert werden, die laut *blastp* mit 70 % höchste Sequenzübereinstimmung auf Aminosäuresequenzebene zum Genprodukt von Rmar\_1682 in *Rhodothermus marinus* DSM4252 hat (Proteinaccession: YP\_003290956.1, *e-value* 0,0). Das Zielgen umfasst 954 Basen und kodiert für ein 317 Aminosäuren langes Protein mit berechnetem Molekulargewicht von 34,93 kDa. Ein für lipolytische Enzyme bekanntes Pentapeptidmuster (G-X-S-X-G) ist vorhanden, darüber hinaus ließen sich die beteiligten Aminosäurereste des aktiven Zentrums vorhersagen (Serin 49 und Asparaginsäure 161). Die Präsenz eines Signalpeptides lies sich nicht nachweisen. Das Zielprotein wird momentan näher charakterisiert (V. T. T. Pham, Lehrstuhl für Mikrobiologie).

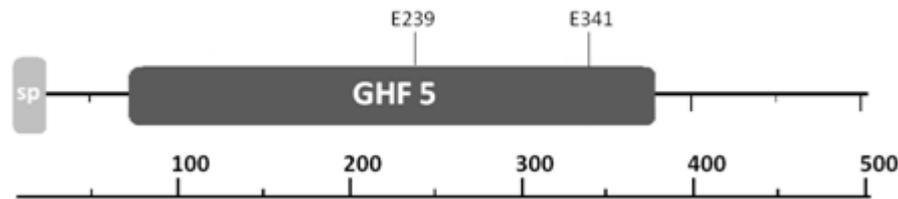
Der hemicellulolytisch aktive Fosmidklon ES-6-B5 konnte *contig* 8 zugeordnet werden und umfasst 30,29 kbp assemblierte metagenomische Sequenz mit höchster Übereinstimmung zu dem marinen Gamma-Proteobakterium HTCC2148 (GC-Gehalt 59,2 %). Es waren insgesamt 23 ORFs vorhersagbar, wovon zwei putative Glycosidhydrolase-Sequenzen identifiziert werden konnten: Eine Chitinase der Familie GHF 18 (ORF 18) sowie eine putative Xylanase (ORF Nummer 23c). ORF23c umfasst 1524 bp Länge und kodiert für ein 507 Aminosäuren großes Protein, im Folgenden Xyn5A benannt. Das berechnete Molekulargewicht entspricht 55,45 kDa. Ein 24 Aminosäuren langes Signalpeptid wurde mittels SignalP-Programm vorhergesagt, wobei voraussichtlich zwischen Ala(24) und Arg(25) abgespalten wird (*D-score* Wert von 0,424).

**Tabelle 24: Zusammenfassung aller sequenzierten und assemblierbaren metagenomischen Fosmide.**

Fosmid	Aktivität (Temperatur)	Assembliertes <i>contig</i> (Größe)	Vorhergesagte Kandidatengene (höchste Ähnlichkeit lt. <i>blastp</i> )
<b>Vollständig sequenzierte pCC1FOS Fosmide mit Kandidaten*:</b>			
HA-cmc-1	Carboxymethyl-cellulose (70 °C)	<i>contig</i> 4 (38,2 kb)	Endo-1,4-β-Glucanase b, GHF12, CBM2 ( <i>Thermococcus</i> sp. AM4)
ES-6-B5	Xylan (30 °C)	<i>contig</i> 8 (30,3 kb)	Putative Glycosidase, GHF5 ( <i>Rhodococcus erythropolis</i> PR4)
TB-est-7	Tributyryn (60 °C)	<i>contig</i> 9 (29,7 kb)	Patatin-ähnliche Phospholipase ( <i>Rhodothermus marinus</i> DSM 4252)
<b>Vollständig sequenzierte Fosmide, aber nicht eindeutig zuordenbare Gene/Proteine:</b>			
TB01-A10	Stärke (60 °C)	<i>contig</i> 1 (43,4 kb)	Kein Kandidat ( <i>Rhodothermus marinus</i> -ähnlich)
TB-xyl-4	Xylan (60 °C)	<i>contig</i> 2 (40,4 kb)	Kein Kandidat ( <i>Rhodothermus marinus</i> -ähnlich)
TB-xyl-3	Xylan (60 °C)	<i>contig</i> 5 (37,3 kb)	Kein Kandidat ( <i>Rhodothermus marinus</i> -ähnlich)
TN-9-B5	Xylan (30 °C)	<i>contig</i> 6 (36,9 kb)	Kein Kandidat ( <i>Opitutus terrae</i> -ähnlich)
ES-7-E10	Tributyryn/Triolein (20 °C)	<i>contig</i> 7 (32,7 kb)	Kein Kandidat
<b>Unvollständig sequenzierte Fosmide mit möglichen Kandidaten:</b>			
MB-est-2 / 3 / 5 / 6	Tributyryn (37 °C)	<i>contig</i> 10 (28,5 kb)	Hypothetisches Protein, α/β-Hydrolase 5 ( <i>Pseudoalteromonas</i> sp. Bsi20429)
MB-est-2 / 3	Tributyryn (37 °C) Stärke (RT)	<i>contig</i> 12 (20,2 kb)	α-Amylase, GHF15 ( <i>Pseudoalteromonas haloplanktis</i> )
TN-2-A8	Xylan (65 °C)	<i>contig</i> 17 (9,1 kb)	Glycosidhydrolase, GHF28 ( <i>Opitutus terrae</i> PB90-1)
MB-est-7 / 8	Tributyryn (37 °C)	<i>contig</i> 25 (6,4 kb)	Carboxylesterase Est50 ( <i>Geobacillus thermoleovorans</i> CCB_US3_UF5)
<b>Nicht zuordenbare contig-Sequenzen(kein mapping möglich):</b>			
<i>contig</i> 27 (6,1 kb)	Hypothetisches Protein, Protein unbekannter Funktion (DUF 3089), Ähnlichkeit zu α/β-Hydrolase 1 und 6 ( <i>Eubacterium</i> sp. 14-2)		
<i>contig</i> 33 (4,5 kb)	Glycosidhydrolase, GHF5 ( <i>Kribbella flavida</i> DSM 17836)		
<i>contig</i> 41 (3,8 kb)	Xylanase XIB, GH11, DUF1080 ( <i>Cellulomonas flavigena</i> )		
<i>contig</i> 49 (3,5 kb)	Putative Endo-1,4-β -Xylanase, GHF43, 2x CBM6 ( <i>Fibrobacter succinogenes</i> subsp. <i>succinogenes</i> S85)		
<i>contig</i> 61 (2,2 kb)	Endo-1,4-β -Xylanase, GHF10, Xylosidase/Arabinosidase, GHF43 ( <i>Bacteroides</i> Xylosidase/Arabinosidase ( <i>Bacteroides</i> sp.))		
<i>contig</i> 88 (1,4 kb)	Xylanase, GHF10 ( <i>Streptomyces hydroscopicus</i> )		
<i>contig</i> 100 (1,1 kb)	Glycosidhydrolase, GHF10 ( <i>Rhodothermus marinus</i> DSM 4252)		

\*: Die pCT3FK-Fosmidsequenz von *contig* 3 und 11 werden im späteren Abschnitt erwähnt (siehe Tabelle 27).

Die Analyse der beteiligten Proteinmodule sagten eine Glycosidhydrolase der Familie 5 voraus (schematische Repräsentation des Proteins in Abbildung 19).



**Abbildung 19: Vorhersage der Domänenzusammensetzung von Xyn5A.** Das N-terminale Signalpeptid umfasst 24 Aminosäuren (Abspaltung zwischen Aminosäure NGA | RD). Die in Pfam vorhergesagte Glycosidhydrolasedomäne der Familie 5 umfasst die Aminosäurepositionen 72 bis 376 ( $e$ -value  $3,4 \cdot e^{-22}$ ). Die beiden Glutaminsäuren an Position 239 und 341 gehören zum katalytischen Zentrum des Enzyms.

Zur Expression wurde Xyn5A mit C-terminalen 6xHis-*tag* in den pET21a(+)-Expressionsvektor mittels isothermaler Assemblierung (Gibson Assembly) über *NdeI* und *XhoI* kloniert (Primerpaar ES6-B5-23c\_F/R, resultierender Vektor pET21a-ES6B5-23chis). *E. coli* BL21 Star Expressionskulturen wurden mit 1,0 mM IPTG induziert und über Nacht inkubiert. Zusätzlich zum Kandidatengen wurde die zweite putative Glycosidhydrolase Chi18A (kodiert durch ORF 18) parallel zu ORF 23c wie oben beschrieben kloniert und exprimiert (Primerpaar ES6-B5-18\_F/R, Vektor pET21a-ES6B5-18his). Im Gegensatz zur Leervektorkontrolle zeigte der Rohextrakt der Xyn5A-Expression nachweisbare Freisetzung reduzierender Zuckerverbindungen im kolorimetrischen DNSA-Test auf Xylan (Daten nicht gezeigt). Aufgrund der geringen Proteinausbeuten scheiterte eine Aufreinigung mittels FPLC und HisTrap-Säulen. Das Genprodukt von ORF 18 wies keine nachweisbare Aktivität auf Xylan, CMC oder Chitin auf (Daten nicht gezeigt).

Ein weiterer Xylan-abbauender Klon TB-xyl-3 lag ebenfalls vollständig sequenziert vor (*contig 5*, 37,3 kbp Größe und 31 kodierende ORFs). Der GC-Gehalt liegt bei 58,1 %. Mittels bioinformatischer Analyse konnte nur ORF 09c identifiziert werden, welcher für eine potentielle Glycosidhydrolase der Familie 18 kodiert. Die Nukleotidsequenzlänge beträgt 1137 Basen, das 378 Aminosäuren lange Produkt zeigt höchste Ähnlichkeit mit einer Glycosidhydrolase (YP\_003290622.1) aus *Rhodothermus marinus* DSM 4252 (76 % Sequenzübereinstimmung,  $e$ -value 0,0). Die Präsenz eines N-terminalen Signalpeptides war nicht signifikant, jedoch nahe an der Detektionsgrenze ( $D$ -score von 0,566 gegenüber dem Grenzwert von 0,570). Das putative thermostabile Enzym (Xyn18A) wurde als Protein voller Länge mit Signalpeptid (42,55 kDa) ohne His-*tag* in pET21a(+) kloniert (PCR-Produkt von Primerpaar TB-xyl-3-9c\_F/R mittels Gibson Assembly in einen mit *NdeI* und *NotI* geschnittenen pET21a(+)-Vektor) und in *E. coli* Stamm BL21 Star exprimiert (Vektor pET21a-TBxyl3-9c). Nach über Nacht Induktion mit 1,0 mM IPTG bei 37 °C konnte das Zielprotein mittels SDS-PAGE nachgewiesen werden. Im Zuge der Aufreinigung wurde der Großteil des Proteins jedoch als Präzipitat in der unlöslichen Fraktion des hitzegefällten Rohextraktes aufgefunden (Daten nicht gezeigt).

In vier Fällen führte die *in silico* Analyse der erhaltenen Sequenzdaten zu keinen plausiblen Kandidatengen. Für den lipolytischen Klon ES-7-E10 konnte mittels Verdünnungsausstrich eine Kontamination der Kultur festgestellt werden, die bereits zum Zeitpunkt der funktionellen Durchmusterungen vorhanden war und für die beobachtete Hydrolyse von Tributyrin und Triolein verantwortlich ist. Da der vereinzelte Klon in Reinkultur keine nachweisbare lipolytische Aktivität zeigte, wurde er daher nicht weiter untersucht.

Um weitere Kandidatengene zu identifizieren wurde das Fosmid des hemicellulolytischen Klons TB-xyl-4 mittels Transposonmutagenese analysiert. Unter mehreren Dutzend analysierten Klonen zeigte eine dieser Mutanten (1-D4) einen unerwarteten und deutlich erhöhten Hydrolysehof auf Xylan-Substratplatten gegenüber den anderen Kolonien bei 60 °C. Die Mutante wurde von den Enden des Transposons sequenziert, um die Fosmidsequenz des betroffenen Locus festzustellen. Es konnte keine Sequenzübereinstimmung zur Fosmidsequenz von TB-xyl-4 festgestellt werden, jedoch konnte die Transposonsequenzierung mit der Genomsequenz aus *R. marinus* DSM4252 erfolgreich assembliert werden. In ihrer direkten chromosomalen Sequenzumgebung wurden mehrere Glycosidhydrolasen als Kandidatengene identifiziert, welche für Xylanasen kodieren können.

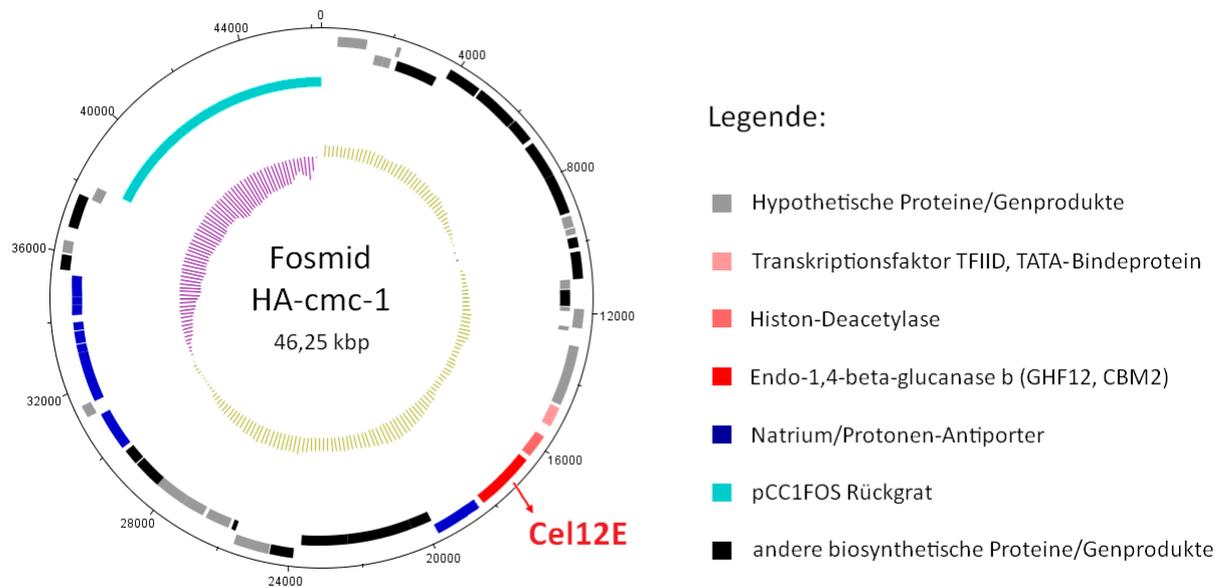
Der amylolytische Fosmidklon TB01-A10 wurde mittels Mauve auf vergleichende Sequenzabschnitte hin analysiert. Die vorhandene Sequenz lies sich in drei Homologieblöcke (*similarity blocks*) aufteilen, die die höchste Übereinstimmung gegenüber nicht zusammenhängenden Sequenzabschnitten des *R. marinus* DSM4252 Genoms zeigten. Es war kein ORF als Kandidat bestimmbar.

Für den Xylan-abbauenden Klon TN-9-B5 konnten mittels Mauve ebenfalls drei übereinstimmende *similarity blocks* festgestellt werden, die ähnlich zum *Opitutus terrae* PB90-1 Genom sind. Die Suche nach spezifischen Gensequenzen, die für hemicellulolytische Enzyme kodieren, verlief ergebnislos.

### **2.3. Charakterisierung einer neuartigen, archaellen und hyperthermostabilen Endoglucanase Cel12E**

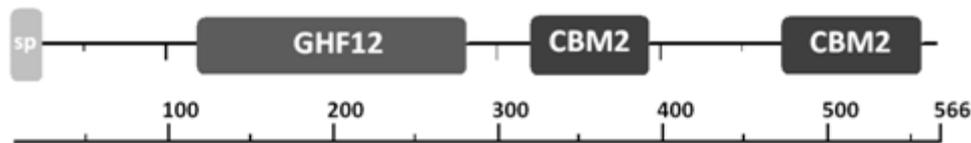
#### **2.3.1. Analyse des Fosmids HA-cmc-1 und des beteiligten ORFs**

Eines der in den funktionellen Durchmusterungen isolierte und vollständig sequenzierte Fosmid aus dem cellulolytischen *E. coli*-Klon HA-cmc-1 wurde nach erfolgreicher Genvorhersage weiter analysiert. Das insgesamt zu 38.175 bp assemblierte Fosmid zeigte lt. *blastn* höchste Sequenzübereinstimmung auf Nukleinsäureebene zu dem Archaeon *Thermococcus sp.* AM4, bei einem GC-Gehalt von 55,2 %. Es waren insgesamt 48 ORFs vorhersagbar, welche in Abbildung 20 schematisch wiedergegeben sind.



**Abbildung 20:** Darstellung des sequenzierten Fosmids HA-cmc-1, welches für cellulolytische Aktivität kodiert. Schwarze Markierungen stellen vorhergesagte ORFs dar, graue Markierungen kodieren für hypothetische Proteine bzw. für Proteine ohne Funktionszuordnungen. Das Zielgen Nummer 23 (Endo-1,4- $\beta$ -Glucanase b) sowie seine unmittelbare Sequenzumgebung sind farblich markiert. Alle Sequenz-basierten Ähnlichkeiten sowie Proteinklassifizierungen sind im Anhang in Tabelle 45 zusammengefasst. Die Fosmidkarte wurde mit dem Programm DNAPlotter visualisiert, die Spuren von Außen nach Innen zeigen ORFs in Vorwärts-Richtung (*coding strand, forward*), ORFs in Rückwärtsrichtung (*complementary strand, reverse*), den Sequenzbereich des pCC1FOS-Rückgrates sowie einen GC-Plot (GC-Gehalt über dem Durchschnitt: Ockerfarben, GC-Gehalt unter dem Durchschnitt: violett). Abbildung modifiziert nach Leis *et al.*, (2015b).

Der für die Aktivität verantwortliche ORF Nummer 23 umfasst 1.701 bp Länge und kodiert für ein 566 Aminosäuren großes Protein, im Folgenden Cel12E benannt. Das berechnete Molekulargewicht entspricht 62,3 kDa, der isoelektrische Punkt (pI) wurde auf 4,26 berechnet. Ein Signalpeptid mit 24 Aminosäuren Länge wurde vorhergesagt, welches lt. SignalP-Server wahrscheinlich zwischen Ala(24) und Gln(25) gespalten wird (*D-score* Wert von 0,72). *Blastp* gab als höchste Übereinstimmung eine Endo-1,4- $\beta$ -Glucanase b (Accessionnummer YP\_002581913.2) aus *Thermococcus sp.* AM4 mit 45 % Übereinstimmung auf Aminosäuresequenzebene (*e-value*  $2 \times e^{-82}$ ) aus. Die Analyse der Proteindomänen mittels Pfam wies eine N-terminale Glycosidhydrolase der Familie 12 mit zwei C-terminalen CBM2 Bindemodulen auf (schematische Repräsentation des Proteins in Abbildung 21).



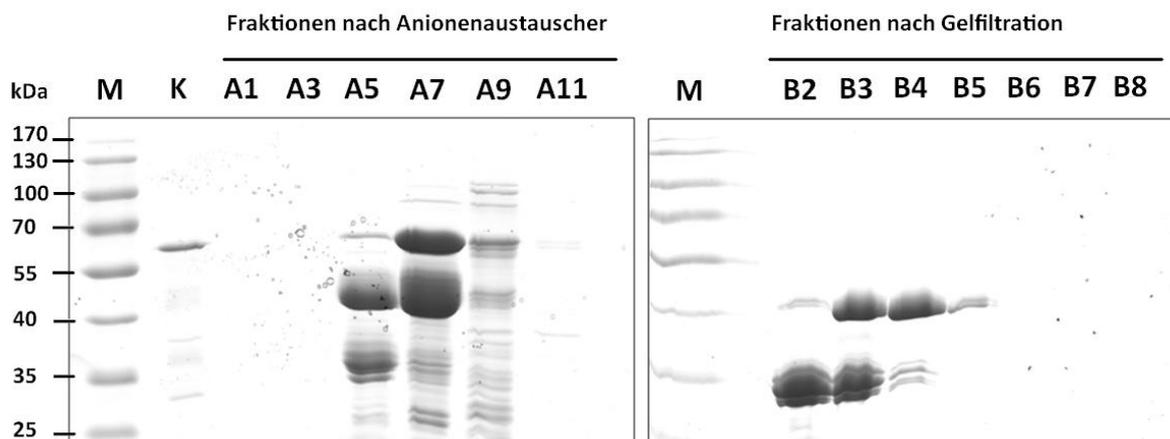
**Abbildung 21:** Vorhersage der Domänenzusammensetzung von Cel12E. Das Signalpeptid am N-terminus umfasst 24 Aminosäuren. Die enzymatisch aktive Glycosidhydrolase der Familie 12 umfasst die Aminosäurepositionen 120 bis 281 ( $e\text{-value } 3 \times e^{-32}$ ), die beiden Bindedomänen der CBM2-Familie liegen an Positionen 323 bis 392 ( $e\text{-value } 3,8 \times e^{-7}$ ) und 475 bis 557 ( $e\text{-value } 0,00035$ ) am C-terminalen Ende des Proteins. Abbildung modifiziert nach Leis *et al.*, (2015b).

Zu Expressionszwecken von Cel12E wurden zwei Genvarianten – jeweils mit und ohne Signalpeptid – amplifiziert und ohne Anfügen eines poly-His *tags* in das pET21a(+)-Expressionsplasmid über *NdeI* und *NotI* mittels isothermaler DNA-Assemblierung (Gibson Assembly) kloniert (Primerpaare SP-HACMC-F und HACMC-F mit HACMC-R). Die sequenzierten Konstrukte (pET21a-HACMC1 und pET21a-spHACMC1 mit Signalpeptid) wurden in den BL21(DE3)-Stamm übertragen und während der exponentiellen Wachstumsphase mit 1,0 mM IPTG induziert und nach drei Stunden geerntet. Die Ansätze wurden nach Gewinnung des Rohextraktes und des Kulturüberstandes auf cellulolytische Aktivität mit CMC in einer DNSA-Reaktion über 1 Stunde bei 96 °C im Eppendorf-Thermomixer getestet. Die Aktivität der Enzymvariante mit dem N-terminalen Signalpeptid betrug  $42,4 \pm 3,1$  % der Aktivität der abgeschnittenen Variante. Die Aktivität des Überstandes war bei beiden Varianten vernachlässigbar. Im Zuge der biochemischen Untersuchung wurde mit Cel12E ohne Signalpeptid weitergearbeitet.

### 2.3.2. Heterologe Expression und Aufreinigung

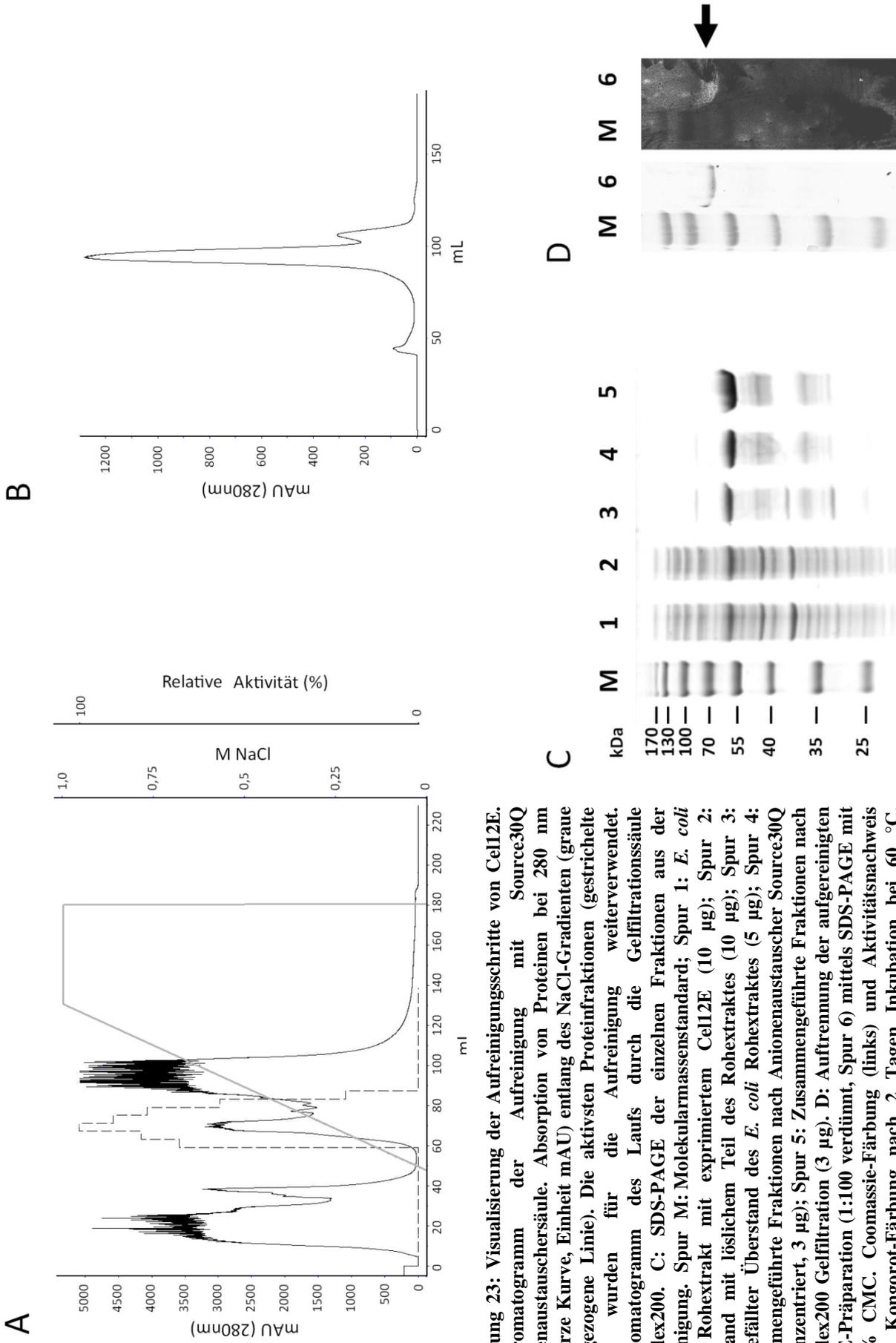
Zur Expression in BL21(DE3) im größeren Volumen wurden zwei 250 mL LB Medium in jeweils zwei 1-Liter Schüttelkolben mit Schikanen bei 37 °C und 180 Upm geschüttelt, ehe ab einer optischen Dichte ( $OD_{600\text{nm}}$ ) von 0,6 bis 0,7 mit 1,0 mM IPTG die Expression für 4 Stunden induziert wurde. Nach Ernte der Zellen wurden die Zellpellets in 50 mM Tris-HCl (pH 8,0) gewaschen und in einer Frenchpress Zelle aufgeschlossen. Der erste Aufreinigungsschritt umfasste eine Inkubation des Rohextraktes bei 80 °C für 20 Minuten im Wasserbad. Präzipitiertes Protein wurde abzentrifugiert und der Überstand mit dem Zielprotein für die nachfolgenden Aufreinigungsschritte über Nacht im Kühlraum bei 4 °C dialysiert. Die Proteinaufreinigung wurde über eine Source30Q Anionenaustauschersäule und Superdex 200 Gelfiltrationssäule weitergeführt, und gesammelte Fraktionen mittels DNSA Test mit CMC-Substrat getestet.

Bereits nach dem ersten Aufreinigungsschritt nach dem Anionenaustauscher konnte eine Fragmentierung von Cel12E beobachtet werden (Abbildung 22): Die einzelnen Fraktionen der Source30Q-Säule zeigen das Zielprotein mit 60 kDa Größe sowie dessen Abbauprodukte zwischen 55 und 40 kDa und 35 kDa. Nach Zusammenführen der Fraktionen mit intaktem Cel12E (Fraktionen A5 bis A9) war nach Gelfiltration kein Volllängen-Protein mehr nachweisbar. Die Fraktionierungen mit der Superdex200 Säule zeigen, dass die Degradation des Proteins definierte Fragmente lieferte. Selbst nach ausgiebigem Waschen der Pumpen, Schlauchsysteme und Säulen der FPLC-Anlage konnte nach einer weiteren Aufreinigung die zunehmende Fragmentierung von Cel12E nicht verhindert werden. Um die möglichen Faktoren zu bestimmen, die zum Abbau des Proteins führen, wurden NaCl-Konzentration zwischen 0 und 200 mM sowie die Präsenz von bestimmten Cellulose- und Dextrin-haltigen Materialien untersucht. Weder die Salzkonzentration, noch Avicel/Source30Q Säulenmaterial konnten für die beobachtete Fragmentierung von Cel12E verantwortlich gemacht werden. Erst durch die Zugabe von Protease-Inhibitor-Cocktailmix Tabletten (Roche) konnte eine überwiegende Fragmentierung gestoppt werden.



**Abbildung 22: Fragmentierung von Cel12E während der Proteinaufreinigung. Die Fraktionen nach Source30Q-Säule (linkes Gelbild) und Superdex200 (rechtes Bild) zeigen die zunehmende Zersetzung des Zielproteins in kleinere Fragmente. Spur M ist der Molekularmassen-Standard, Spur K ist die Kontrollprobe von Cel12E nach Hitzepräzipitation des Rohextraktes vor der Säulenaufreinigung.**

Durch die wiederholte Zugabe von Protease-Inhibitor konnte Cel12E über alle Reinigungsstufen hinweg stabilisiert werden. Wie aus der Gelkontrolle mittels SDS-PAGE (Abbildung 23) ersichtlich ist, konnte das 543 Aminosäuren-lange Zielprotein ohne Signalpeptid mit einer Größe von ca. 60,0 kDa aufgereinigt und aufkonzentriert werden. Leichte Degradationsbanden waren allerdings während allen Reinigungsschritten sichtbar und konnten nicht verhindert werden. Unter diesen Bedingungen war Cel12E auch über die Dauer der Enzymtests für mindestens mehrere Wochen bis Monate im Kühlschrank stabil.



**Abbildung 23: Visualisierung der Aufreinigungsschritte von Cell12E.**

**A:** Chromatogramm der Aufreinigungsschritte mit Source30Q Anionenaustauschersäule. Absorption von Proteinen bei 280 nm (schwarze Kurve, Einheit mAU) entlang des NaCl-Gradienten (gestrichelte graue Linie) wurden für die Aufreinigung weiterverwendet.

**B:** Chromatogramm des Laufs durch die Gelfiltrationssäule Superdex200. **C:** SDS-PAGE der einzelnen Fraktionen aus der Aufreinigung. Spur M: Molekularmassenstandard; Spur 1: *E. coli* BL21 Rohextrakt mit exprimiertem Cell12E (10 µg); Spur 2: Überstand mit löslichem Teil des Rohextraktes (10 µg); Spur 3: Hitzegefällter Überstand des *E. coli* Rohextraktes (5 µg); Spur 4: Zusammengeführte Fraktionen nach Anionenaustauscher Source30Q (aufkonzentriert, 3 µg); Spur 5: Zusammengeführte Fraktionen nach Superdex200 Gelfiltration (3 µg). **D:** Auftrennung der aufgereinigten Cell12E-Präparation (1:100 verdünnt, Spur 6) mittels SDS-PAGE mit 0,15 % CMC. Coomassie-Färbung (links) und Aktivitätsnachweis mittels Kongorot-Färbung nach 2 Tagen Inkubation bei 60 °C. Abbildung modifiziert nach Leis *et al.*, (2015b).

Die Zusammenfassung aller Aufreinigungsschritte (Aufreinigungstabelle) von Cel12E ist in Tabelle 25 wiedergegeben.

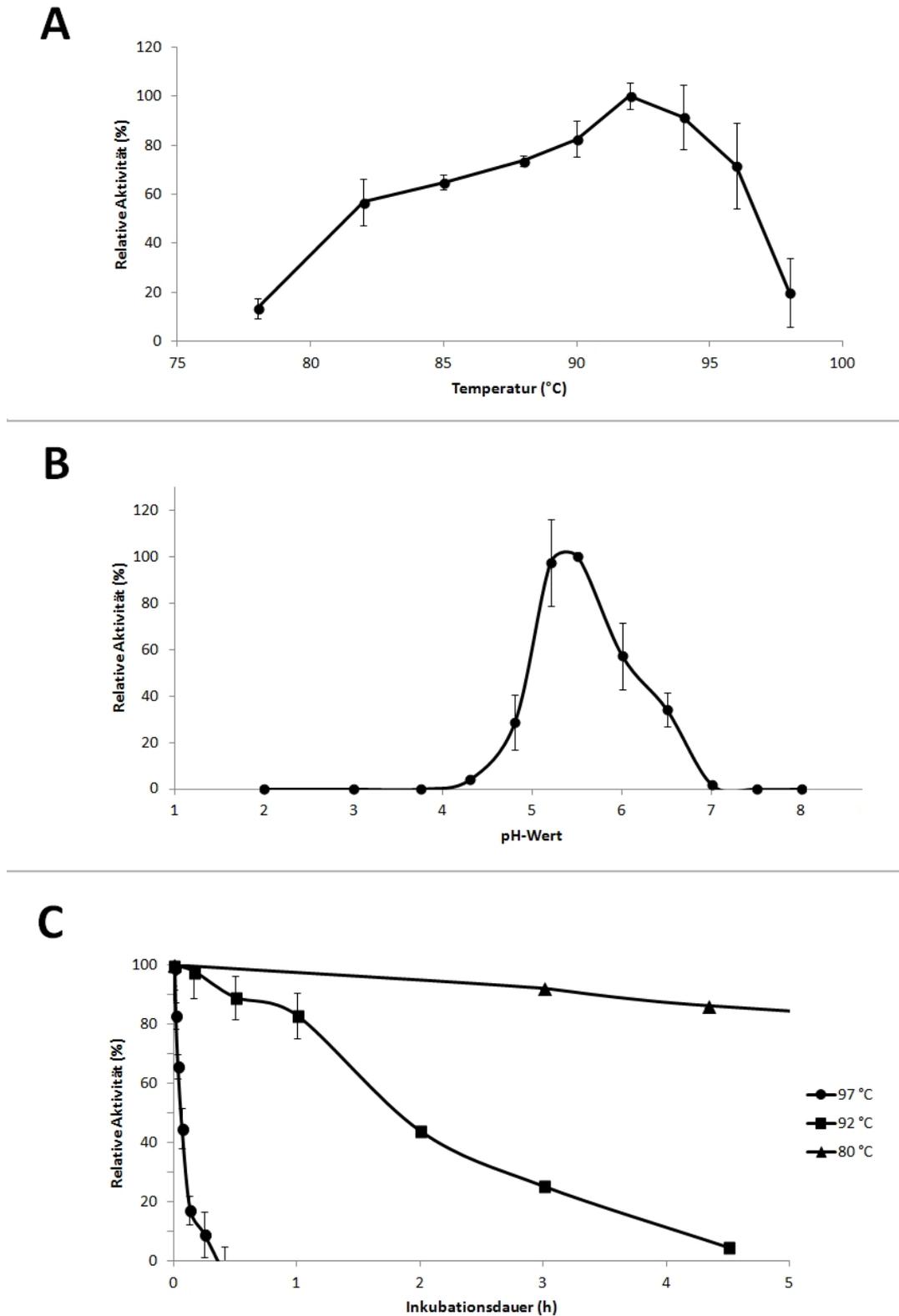
**Tabelle 25: Aufreinigungstabelle aus Präparation der heterolog exprimierten Glycosidhydrolase Cel12E.**

Fraktion der Aufreinigung	Volumen (mL)	Gesamtprotein (mg)	Gesamtaktivität (U) × 10 <sup>3</sup>	Spezifische Aktivität (U/mg)	Ausbeute (%)	Aufreinigungs-Faktor
Rohextrakt	27,5	361,1 ± 3,9	14,19 ± 2,62	39,3 ± 7,5	100 ± 18,6	1,0 ± 0,19
Hitzegefällter Rohextrakt	22,5	60,3 ± 3,7	9,78 ± 0,86	162,3 ± 9,5	68,9 ± 6,1	4,13 ± 0,24
Anionen-austauscher	0,4	17,7 ± 0,09	7,81 ± 0,55	440,2 ± 31,9	53,9 ± 6,0	11,2 ± 0,81
Gelfiltration	1,6	8,0 ± 1,14	5,60 ± 1,12	700,4 ± 75,8	39,6 ± 7,9	17,8 ± 1,93

Angegeben sind Mittelwerte ± Standardabweichung aus zwei Aufreinigungsprozessen.

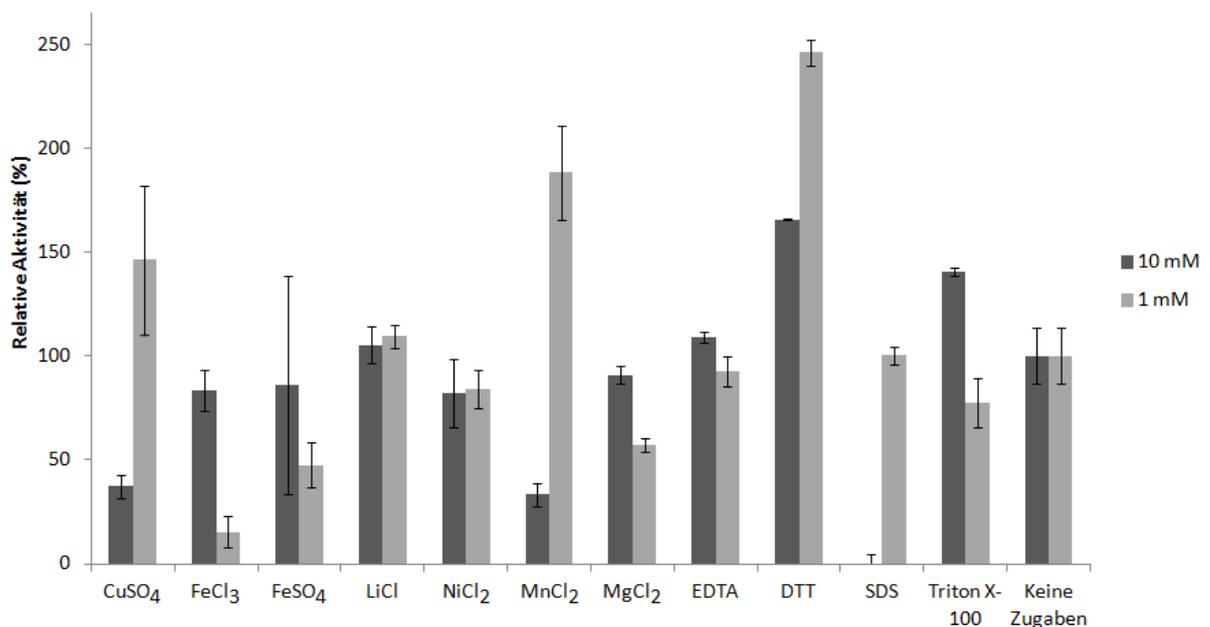
### 2.3.3. Biochemische Charakterisierung von Cel12E

Die archaeelle Glycosidhydrolase Cel12E wurde bezüglich biochemischer Eigenschaften charakterisiert. Das Aktivitätsprofil des thermostabilen Enzyms wurde bei Temperaturen zwischen 78 und 97 °C bestimmt, wobei die Inkubationszeiten bei allen Proben identisch waren (jeweils 10 Minuten). Das Enzym hat sein Temperaturoptimum im Bereich von 85 °C bis 96 °C, wobei die höchsten gemessenen Aktivitäten bei 92 °C ermittelt werden konnten (Abbildung 24A). Bei Temperaturen unter 78 °C und am Siedepunkt des Wassers (ca. 98 °C) fiel die Aktivität auf unter 20 % des Aktivitätsoptimums zurück. Der ermittelte optimale pH-Wert von Cel12E liegt bei etwa 5,5 bei 92 °C (Abbildung 24B). Zur Feststellung der thermischen Stabilität von Cel12E, wurden Ansätze des aufgereinigten Enzyms bei 80, 92 und 97 °C über mehrere Stunden vorinkubiert, bis die Auswertung der Restaktivität im DNSA-Test unter Idealbedingungen (siehe oben) durchgeführt wurde (Abbildung 24C).



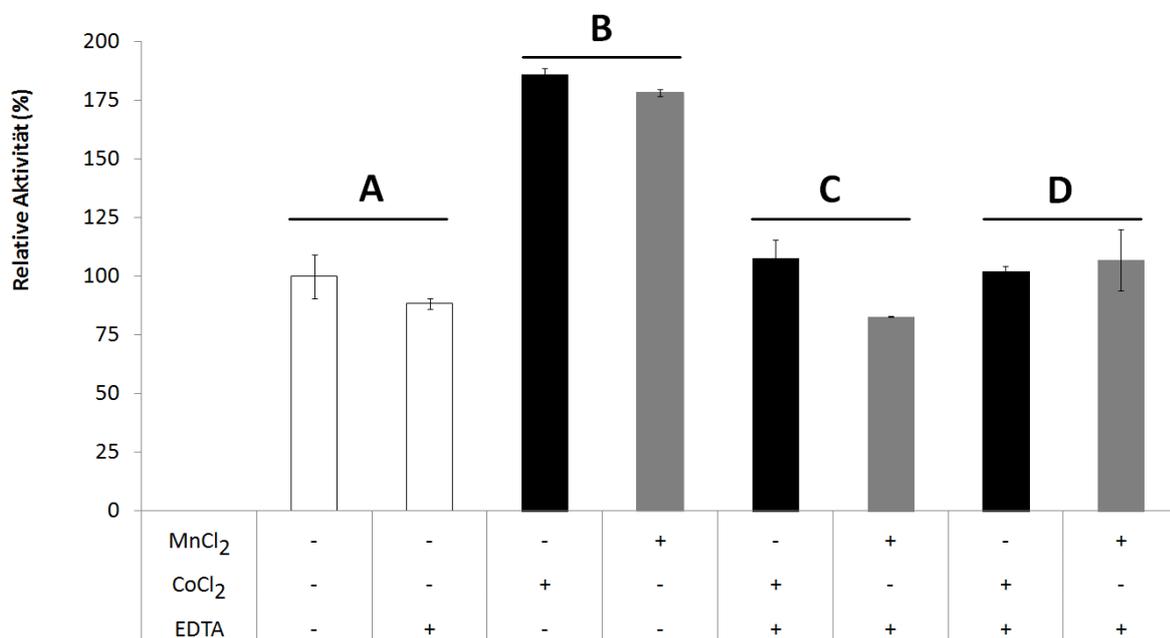
**Abbildung 24: Zusammenfassung der Aktivitätsprofile von Cel12E. A: Das Temperaturspektrum von Cel12E; B: Bestimmung des pH-Optimums des Enzyms mit verschiedenen Puffersystemen: Glycin-HCl-Puffer (pH-Bereich von 2-3), Natriumacetat-Puffer (pH 3,8-5,5) und Phosphatpuffer (pH 5,5-8), jeweils bei idealer Inkubationstemperatur getestet; C: Darstellung der thermischen Stabilität von Cel12E. Die Messpunkte geben die relativen Aktivitäten als Mittelwerte aus Duplikatmessungen  $\pm$  einfache Standardabweichung. Unter Idealbedingungen (92 °C, pH 5,5) entspricht die rel. Aktivität von Cel12E 100 %. Abbildung modifiziert nach Leis *et al.*, (2015b).**

Der Einfluss bestimmter Salze und Inhibitoren (im Bereich von 1,0 bis 10 mM der Zusätze) auf die Aktivität von Cel12E wurden getestet (Abbildung 25). Zweiwertige Schwermetallionen wie Kupfer und Mangan wirkten bei 10 mM Endkonzentration stark inhibierend, während geringere Konzentrationen zu einer teils signifikanten Zunahme der Aktivität beitrugen. Zugaben von Ethylendiamintetraessigsäure (EDTA) hatten unter diesen Bedingungen keinen Einfluss auf die Aktivität von Cel12E. Aktivitätssteigernd wirkten niedrige Konzentrationen an Dithiothreitol (DTT) und höhere Beigaben an Triton X-100. Zugabe von mehr als 1,0 mM Natriumdodecylsulfat (SDS) inaktivierte Cel12E vollständig. Der Einfluss von Natrium- und Kaliumchlorid bei Konzentrationen von 1 mM bis 10 mM war vernachlässigbar (nicht gezeigt). Die höchsten Aktivitäten konnten erst ab Salzkonzentrationen zwischen 200 mM (KCl) bis 400 mM (NaCl) pro Reaktion gemessen werden, wobei bei noch höheren Salzkonzentrationen keine signifikante Aktivitätssteigerung mehr zu beobachten war (saturierender Effekt). Die höchsten gemessenen relativen Aktivitäten waren  $188,2 \pm 1,9 \%$  (NaCl) und  $177,2 \pm 3,3 \%$  (KCl) bei einer Konzentration von je 500 mM der Chlorid-Ionen.



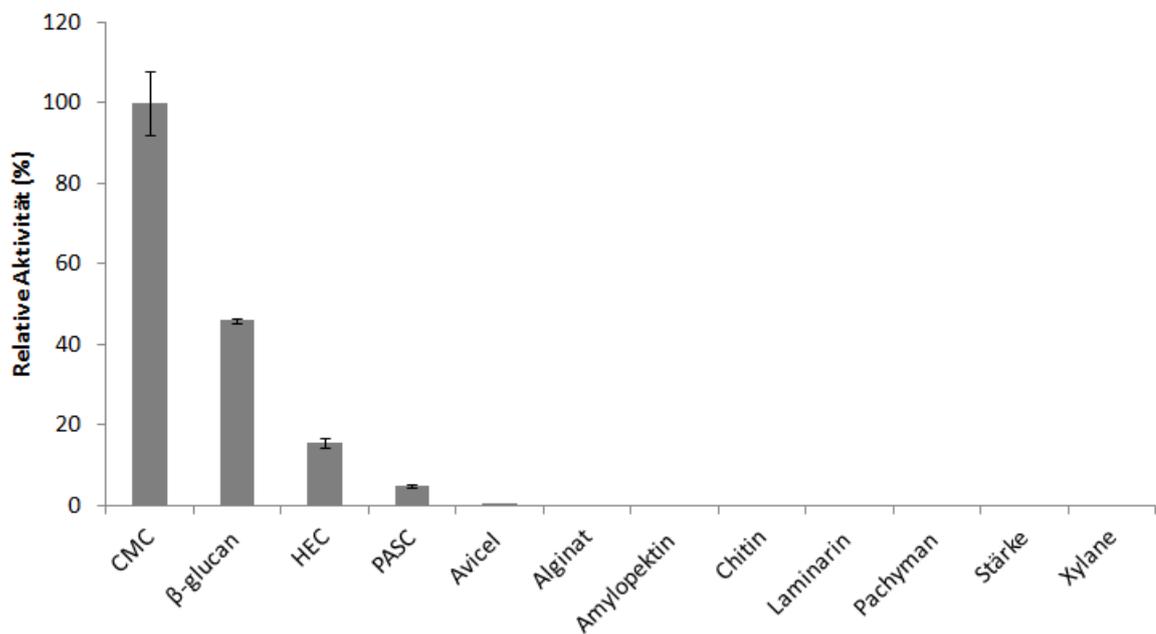
**Abbildung 25: Einfluss verschiedener Salze, Metallionen und ausgewählter Detergenzien auf die Aktivität von Cel12E.** Die Balkendiagramme zeigen Mittelwerte der relativen Aktivitäten aus Doppelbestimmungen, zugehörige Fehlerbalken sind als  $1 \times$  Standardabweichung repräsentiert. Zur Inkubation wurden  $0,05 \mu\text{g}$  Enzym eingesetzt, die Bedingungen waren  $92 \text{ }^\circ\text{C}$  in  $50 \text{ mM}$  MES Puffer (pH 5,5). Die rel. Aktivität (aus Duplikatmessungen, einfache Standardabweichung) von Cel12E ohne Additive entspricht  $100 \%$ .

Der positive Effekt zweiwertiger Metallionen auf die Aktivität von Cel12E wurde anhand von Supplementierungsversuchen mit Kobalt- und Manganchlorid untersucht. Die höchsten beobachteten Aktivitätssteigerungen waren bei 0,5 mM CoCl<sub>2</sub> (210,9 ± 5,1 % rel. Aktivität) und zwischen 1,0 und 2,0 mM MnCl<sub>2</sub> (188,5 ± 22,6 % rel. Aktivität) erzielt worden. Höhere Konzentrationen dieser beiden Schwermetallionen führten wieder zu einer Abnahme der Aktivität. Titrationsversuche mit Ethylendiamintetraessigsäure als Chelator zweiwertiger Ionen zeigten kaum negative Effekte auf die Aktivität von Cel12E alleine (Abbildung 26). Dennoch konnte gezeigt werden, dass der positive Effekt durch 1,0 mM Mangan oder Kobalt im Reaktionsansatz durch Komplexierung der Ionen mit Überschuss an EDTA (10 mM) selektiv und reversibel war. Die Reihenfolge der Supplementation von Ionen und Chelator spielte für die Aktivität des Enzyms dabei keine Rolle. Die beobachteten aktivierenden Effekte aus hohen Chloridionenkonzentrationen und bivalenten Metallionen waren additiv: Eine Kombination aus 0,3 M KCl sowie 0,4 M NaCl konnte die Aktivität um Faktor 1,56 steigern, eine weitere Zugabe von 0,2 mM Cobalt-Chlorid bewirkte eine 4,3-fache Steigerung gegenüber der Ausgangsaktivität ohne Supplementationen.



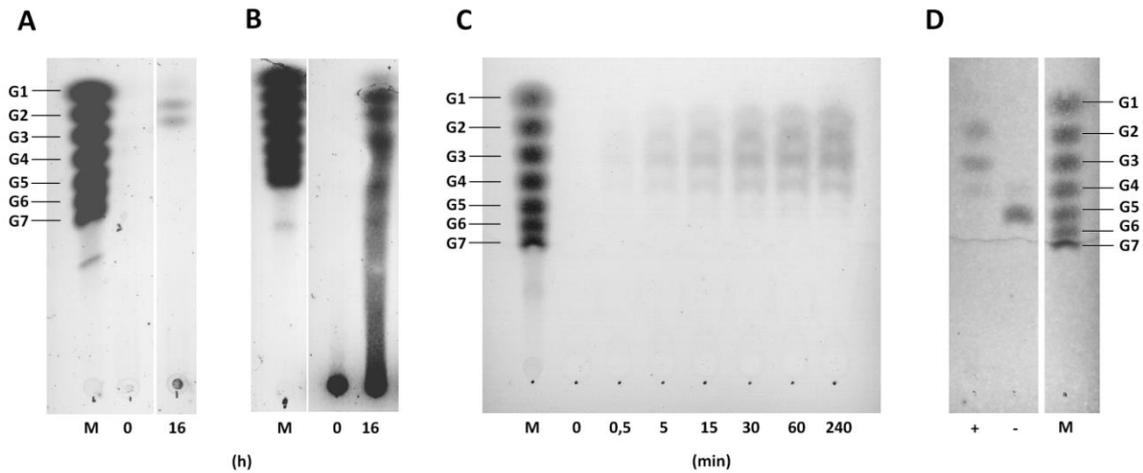
**Abbildung 26: Einfluss von zweiwertigen Kobalt- und Manganionen und EDTA auf die Aktivität von Cel12E.** Gezeigt ist Kobalt (schwarze Balken) und Mangan (graue Balken). Kontrollen (weiße Balken) ohne weitere Zugaben und mit Supplementationen von 10 mM EDTA (A). Einfluss nach Zugabe von 1 mM Kobalt- und Manganionen zu Cel12E (B) sowie nach zusätzlicher Inkubation mit 10 mM EDTA (C). Aktivität von Cel12E, welches zuerst mit EDTA inkubiert und anschließend mit den Ionen supplementiert wurde (D). Das Balkendiagramm zeigt Mittelwerte ± einfache Standardabweichungen (n = 2).

Das Substratspektrum von Cel12E wies eine Präferenz für lösliche und modifizierte sowie für amorphe und unlösliche Cellulose auf (Abbildung 27). Die höchste gemessene Aktivität lag bei  $692,3 \pm 55,7$  U für Carboxymethylcellulose (CMC), gefolgt von  $317,7 \pm 3,8$  U auf  $\beta$ -Glucan (aus Gerste),  $107,7 \pm 8,8$  U auf Hydroxyethylcellulose (HEC) und  $33,6 \pm 3,8$  U mit Phosphorsäure-gequollene Cellulose (PASC). Mikrokristalline Cellulose (Avicel) war nur in sehr geringen Mengen hydrolysiert worden ( $32 \pm 5,0$  mU). Die beobachtete Aktivität konnte selbst durch erhöhte Enzymzugaben nicht mehr gesteigert werden (die Menge des eingesetzten Enzyms ergab zwischen  $0,5 \mu\text{g}$  bis  $5,0 \mu\text{g}$  keinen signifikanten Unterschied in der Bildung reduzierender Zuckeren aus mikrokristalliner Cellulose).



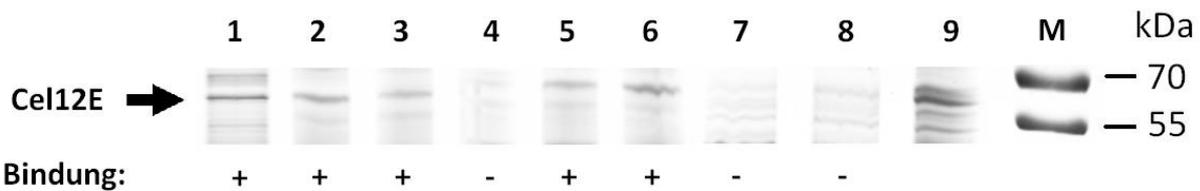
**Abbildung 27: Substratspektrum von Cel12E auf verschiedenen getesteten Substraten. Mittelwerte und einfache Standardabweichung aus Doppelbestimmungen, Daten wiedergegeben als relative Aktivität (auf Carboxymethylcellulose unter Idealbedingungen entsprechend 100 %).**

Die qualitative Untersuchung der Hydrolyseprodukte mittels Dünnschichtchromatographie ergab eine vorwiegende Bildung von Cellobiose und Cellotriose-Molekülen aus Avicel und PASC (Abbildung 28). In geringeren Mengen waren auch Glucose und Cellotetraose nachweisbar, welche sich im Laufe der Inkubation als Nebenprodukte akkumulierten. Zwischenprodukte wie höhermolekulare Zuckerverbindungen (z.B. Cellopentaose) waren nicht nachweisbar. Außerdem konnte gezeigt werden, dass Cel12E in der Lage ist, Cellopentaose als definiertes Substrat zu Cellobiose und Cellotriose abzubauen.



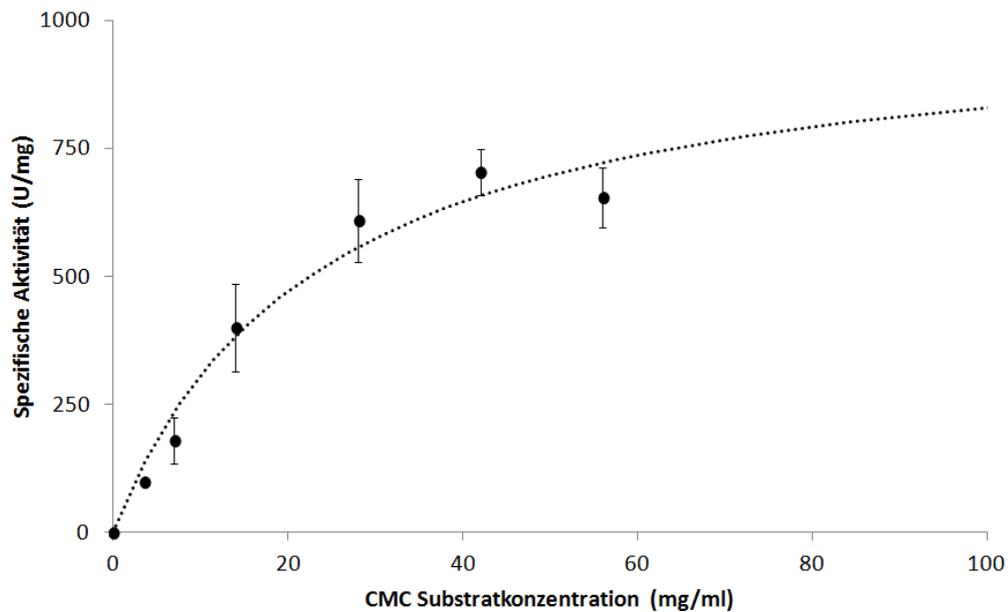
**Abbildung 28: Dünnschichtchromatographische Analyse von Hydrolyseprodukten aus Polysacchariden. Qualitative Produktanalyse von 0,7 % Avicel (A) und Carboxymethylcellulose (B) nach ÜN-Inkubation bei 70 °C mit 10 % (v/v) Hitze-behandelter Rohextrakte des Fosmid-tragenden Klons (HA-cmc-1) in 50 mM Tris-HCl pH 8,0. Abbauproduktanalyse von 0,05 µg/ml aufgereinigtem Enzym Cel12E unter optimalen Reaktionsbedingungen (50 mM MES pH 5,5, 92 °C) mit je 0,7 % PASC (C) und Cellopentaose (G5-Zucker) als definiertes Substrat (D) durch Enzymzugabe (+) und nach Inkubation ohne Cel12E (-). M: Maltooligosaccharid Zuckerstandard mit Monosaccharid (G1) bis Heptasaccharid (G7).**

Die Bindetests von Cel12E an verschiedenen Polysacchariden zeigten eindeutige Präferenzen für mikrokristalline und amorphe Cellulosen (Avicel, PASC) sowie für Chitin und  $\beta$ -verknüpfte Glucane (aus Gerste und Pilz), nicht jedoch für Chitosan und die getesteten Xylane (von Hafer und Lärche) (Abbildung 29).



**Abbildung 29: Bindevverhalten von Cel12E auf verschiedenen Polysacchariden. Rohextrakte (1,5 µg/µL Proteinkonzentration mit exprimiertem Cel12E) wurden mit 0,5 % (w/v) Substraten bei 6 °C ÜN unter ständigem Schütteln inkubiert. Das mit den Substraten pelletierte Protein wurde in 40 µL rekonstituiert, denaturiert und mittels SDS-PAGE visualisiert. Aufgetragen wurden Inkubationen mit  $\beta$ -Glucan der Gerste (1), Avicel (2), Chitin (3), Chitosan (4), Pachyman (5), PASC (6), Xylan vom Haferspelz (7), Xylan vom Lärchenholz (8) sowie die Cel12E-Kontrolle ohne Polysaccharid-Exposition (9), molekularer Größenstandard (M).**

Die Bestimmung der Michaelis-Menten-Kinetik von Cel12E fand unter idealen Reaktionsbedingungen des Enzyms statt. Zur Ableitung der kinetischen Parameter der maximal erreichbaren Reaktionsgeschwindigkeit ( $V_{\max}$ ) und Michaelis-Konstante ( $K_m$ ) wurden Carboxymethylcellulose-Konzentrationen zwischen 3,5 und 56 mg/mL verwendet. Der von der Substratkonzentration charakteristische hyperbelförmige Aktivitätsverlauf konnte mittels Lineweaver-Burk-Methode durch doppelt-reziproke Darstellung sehr gut linearisiert werden (lineare Regressionsgerade mit einem Korrelationskoeffizienten  $r^2 = 0,992$ ). Daraus ließen sich die kinetischen Parameter von Cel12E auf CMC berechnen mit  $V_{\max}$   $1.547,6 \pm 168,4$  U/mg und einem  $K_m$ -Wert von  $51,21 \pm 2,53$  mg/ml. Auf Basis der Gleichung der Michaelis-Menten-Kinetik konnte ebenfalls eine nicht-lineare Regressionskurve ermittelt werden (siehe Abbildung 30), aus der sich für  $V_{\max}$   $1.025,2$  U/mg und ein  $K_m$ -Wert von  $23,48$  mg/ml ergab ( $r^2 = 0,986$ ).



**Abbildung 30: Enzymkinetischer Verlauf der Aktivität von Cel12E mit steigender Substratkonzentration an CMC (schwarze Punkte). Auf Basis der Datenpunkte (Duplikatmessungen) konnte eine nicht-lineare Regressionskurve nach der Gleichung der Michaelis-Menten-Kinetik berechnet werden (gepunktete Linie). Die Enzymtests erfolgten bei optimalen Reaktionsbedingungen in 50 mM MES, pH 5,5 und 92 °C. Abgebildet sind Mittelwerte  $\pm 1 \times$  SD.**

### **3. Vergleichende Durchmusterung und Analyse von Fosmid-Genbanken in *E. coli* und *T. thermophilus* HB27**

Auf Basis des pCT3FK *Shuttle*-Fosmids (Angelov *et al.*, 2009) können diverse Genbanken mit metagenomischer DNA sowohl in *E. coli* als auch in *T. thermophilus* parallel und vergleichend durchmustert werden. Dazu müssen die bestehenden Banken aus *E. coli* in *T. thermophilus* übertragen werden, entweder als einzelne Übertragungen in einem Hochdurchsatzverfahren im 96-Deepwell Block, oder aber als Übertragung einer Mischung (Pool) von Fosmiden. Die zur Verfügung gestellten Banken umfassten eine kompatible Genbank des Genoms von *T. Brockianus* (T17). Die verwendete genomische DNA wurde von K. Sahn und G. Antranikian (TU Hamburg-Harburg) zur Verfügung gestellt und von A. Angelov (TU München) und S. Blank (TU Hamburg-Harburg) in *E. coli* hergestellt. Diese Fosmid-Genombank deckt mit 384 Einzelklonen das vermutlich zwischen 2,1 bis 2,4 Mb große Genom etwa 6,8-fach ab (Leis *et al.*, 2013). Eine weitere thermophile Metagenombank aus Kompostproben (M12) von Holzabfällen (Kompostierwerk Bioenergiezentrum GmbH, Göttingen) wurde von M. Mientus (TU München, unveröffentlichte Ergebnisse) bereitgestellt. Die 1.920 Klone umfassende Fosmidbank wurde bereits auf hemicellulolytische Aktivitäten hin untersucht, phylogenetische 16S rDNA-Daten zur Probe sind darüber hinaus verfügbar. Eine weitere Metagenombank aus heißen Quellen von den Azoren (Probenmaterial bereitgestellt von L. Gonçalves und H. Santos, Universität Lissabon, Portugal) wurde in dieser Arbeit in *E. coli* hergestellt, welche insgesamt über 14.300 Einzelklone aus verschiedenen Habitaten umfassen (Wasser, Biofilme, Sedimente). Die reine Sequenzinformation dieser Bank hat damit einen durchschnittlichen Umfang von 572 Mb, welche zur funktionellen Durchmusterung in beiden Expressionswirten zur Verfügung standen. Phylogenetische Daten sind durch die Sequenzierung der 16S rDNA-Sequenzen vorhanden (Kapitel IV.3.2.4, Tabelle 31).

#### **3.1. Analyse einer Genbank von *T. Brockianus***

##### **3.1.1. Vergleichende Durchmusterung in beiden Wirten**

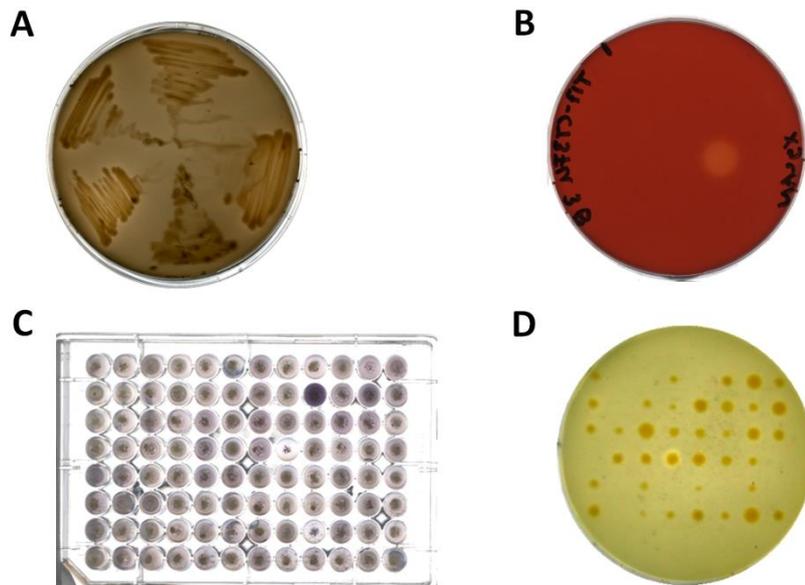
Zur funktionellen Durchmusterung der *T. Brockianus*-Genombank in beiden Wirten wurde die in *E. coli*-vorliegende Fosmidbank zunächst eins-zu-eins in *T. thermophilus* BL03 in Deepwell-Blöcken im Hochdurchsatzformat übertragen. Die Transformationseffizienz der Fosmide im 96 Well-Format

von dem thermophilen Wirt war hoch effizient, sodass von 384 verfügbaren Einzelfosmiden aus *E. coli* 371 Einzelklone in *T. thermophilus* generiert werden konnten. Nachdem die Klone auf Selektivagar angewachsen waren, wurden die Zellen für zwei Tage bei 60 °C in 1,1 mL Deepwell-Blöcken inkubiert. Schließlich konnten die Aktivitätsnachweise auf Agar-Festmedien mit Substraten und unter der Verwendung von lysierten Zellpellets in Mikrotiterplatten durchgeführt werden (beispielhafte Darstellung positiver Fosmidklone in Abbildung 31). Die Zusammenfassung aller erhaltenen Ergebnisse ist in Tabelle 26 ersichtlich. Da *T. thermophilus* HB27 verschiedene Proteasen/Peptidasen (Henne *et al.*, 2004) sowie Esterasen und lipolytische Enzyme (Fuciños *et al.*, 2005a; Levisson *et al.*, 2009) zu produzieren und sekretieren vermag, war der Nachweis der Substratverwertung durch hohe Hintergrundaktivitäten allgemein erschwert. Dennoch konnten mehr aktive Klone in dem thermophilen Bakterium identifiziert werden. In beiden Wirten konnten die zwei identischen Xylanase-aktiven Fosmid-Klone identifiziert werden. Nach analytischem Restriktionsverdau lieferten beide positiven Klone überwiegend überlappende Fragmentmuster (Daten nicht gezeigt). Alle weiteren Aktivitäten waren nur in *T. thermophilus* beobachtet worden, je zwei positive Klone mit AZCL-Galactomannan und auf Magermilchplatten, sowie ein detektierbarer Tributyrase-Klon. Cellulolytische Aktivitäten konnten trotz verschiedenster eingesetzter Substrate (Avicel, CMC, HEC) und verschiedener funktioneller Nachweismethoden in beiden Wirten nicht nachgewiesen werden.

**Tabelle 26: Auswertung der vergleichenden Durchmusterungsergebnisse in *E. coli* und *T. thermophilus* HB27.**

Aktivität	Substrat (Konzentration)	Aktive Klone (2 Tage Inkubation bei 60 °C)	
		<i>E. coli</i> EPI300	<i>T. thermophilus</i> HB27
Cellulase	Substratagarplatte mit 0,1 % (w/v) Avicel PH-101	-	-
	Mikrotiterplatte mit 0,05 % (w/v) AZO-Avicel	-	-
	Substratplatte mit 0,1 % (w/v) Carboxymethylcellulose	-	-
Esterase	Substratagarplatte mit 1,0 % (v/v) Tributyrin	-	1-D4
Protease	Substratplatten mit 2,0 % (w/v) Magermilch	-	1-B2 / 1-B10
Mannanase	Mikrotiterplattenformat mit 0,05 % (w/v) AZCL-Galactomannan	-	1-B9 / 2-D5
Xylanase	Substratagarplatten mit 0,1 % (w/v) Xylan (Hafer) im Mikrotiterplattenformat	3-E11 / 4-F4	3-E11 / 4-F4
Zahl identifizierter, aktiver Klone:		2	7
Zahl durchmusterter Klone (%):		384 (100%)	371 (96,6 %)

Zur Analyse der lipolytischen Fosmidklons T17-1-D4 in *T. thermophilus* BL03 wurde der Rohextrakt einer 5 mL über Nacht Kultur gewonnen, sein Proteingehalt quantifiziert und mit *para*-Nitrophenyl-Substraten (*p*NP-butyrat, -laurat und -palmitat) getestet. Der positive Fosmidklon zeigte eine nur leichte, obgleich messbare Zunahme von  $7,9 \pm 2,9$  mU pro mg eingesetzten Rohextrakt für *p*NP-C<sub>12</sub> gegenüber der Kontrolle mit pCT3FK ohne *Insert* im Expressionswirt.



**Abbildung 31:** Beispielhafte Zusammenstellung ausgewählter Fosmidklone, die in *T. thermophilus* auf Substrat- und Mikrotiterplatten Aktivitäten zeigten. **A:** Magermilch-Substratplatte mit vereinzelt Kandidatenklonen und Hofbildung bei erhöhter Protease-Aktivität. Stärkste Aktivität ging von Klon 1-B2 (auf Platte links unten) und 1-B10 (rechts) aus, die Wildtyp-Kontrolle (unterer Ausstrich) zeigte keinerlei Hofbildung. **B:** Hofbildung Xylanase-aktiver Klon 3-E11, welche nach Kongorotfärbung detektierbar ist. **C:** Mannanase-Aktivität nach Durchmusterung in Mikrotiterplatte. Überstände des sonst unlöslichen AZCL-Galactomannan färbten sich dunkelblau (siehe Klon 1-B9). **D:** Sichtbare Hydrolyse durch Hofbildung aufgrund verstärkter lipolytischer Aktivität durch das Fosmid 1-D4.

### 3.1.2. Bioinformatische Analyse

Nachdem redundante Fosmidklone durch analytische Restriktionsverdau ausgeschlossen werden konnten, wurden mehrere Kandidatenfosmide (Tabelle 26) mittels 454-NGS am Genomlabor in Göttingen (G<sub>2</sub>L) sequenziert (Projekt ANG48, A. Thürmer). Aus den fünf sequenzierten Fosmiden der *T. Brockianus*-Bank, Klon 1-B2 (Magermilch), 1-B9 (Galactomannan), 1-D4 (Tributyryn), 2-D5 (Galactomannan) und 3-E11 (Xylan) konnten lediglich zwei Fosmide auf größere *contigs* zugeordnet werden (Tabelle 27). Die vollständige Sequenz von Fosmid T17-1-D4 konnte auf 39,9 kb Länge (*contig* 3) assembliert und gemappt werden. ORF Nummer 26 kodiert für eine 237 Aminosäuren umfassende  $\alpha/\beta$ -Hydrolase der Familie 5, welche laut *blastp* Ähnlichkeiten zu *T. thermophilus* Stämmen HB27 und JL-18 zeigen (je 83 % Sequenzübereinstimmung zu YP\_005244.1 und YP\_006058102.1 mit identischem *e-value* =  $3 \times e^{-129}$ ). Das nächste bereits kristallisierte Homolog in

*T. thermophilus* HB8 (ORF Tt1662, PDB Accession-Nummer 1UFO\_A) zeigte ähnlich hohe Übereinstimmungen. Das andere, jedoch nicht vollständig sequenzierte Fosmid des Mannanase-aktiven Klons T17-1-B9 konnte zu zwei *contig*-Sequenzen zugeordnet werden. *contig* 11 war das größte assemblierte Fragment mit 27.035 bp ohne vorhersagbare Kandidatengene. Das wesentlich kleinere *contig* 145 trug einen partiellen ORF, der für einen 203 Aminosäuren-langen N-terminalen Proteinteil kodierte. Das vorhergesagte Protein hat höchste Ähnlichkeit zu einer Glycosid Hydrolase Familie 57 (GHF 57) aus *Thermus igniterrae* (88 % zu Protein WP\_018110714.1,  $e\text{-value} = 2 \times e^{-120}$ ) und hatte ebenso hohe Identität mit einem annotierten hypothetischen Protein TT\_C1540 aus *T. thermophilus* HB27 (YP\_005509.1).

**Tabelle 27: Sequenzdaten und Kandidatengene aus *T. Brockianus* Genombank**

Fosmid	Aktivität (Temperatur)	Assembliertes <i>contig</i> (Größe)	Vorhergesagte Kandidatengene (höchste Übereinstimmung lt. <i>blastp</i> )
T17-1-D4	Tributylin (60 °C)	<i>contig</i> 3 (39,9 kb)	$\alpha/\beta$ -Hydrolase Familie 5 Hypothetisches Protein TT_C1275 ( <i>T. thermophilus</i> HB27) Esterase ( <i>Thermus thermophilus</i> JL-18)
T17-1-B9	AZCL- Galactomannan (60 °C)	<i>contig</i> 11 und <i>contig</i> 145 (27,0 + 0,7 kb)	Partieller ORF, GHF 57 1,4- $\alpha$ -Glucan verzweigendes Protein ( <i>Thermus igniterrae</i> ) Hypothetisches Protein TT_C1540

### 3.2. Konstruktion einer metagenomischen Genbank aus heißen Quellen in *E. coli*

#### 3.2.1. Analyse des Probenmaterials

Probenmaterial von heißen Quellen (Azoren, Portugal) wurden freundlicherweise von L. Gonçalves und H. Santos (Universität Lissabon, Portugal) zur Verfügung gestellt (Tabelle 28). Die Umweltproben stammen von einer natürlich erhitzten Quelle, in der Mais von einheimischen Dorfbewohnern vor Ort gekocht wurde. Es wurden eine siedend heiße Wasserprobe aus dem Inneren dieser Quelle (Probenbezeichnung AZ 1), sowie zwei weitere Proben aus dem Sickerwasser (AZ 2 und AZ 4) entnommen. Eine weitere Probe wurde aus einer anderen heißen Quelle entnommen, welche mit

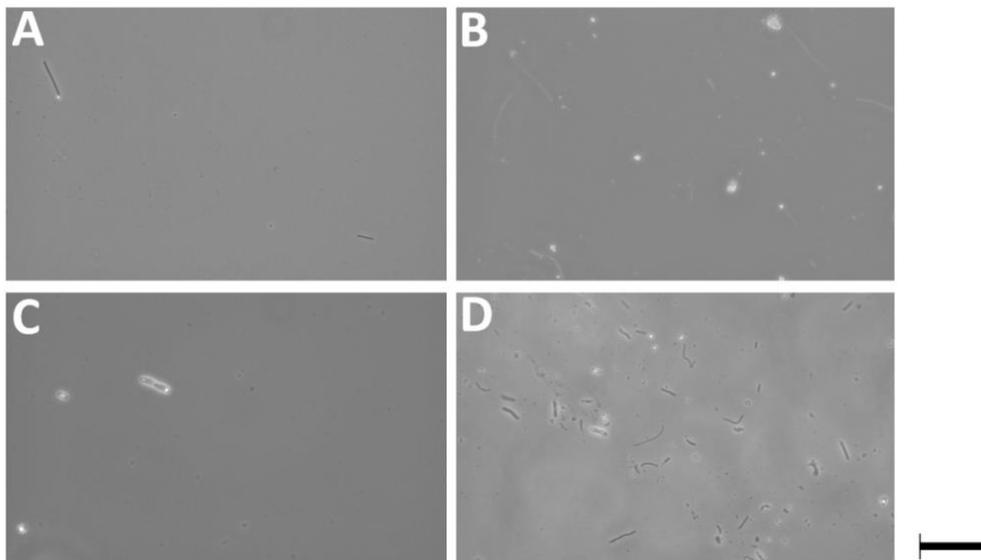
organischem Material (z.B. abgefallene Blätter und Zweige) angereichert war (Probe AZ 3). Das gesammelte Material umfasste drei Schott-Flaschen mit je etwa 400 mL Gesamtvolumen an Wasser und Sedimenten, sowie ein weiteres Gefäß mit Biofilmproben in einem Umfang von etwa 15 mL.

**Tabelle 28: Eigenschaften der Proben von heißen Quellen (Azoren)\*.**

Probenbezeichnung	Probenmaterial	pH	Temperatur
AZ1	Wasserprobe	7,0	95 °C
AZ2	Wasser und Sediment	6,0	60 °C
AZ3	Wasser und Sediment	2,0	62 °C
AZ4	Biofilmprobe	7,0	60 °C

\*: Lt. Angaben von L. Gonçalves (Universität Lissabon).

Die Präsenz von Mikroorganismen in den zum Teil sehr extremen Habitaten konnte mittels Mikroskopie gezeigt werden (Abbildung 32). Neben länglichen und gedrunenen Stäbchen waren stark lichtbrechende Sporen erkennbar, welche in Anzahl und Auftreten von Probe zu Probe unterschiedlich waren. In der Probe aus siedendem Quellwasser (AZ1) waren nur sehr wenige Stäbchen zu sehen, während aus der Biofilmprobe (AZ4) verschiedene Zellmorphologien zu erkennen waren.



**Abbildung 32: Mikroskopie-Aufnahmen der Proben von heißen Quellen. Gezeigt sind die Wasserproben AZ1 (A), AZ2 (B) und AZ3 (C) sowie eine Biofilmprobe AZ4 (D) in 1000-facher Vergrößerung. Der schwarze Balken entspricht einer Länge von 200 µm (Axio Scope Lichtmikroskopie mit AxioVision Software, Carl Zeiss AG).**

### 3.2.2. Gewinnung von metagenomischer DNA

Die Oberflächen von Blättern und Zweigen wurden abgeschabt, um etwaige anhaftende Zellen physikalisch zu entfernen. Durch Filtration der Wasserproben waren die Zellen auf den Filtermembranen gebunden. Diese wurden in der Lyse der Zellen mit den o.g. Methoden eingesetzt, und die Isolierung und Aufreinigung der Nukleinsäuren führte zu hochmolekularer, metagenomischer DNA. Je nach angewandter Methode und Umweltprobe variierten die Ausbeuten (Tabelle 29) und Qualitäten (Abbildung 33) der gewonnenen Nukleinsäuren (die Daten aus der Tabelle und Abbildung stammen jeweils von M. Heuchel, Praktikumsreport, 2011). Von den vier verschiedenen Umweltproben konnte hochmolekulare metagenomische DNA vor allem von Probe AZ3 (Sediment und Wasser) mit verschiedenen Kits zur DNA-Extraktion (PowerMax Soil, Meta-G-nome und UltraClean Water) isoliert werden. Stark fragmentierte Nukleinsäuren waren vor allem bei der Biofilmprobe beobachtet worden, oder wenn die DNA mit physikalischen Methoden (*bead beating*) behandelt worden waren. Bei der Verwendung von SoilMaster DNA Extraction Kit war keine DNA mehr im Agarosegel nachweisbar (Daten nicht gezeigt), obwohl bei der DNA Quantifizierung erhebliche Mengen gemessen werden konnten (200 ng/g Probenmaterial). Die qualitativ besten Ergebnisse lieferte der Meta-G-nome Kit mit einer Kombination enzymatischer und chemischer Lyse für die Sedimentproben mit sauberer, hochmolekularer genomischer DNA von über 20 kb Größe. Die beste mengenmäßige Ausbeute mit leicht fragmentierter DNA (sichtbar als Schmier im Gelbild) lieferte der UltraClean Water DNA Isolation Kit. Der PowerMax Soil Kit zeigte zwar die höchsten Ausbeuten (mehr als 1 µg pro g Probe), jedoch waren Nukleinsäuren hochgradig fragmentiert, mit wenig hochmolekularem Anteil. Geringe Ausbeuten waren von PowerSoil und UltraClean Fecal DNA Isolation Kit beobachtet worden. Generell war die Gewinnung reiner hochmolekularer DNA aus Biofilmen mit den angewendeten Methoden nur schwer möglich. Selbst ein Aufbrechen der für die Extraktion nachteiligen Exopolysaccharide (EPS) durch Flüssigstickstoff (N<sub>2</sub>) erbrachte keine signifikante Steigerung der Ausbeuten. Aus Probe AZ1 konnten trotz Anwendung verschiedener Extraktionsmethoden keine detektierbaren Mengen DNA in der Gelelektrophorese oder per Nanodrop-Quantifizierung festgestellt werden. Für die Wasserprobe von AZ2 waren nach Endreparatur der DNA für die Erstellung der pCT3FK Fosmidbank ebenfalls keine nachweisbaren Mengen mehr vorhanden. Um die Ausbeute der metagenomischen Nukleinsäuren zu erhöhen, konnte die mit den verschiedenen Isolationskits extrahierbare DNA für jede Probe einzeln zusammengeführt und weiterverarbeitet werden. Im Falle von gescherter DNA konnten per Gelelektrophorese niedermolekulare Fragmente unter 10 kb abgetrennt werden. Insbesondere bei der Biofilmprobe mussten größere Fragmente aus dem Gel ausgeschnitten werden, ehe sie zur Herstellung einer *large-insert* Fosmidbank in Frage kamen.

Tabelle 29: Ausbeuten der isolierten metagenomischen Nucleinsäuren aus heißen Quellen.

Konsistenz der Probe	Verwendetes Kit	Ausbeuten metagenomischer DNA pro g oder mL Umweltprobe			
		AZ1	AZ2	AZ3	AZ4
flüssig (Wasser)	UltraClean Water DNA Isolation Kit (MO BIO)	< 0,1 ng/mL	~ 2,5 ng/mL	~ 13,5ng/g	-
	Meta-G-nome DNA Isolation Kit (Epicentre)	-	~ 0,75 ng/mL	~ 7,5 ng/mL	-
fest (Sedimente und Biofilme)	SoilMaster DNA Extraction Kit (Epicentre)	-	< 0,2 µg/g	< 0,2 µg/g	< 0,2 µg/g
	Power Soil und PowerMax Soil DNA Isolation Kit (MO BIO)	-	~ 1,5 µg/g	~ 1 µg/g	~ 1 µg/g
	UltraClean Fecal DNA Isolation Kit (MO BIO)	-	~ 360 ng/g	~ 200 ng/g	~ 0,5 µg/g

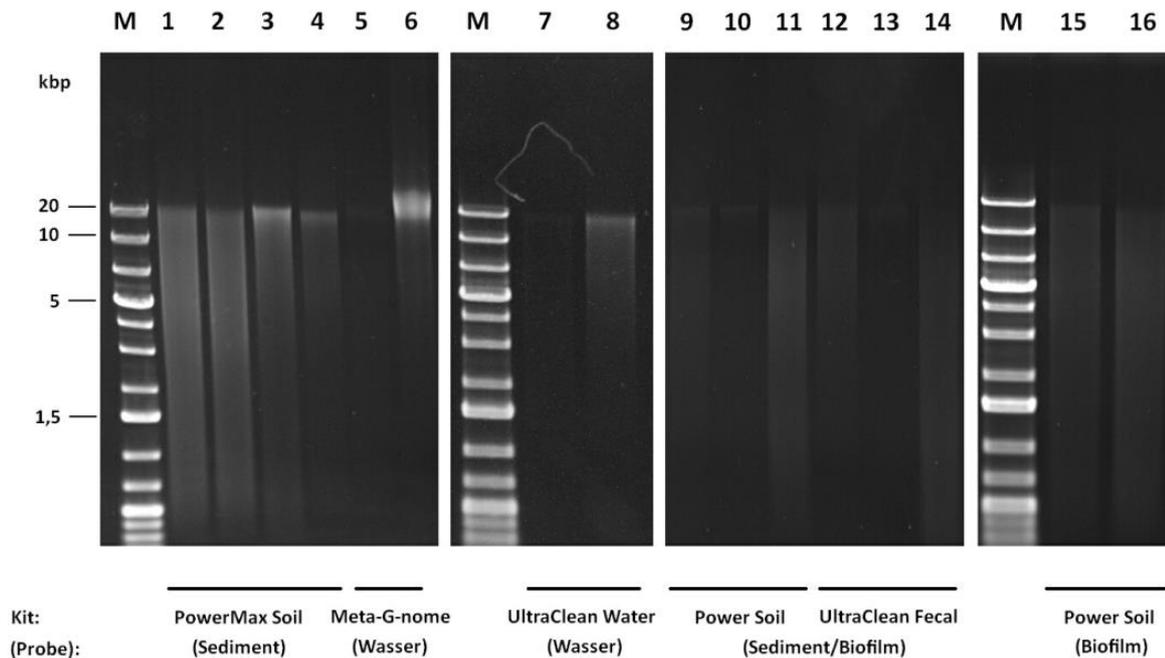
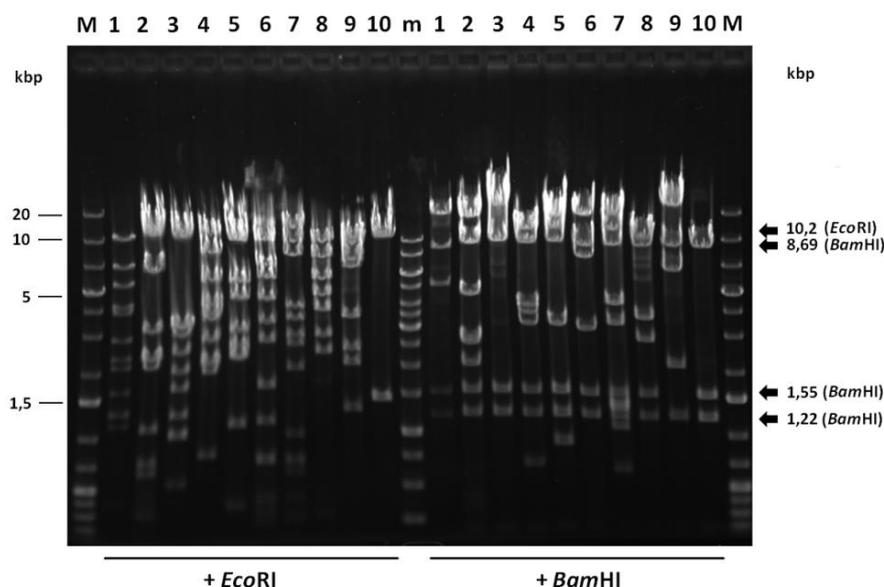


Abbildung 33: Qualität der gewonnenen metagenomischen DNA im Agarosegel (0,8% w/v). Analysiert wurden aufgereinigte Nucleinsäuren aus Sediment-Proben AZ2 (Spuren 1 und 2) und AZ3 (Spuren 3 und 4) jeweils mit (2µL) und ohne (5 µL) Aufkonzentrierung mit Glycogen (PowerMax Soil Kit). Wasserproben von AZ2 und AZ3 wurden mit dem Meta-G-nome Kit (Spuren 5 und 6, je 5 µL pro Spur) sowie mit dem UltraClean Water Kit (Spuren 7 und 8, je 3 µL geladen) aufgereinigt. Sedimente und Biofilme aus Proben AZ2, AZ3 und AZ4 wurden ebenfalls mittels Power Soil Kit (Spuren 9, 10 und 11) und UltraClean Fecal Kit (12, 13 und 14) isoliert (es wurden je 12 µL geladen), sowie die Biofilmprobe AZ4 mit und ohne Flüssigstickstoffbehandlung (Spuren 15 und 16 mit 5 µL beladen). Größenstandard (M): GeneRuler 1 kb Plus (Fermentas).

### 3.2.3. Erstellung der pCT3FK-Fosmidbank

Die aufgereinigte metagenomische DNA der Umweltproben hatte eine durchschnittliche Größe von mindestens 10 bis 20 kb, welche in der Ligationsreaktion mit pCT3FK Fosmidvektor verbunden werden konnten. Der Vektor war zuvor mit *Eco*72I (*Pml*I) linearisiert und die Enden dephosphoryliert worden. Das molare Verhältnis *Insert*:Vektor in der Ligation war 1:5, wobei circa 500 ng *Insert* mit entsprechend 1,7 µg Fosmid über Nacht bei 16 °C ligiert wurden. Nach der Ligation wurde der aufgereinigte Ansatz mit 25 µL MaxPlax Lambda Packaging Extract vollständig gemischt und nach zwei Stunden Inkubation bei 30 °C mit dem restlichen Extrakt versetzt. Nach weiteren zwei Stunden wurden die *E. coli* EPI300-Zellen mit den Phagen infiziert und die maximale Anzahl transduzierbarer Klone auf LB mit Chloramphenicol (12,5 µg/mL) bestimmt.

Aus der Wasserprobe AZ3 wurden im ersten Schritt etwa 10.000 Kolonien gewonnen, nach einer zweiten Klonierungsreaktion waren es weitere 6.000. Für die Sedimentprobe AZ3 konnten tausende Einzelfosmid-Klone generiert werden (> 10.000). Für die Proben mit erhöhten Anteilen an fragmentierter DNA (hoher Scherungsgrad bei Probe 2 Sediment und Probe 4 Biofilme) konnte nur ein Bruchteil der Ausbeute erreicht werden (insgesamt 672 Kolonien). Von den erhaltenen Fosmidklonen wurde die Diversität der Fosmid-*inserts* mittels *Bam*HI und *Eco*RI-Verdau bestimmt, um Redundanzen und religiertes pCT3FK-Fosmid ausschließen zu können (Abbildung 34).



**Abbildung 34:** Analyse der Diversität der Metagenombank, beispielhaft gezeigt an der Wasserprobe AZ3. Verdauemuster von neun zufällig ausgewählten Fosmidklonen (Spalten 1-9) mittels *Eco*RI (links) und *Bam*HI (rechts). Als Kontrolle diente das pCT3FK-Leerfosmid (Spur 10). Die für das pCT3FK-Fosmid konstanten Restriktionsfragmentlängen sind als Pfeile markiert. Fosmide 1 bis 9 entsprechen unterschiedlichen Verdauemustern mit variabler metagenomischer *Insert*-Sequenz. Verwendete Größenstandards: GeneRuler 1 kb Ladder (m) und 1 kb Plus Ladder (M) (aus M. Heuchel, Praktikumsreport, 2011).

Alle der getesteten Fosmidklone ( $n = 9$ ) zeigten unterschiedliche Restriktionsfragmente der metagenomischen *Inserts*. Da die Diversität der Genbanken damit ausreichend hoch war, wurden von den Proben AZ2 und AZ4 alle gewachsenen Kolonien, und von AZ3 Wasser und Sediment 13.824 Einzelklone gepickt und in 96-Well Mikrotiterplatten angeordnet. Die gesamte Bank aus heißen Quellen von den Azoren umfasste damit 151 Mikrotiterplatten, welche bei  $-80\text{ °C}$  gelagert wurden (Tabelle 30).

**Tabelle 30: Generierte einzelne Fosmidklone aus Metagenomprobe (Azoren).**

	AZ1	AZ2		AZ3		AZ4
	Wasser	Wasser	Sediment	Wasser	Sediment	Biofilm
Überführte <i>E. coli</i> EPI300 Klone	0	0	480	8.352	5.472	192
Entsprechend Mikrotiter- platten (je 96 Kavitäten)	0	0	5	87	57	2

### 3.2.4. Phylogenetische Analyse mittels 16S rDNA-Sequenzen

Die archaellen und bakteriellen Konsortien aus den Umweltproben der heißen Quellen der Azoren wurden auf Basis der 16S ribosomalen DNA-Sequenzen hin untersucht. Hierfür boten sich die Azorenprobe Nummer 3 an, da sie mitunter die meisten Fosmidklone der Metagenombank umfasste. Nach erfolgreicher PCR-Amplifikation mit den teils degenerierten 16S-Primern konnte je eine 96 Einzelkolonien umfassende Genbank des Sediments und der filtrierten Wasserproben erstellt werden. Hiervon wurden zunächst je 5 Einzelklone der archaellen und bakteriellen 16S rDNA vollständig sequenziert und mittels *blastn* nach den nächsten Vertretern analysiert. Für die archaellen 16 rDNA-Sequenzen konnte ausschließlich höchste Ähnlichkeit zu *Stygiolobus azoricus* DSM 6296 mit etwa 99 % Übereinstimmung aufgefunden werden. Die bakteriellen 16S-Sequenzen wiesen in drei Fällen Homologien gegenüber dem pathogenem *Escherichia fergusonii* ATCC 35469 auf, in einem Fall konnte höchste Übereinstimmung gegenüber *Methylobacterium radiotolerans* JCM 2831 nachgewiesen werden. Die Analyse mit partiellen Sequenzierungen von 32 weiteren bakteriellen 16S-Klonen mit Hilfe der RDP-II Datenbank (Ribosomal Database Project, Cole *et al.*, 2003) lieferte 19 weitere Zuordnungen, die in Tabelle 31 zusammengefasst sind.

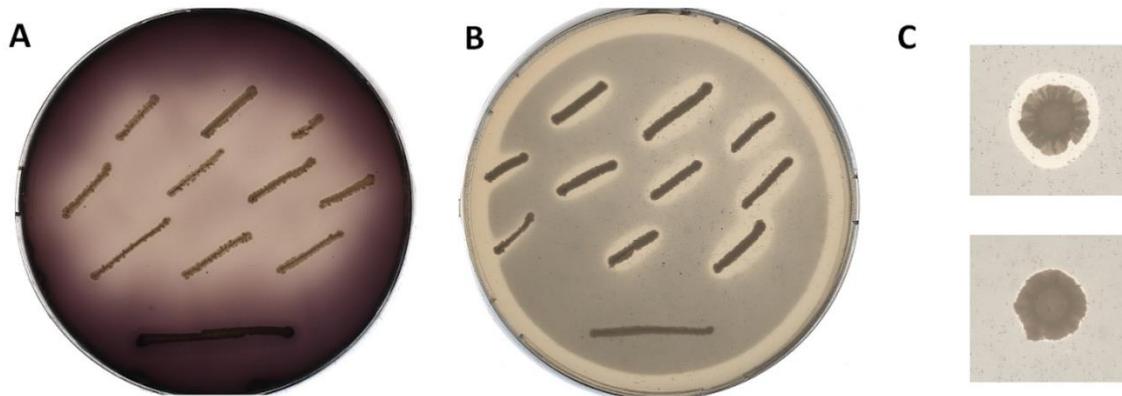
**Tabelle 31: Phylogenetische 16S rRNA-Analyse der Azorenprobe.**

Phylum	Klasse	Ordnung	Familie/Gattung	Anzahl der Sequenzen*
Crenarchaeota	Thermoprotei	Sulfolobales	Sulfolobaceae/ <i>Stygiolobus</i>	5
Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae/ <i>Ulvibacter</i>	2
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae/ <i>Methylobacterium</i>	4
		Rhodospirillales	Acetobacteraceae/ <i>Acidicaldus</i>	1
		Sphingomonadales	Sphingomonadaceae/unklassifiziert	1
	Betaproteobacteria	Burkholderiales	Burkholderiaceae/ <i>Burkholderia</i>	1
	Gammaproteobacteria	Acidithiobacillales	Acidithiobacillaceae/ <i>Acidithiobacillus</i>	1
		Enterobacteriales	Enterobacteriaceae/ <i>Escherichia-Shigella</i>	11
Cyanobacteria/Chloroplast		Cyanobacteria	Familie II/GpIIa	2

\*: Insgesamt standen 28 auswertbare Sequenzierungen zur Verfügung, welche bezüglich ihrer höchsten Übereinstimmung phylogenetisch zugeordnet werden konnten.

### 3.3. Funktionelle Durchmusterung thermophiler Metagenombanken

Die etablierten pCT3FK-basierten Metagenombanken von den Azoren sowie von Kompostproben (Miete M12, M. Mientus) in *E. coli* standen für funktionelle Untersuchungen nach aktiven Klonen bereit. Da die Kompost-Bank bereits auf (hemi-)cellulolytische Aktivitäten in *E. coli* untersucht worden war (M. Mientus, persönliche Mitteilung), wurde das Hauptaugenmerk der Durchmusterungen auf lipolytische Aktivitäten begrenzt. Für die heißen Quellen der Azorenbank wurde eine breite Palette an verfügbaren Substraten in *E. coli* im Hochdurchsatzformat (entweder gestempelt oder als gemischte Fosmidpools) bei 60 °C untersucht, darunter Tributyrin, Triolein, Xylan, Carboxymethylcellulose (CMC), Methylumbelliferyl-Cellobiosid (MUC), lösliche Stärke, sowie Magermilch als Overlay, auf Substratplatten oder in Mikrotiterplatten. In *Thermus* war der Abbaunachweis von MUC und Stärke aufgrund zu hoher Hintergrundaktivität nicht möglich. Selbst die Mehrfachmutante BL03, die zu Durchmusterungszwecken verwendet worden ist, zeigte auf Tributyrin eine relativ hohe Restaktivität, welche die Identifizierung potentieller Kandidaten im Hochdurchsatzformat erschwerte. In *E. coli* konnten von insgesamt 8.352 untersuchten Einzel-Fosmidklonen der Azorenbank nur ein amylolytischer Klon in der Biofilmprobe (AZ4-1-H1) sowie eine Tributyrase-Aktivität in der Sedimentprobe 2 (AZ2-4-B6) bei 60 °C identifiziert werden. In der Kompostprobe konnte von 1.920 Einzelfosmiden ein Klon (M12-4-D9) mittels Bildung eines hydrolytischen Hofes um die Kolonie auf Tributyrin festgestellt werden.



**Abbildung 35: Identifizierung positiver Fosmidklone in *E. coli* aus den durchmusterter thermophilen Metagenombanken der Azoren und aus Kompost. Gezeigt sind mehrere retransformierte Klone der analysierten Originalfosmide. Die beobachteten und bestätigten Aktivitäten umfassen Stärkeabbau (visualisiert nach Färbung der Platte mit Lugol'scher Lösung) des Klons AZ4-1-H1 (A) sowie zwei Tributyrasen (visualisiert durch hydrolytische Höfe um die trübe Substratemulsion) des Klons AZ2-4-B6 (B) und M12-4-D9 (C) bei 60 °C. *E. coli* EPI300 Zellen mit Leerfosmid sind jeweils ausgestrichen worden (unten) und zeigen keine Aktivität.**

### 3.3.1. Übertragung der Fosmide in *T. thermophilus*

Nach erfolgreicher funktioneller Durchmusterung der Fosmidklone in *E. coli* wurden mehrere hundert Einzelfosmide im 96-Well Plattenformat in *Thermus* übertragen. Von den ursprünglich 14 Mikrotiterplatten (entsprechend 1.344 *E. coli*-Klone) konnten mittels natürlicher Transformation im Deepwell Block 1.175 Einzelfosmide übertragen werden (87,4 %). Diese Klone standen für funktionelle Untersuchungen auf den oben genannten Substraten zur Verfügung. Es konnte kein aktiver Klon unter diesen hundert Transformanten festgestellt werden. Da die Übertragung der Fosmide in *Thermus* zeit- und arbeitsaufwändig ist (384 Fosmide pro Woche), wurden ebenfalls Fosmidmischungen (Pools) zur Transformation verwendet, um mehr Klone parallel durchmusterung zu können. Dabei wurden pro Transformationsreaktion 96 gemischte Einzelfosmide (entsprechend einer 96-Well Mikrotiterplatte) aus *E. coli* in *Thermus* übertragen. Diese Mischungen aus je 96 Einzelfosmiden wurden auf TB Medium mit Kanamycin (20 µg/mL) ausplattiert, und angewachsene Kolonien abgewaschen. Kleine Volumina (5 bis 10 µL) dieser Bakteriensuspension wurde nach einem mehrmaligen Waschschrift (3 bis 4-malige Wiederholung) in Puffer auf die Substratplatten pipettiert. Aus den restlichen 45 Mikrotiterplatten konnten in den klassischen Detektionsmethoden ebenfalls keine funktionell exprimierten Enzymaktivitäten in *T. thermophilus* nachgewiesen werden.

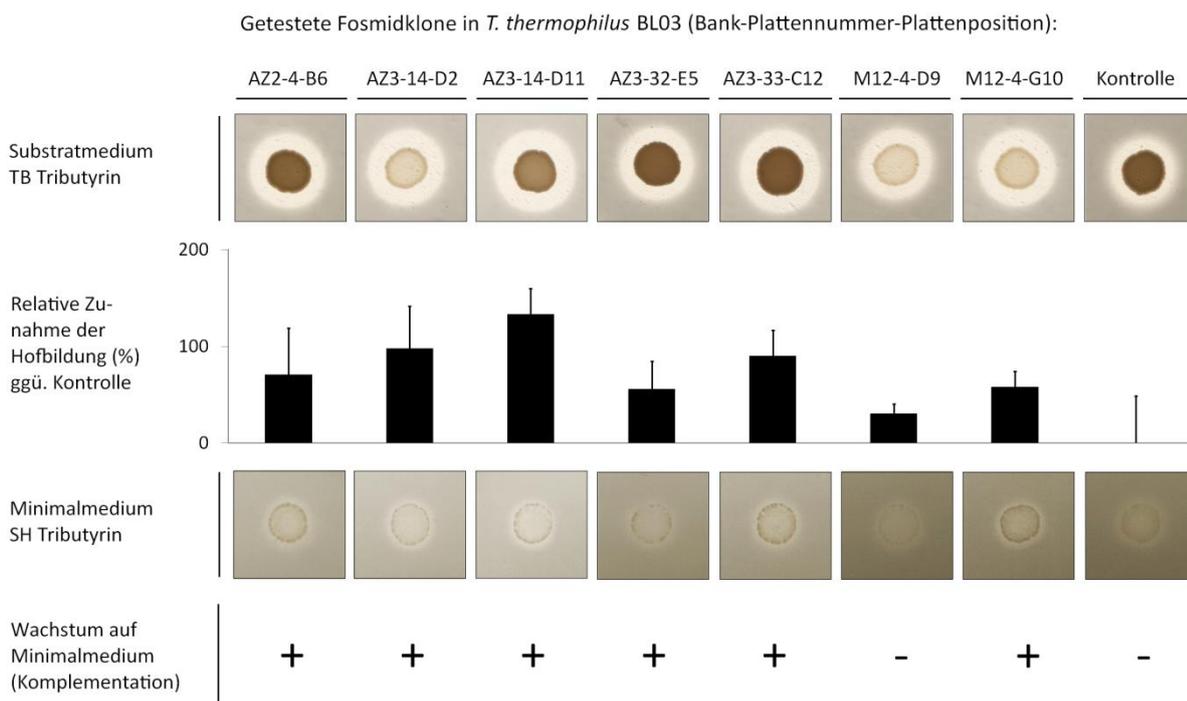
### 3.3.2. Komplementation der Esterase-defizienten *T. thermophilus* Mutante BL03

Die heterologe Komplementation basiert auf der Strategie, einen genetischen Defekt unter selektiven Bedingungen wiederherzustellen. Durch gezielte Gendeletion oder Mutation können Gene bestimmter metabolischer Stoffwechselwege deaktiviert werden, sodass die Mutante nicht mehr in der Lage ist, ein bestimmtes Substrat abzubauen, welches zum Wachstum notwendig ist. Durch die funktionelle Expression metagenomischer DNA kann der Wirt genetisch und metabolisch komplementiert werden, sodass er wieder in der Lage ist, zu wachsen oder andere phänotypische Merkmale zu erlangen. Dieses Prinzip wurde auf die Esterase-defiziente multiple *T. thermophilus*-Mutante BL03 übertragen, welche nicht in der Lage ist, auf Minimalmedium mit Tributyrin zu wachsen. Durch gezieltes Einbringen der verfügbaren thermophilen pCT3FK Fosmidbanken sollten solche Fosmidklone identifizierbar sein, die den defizienten Stamm BL03 wieder zu Wachstum befähigten. Dazu wurden Mischungen aus 96 Fosmiden aller Fosmidbanken aus *E. coli* isoliert, und in Einzeltransformationen von *Thermus* überführt. Nach erfolgter Transformation wurden die gewachsenen und auf 20 µg/ml Kanamycin selektierten Kolonien abgewaschen und mehrmals in Puffer gewaschen (mindestes drei mal). Insgesamt wurden von der Azorenbank 6.048 Fosmide (Sedimentproben AZ2 und AZ3 sowie Biofilme AZ4) und von der Kompostbank (M12) 1.920 Fosmide funktionell auf wachstumsbefähigte Klone durchmustert (entsprechend 83 Mischungen bestehend aus je 96 Einzelfosmidklonen). Von diesen konnten 21 Fosmidpools identifiziert werden, von welchen auf dem Minimalmedium mit Tributyrin bei 60 °C Kolonien angewachsen sind (Abbildung 36). Nach einer ersten Komplementationsrunde konnten so die Fosmidpools (je abgekürzt als „Bankname-Plattenummer“) von AZ2-4, AZ3-14, AZ3-25, AZ3-30, AZ3-32, AZ3-33, AZ3-36, AZ3-37, AZ3-38, AZ3-44, AZ3-47, AZ3-51, AZ3-55, AZ3-56, AZ3-57, AZ4-1, AZ4-2, M12-3, M12-4, M12-6 und M12-11 selektiert werden. Nach mindestens dreimaligen Passagieren der Zellen auf frischen Minimalmedien blieben von den Kandidaten lediglich 13 Fosmidpools übrig, die einen stabilen Phänotyp aufwiesen.

**Abbildung 36: Komplementationsexperiment in *T. thermophilus* nach 3 Tagen Inkubation bei 60 °C auf SH Minimalmedium mit 1,0 % (v/v) Tributyrin supplementiert (Selektion auf 20 µg/ml Kanamycin). Beispielhaft sind zwei Dutzend Fosmidmischungen gezeigt (BL03 mit AZ2-1 bis AZ2-4 (Nr. 1-4), AZ3-1 bis AZ18 (Nr. 5-22), AZ4-1 bis AZ4-2 (Nr. 23-24) sowie die Negativ-Kontrolle (BL03) und die (WT) Positiv-Kontrolle). Der Kontrollstamm BL03 mit pCT3FK-Fosmid ohne *Insert* ist nicht befähigt, auf dem Substrat zu wachsen, der *Thermus* Wildtyp mit pCT3FK-Leervektor zeigt hingegen starke Koloniebildung. Die positiv selektionierten Kandidaten waren AZ2-4 (Nr. 4), AZ3-14 (Nr. 18), AZ4-1 (Nr. 23) und AZ4-2 (Nr. 24), welche sich durch leichtes bis mäßiges Wachstum auszeichneten.**



Aus den entsprechenden 96-Well Mikrotiterplatten wurden dann 96 einzelne Fosmide im Blockformat in *Thermus* übertragen, und mögliche positive Positionen für die Platten AZ2-4, AZ3-14, AZ3-25, AZ3-30, AZ3-32, AZ3-33, AZ3-37, AZ3-38, AZ3-47, AZ4-2, M12-3, M12-4 und M12-6 ermittelt. Nach der Selektion auf Minimalmedien analog zu oben wurden zusätzliche *pNP*-Assays der Rohextrakte durchgeführt, um die lipolytische Aktivität der Kandidatenpositionen zu bestätigen. Von den Fosmidmischungen aus AZ2-4, AZ3-14, AZ3-32, AZ3-33, AZ4-2 und M12-4 konnten schließlich die Einzelpositionen zugeordnet werden, welche für die Esterase-Aktivität verantwortlich sein mussten. Mit Ausnahme von Pool AZ4-2 konnte erneut mittels Selektion durch Komplementation auf Minimalmedium für jede Mikrotiterplatte mindestens eine spezifische Position identifiziert werden, sodass sich die Zahl der Kandidatenfosmide in *Thermus* auf sechs reduzierte (Abbildung 37). Zusätzlich wurde der aus den *E. coli*-Screenings identifizierte Tributyrase-aktive Fosmidklon M12-4-D9 in *T. thermophilus* BL03 einzeln übertragen, welcher jedoch nicht das Wachstum auf Minimalmedium komplementieren konnte.



**Abbildung 37:** Aktivitäten einzelner Fosmidklone nach funktioneller Durchmusterung der thermophilen Metagenombanken in *T. thermophilus* BL03. Gezeigt sind klassische Substrattests auf TB-Medium mit 1,0 % (v/v) Tributyrin, welche eine Zunahme der Hoffläche zeigen (oben). Die relative Zunahme der Hofbildung wurde gegenüber der Kontrolle berechnet (BL03 Stamm mit Leerfosmid entsprechend 0,0 %). Die untere Hälfte der Abbildung zeigt die Fosmidklone und ihr Wachstumsverhalten auf Minimalmedium mit 1,0 % (v/v) Tributyrin. Als Kontrolle diente einerseits Klon M12-4-D9, welcher bereits aus *E. coli*-Screenings aufgefunden worden war, jedoch nicht im Zuge der Komplementationsstrategie identifiziert werden konnte. Als weitere Kontrollte diente ein Fosmid ohne metagenomisches Insert aus Klon AZ4-2-H10. Abbildung modifiziert nach Leis *et al.*, (2015a).

Wie aus der Abbildung ebenfalls ersichtlich ist, zeigten die bestätigten Fosmideinzelklone verschieden ausgeprägte Hydrolysehöfe auf dem Substrat. Trotz der noch immer beachtlichen Hintergrundaktivität des BL03 Stammes waren die beobachteten Hofbildungen durchschnittlich größer zur Kontrolle mit pCT3FK Leerfosmid (Klon AZ4-2-H10:  $0,0 \pm 48,4$  %): Die stärksten Hofbildungen auf TB Substratmedien mit Tributyrin waren bei den Klonen AZ3-14-D2 ( $97,8 \pm 44,0$  %) und AZ3-14-D11 ( $133,1 \pm 26,8$  %) gegenüber der BL03-Kontrolle beobachtet worden. Mittelstarke Hofbildungszunahme zeigten die Klone AZ2-4-B6 ( $70,6 \pm 48,1$  %), AZ3-32-E5 ( $55,8 \pm 28,9$  %), AZ3-33-C12 ( $89,9 \pm 26,6$  %) und M12-4-G10 ( $58,1 \pm 15,9$  %). Keine eindeutige Zunahme der Hofbildung konnte für M12-4-D9 ( $30,5 \pm 9,9$  %) gegenüber dem Leerfosmid verzeichnet werden.

### 3.3.3. Aktivität von Rohextrakten aus positiven Klonen auf *pNP*-Substraten

Als Bestätigung der in den Durchmusterungen identifizierbaren Fosmidkandidaten, wurde die lipolytische Aktivität der einzelnen Rohextrakte mittels *pNP*-Tests in *E. coli* EPI300 und *T. thermophilus* BL03 bestimmt. Dabei wurden die erhaltenen Aktivitäten in mU von der Hintergrundaktivität des Expressionswirts mit Leerfosmid subtrahiert und in Abbildung 38 zusammengefasst. Die in *E. coli* identifizierten Fosmidklone konnten im *pNP*-Test bestätigt werden, wobei die gemessenen Aktivitäten der Rohextrakte für diesen Wirt allgemein gering ausfielen (höchste Werte bei  $6,4 \text{ mU} \times \text{mg}^{-1}$  Rohextrakt für AZ2-4-B6 und  $8,0 \text{ mU} \times \text{mg}^{-1}$  für M12-4-D9 für *para*-Nitrophenyl-caprylat). Mit Ausnahme von Klon M12-4-D9 waren für *pNP*-C<sub>4</sub> und *pNP*-C<sub>12</sub> lediglich Aktivitäten am Detektionslimit feststellbar. Für länger-kettige Fettsäuren wie Palmitinsäure (C<sub>16</sub>) waren keine Aktivitäten nachweisbar. Im Vergleich dazu konnten in *Thermus* wesentlich höhere Aktivitäten festgestellt werden, was ebenfalls auf die relativ hohe, verbleibende lipolytische Hintergrundaktivität des BL03-Expressionstammes mit Leerfosmid zurückzuführen war. Nach Korrektur durch Subtraktion der Hintergrundaktivitäten wies Fosmidklon AZ3-14-D2 über alle getesteten Substrate die höchsten gemessenen Aktivitäten auf, wobei *pNP*-laurat bevorzugt abgebaut wurde ( $64,3 \text{ mU} \times \text{mg}^{-1}$ ). Fosmidklone AZ2-4-B6, AZ4-14-D11 und M12-4-D9 zeigten ähnliche Substratpräferenzen wie in *E. coli* für länger-kettige Fettsäuren (Caprylin- und Laurinsäure), alle anderen Rohextrakte wiesen Präferenzen gegenüber kürzerkettigen *pNP*-Fettsäureestern auf (AZ3-33-C12, AZ3-32-E5 und M12-4-G10 mit je  $25,1 \text{ mU} \times \text{mg}^{-1}$ ,  $11,7 \text{ mU} \times \text{mg}^{-1}$  und  $16,3 \text{ mU} \times \text{mg}^{-1}$  auf *para*-Nitrophenyl-butyrate, C<sub>4</sub>). Auf *pNP*-caprylat (C<sub>8</sub>) war die Reihenfolge (absteigend angeordnet) wie folgt: AZ3-14-D2 > M12-4-D9 > AZ3-33-C12 > M12-4-G10 > AZ4-D11 > AZ2-4-B6 > AZ3-32-E5.

Expressionswirt	Substrat	Getestete Fosmidklone (Bank-Plattennummer-Plattenposition):							Kontrolle
		AZ2-4-B6	AZ3-14-D2	AZ3-14-D11	AZ3-32-E5	AZ3-33-C12	M12-4-D9	M12-4-G10	
<i>E. coli</i> EPI300	<i>p</i> NP-butyrat (C <sub>4</sub> )	0,0	0,3	0,0	n.b.	0,1	2,4	n.b.	0,0
	<i>p</i> NP-caprylat (C <sub>8</sub> )	6,4	1,5	2,9	3,0	1,9	8,0	3,1	0,0
	<i>p</i> NP-laurat (C <sub>12</sub> )	0,6	0,5	1,2	1,0	0,3	1,7	1,0	0,0
	<i>p</i> NP-palmitat (C <sub>16</sub> )	n.b.	n.b.	n.b.	n.b.	n.b.	0,1	n.b.	n.b.
<i>T. thermophilus</i> BLO3	<i>p</i> NP-butyrat (C <sub>4</sub> )	6,7	31,9	0,1	11,7	25,1	11,7	16,3	0,0
	<i>p</i> NP-caprylat (C <sub>8</sub> )	8,4	37,9	9,7	1,9	16,6	30,2	11,7	0,0
	<i>p</i> NP-laurat (C <sub>12</sub> )	10,6	64,3	3,8	n.b.	9,8	n.b.	13,3	0,0
	<i>p</i> NP-palmitat (C <sub>16</sub> )	0,2	9,9	0,5	n.b.	n.b.	n.b.	0,6	0,0

**Abbildung 38: Zusammenfassung der lipolytischen Aktivitäten der Fosmidklone, die in den beiden Expressionwirten mittels *p*NP-Assay bestimmt wurden. Die *Heatmap*-Representationen zeigen die höchsten relativen Aktivitäten eines jeden verwendeten *p*NP-Substrats im Vergleich zwischen *E. coli* und *T. thermophilus*, die einzelnen Werte geben die spezifische Aktivität in mU pro mg Rohextrakt an (Reaktionsbedingungen 50 mM Tris-HCl-Puffer pH 8,0 bei 60 °C mit 0,125 bis 1,25 mM Substratkonzentration). Die Rohextrakte der jeweiligen Wirte stammen von über Nacht Kulturen unter idealen Inkubationsbedingungen (*E. coli*: LB-Medium mit Chloramphenicol 12,5 µg/mL mit Fosmid Autoinduction Solution-Zugabe; *T. thermophilus*: TB-Medium mit Kanamycin 20 µg/mL bei 60 °C Inkubation). Die gezeigten Durchschnittswerte stammen von mindestens zwei Messungen biologischer Replikate. Als Kontrolle diente Leerfosmid Klon AZ4-2-H10. Abkürzung n.b.: nicht bestimmbar.**

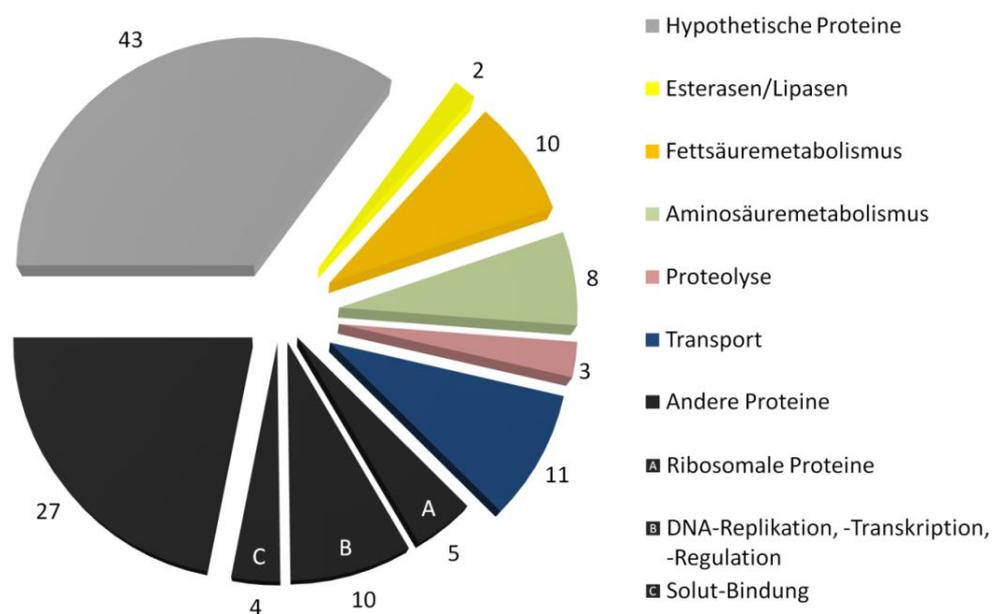
### 3.4. Sequenzierung und *in silico* Analyse der Fosmide der positiven Klone des Screenings

Die erfolgreich durchmusterten und bestätigten Fosmide, die lipolytische Aktivität vermittelten, wurden zur weiteren Analyse mittels 454-NGS Pyrosequencing am G<sub>2</sub>L in Göttingen (Sequenzierprojekt *ANG48*, A. Thürmer) sowie bei LGC-Genomics (Projektname *EST6*) sequenziert. Von insgesamt 8 verschiedenen Fosmidklonen wurden 214,93 kbp assemblierte Sequenzinformationen gewonnen. Dabei wurden 15 *contigs* aus *EST6*, ein *contig* aus *ANG48* und eine *shotgun*-Sequenz erhalten. Die beiden vollständig sequenzierten Fosmidklone AZ3-33-C12 und M12-4-G10 wiesen mit metagenomischer DNA und pCT3FK-Fosmidrückgrat eine Größe von 39,46 bzw. 50,87 kbp auf. Drei Fosmidsequenzen waren nur partiell in zwei bis drei *contigs* assembliert worden und wiesen Größen zwischen 28,0 kbp (AZ3-32-E5), 32,71 kbp (AZ3-14-D11) und 33,25 kbp (AZ3-14-D2) Sequenzinformation auf. Für Fosmid AZ4-2-H10 war kein *Insert* nachweisbar, der Restriktionsverdau

und die Sequenzierung bestätigte dieses Ergebnis. Aus dem Sequenzierprojekt *ANG48* konnte das 10,5 kbp große *contig* 14 dem Fosmidklon AZ2-4-B6 zugeordnet werden. Von Klon M12-4-D9 war lediglich die Sequenzinformation aus einem subklonierten Fragment im Klonierungsvektor pCR-TOPO vorhanden.

Die bioinformatische Auswertung der erhaltenen Sequenzdaten umfasste insgesamt 123 vorhergesagte ORFs, welche bezüglich ihrer GO-Annotation zusammengefasst wurden (Abbildung 39). Etwa ein Drittel davon (35,2 %) waren als hypothetische Proteine vorhergesagt worden, gefolgt von in den Fettsäurestoffwechsel involvierte Genprodukte und zwei  $\alpha/\beta$ -Hydrolasen (9,8 %), sowie Transportrelevante (8,9 %), am Aminosäuremetabolismus beteiligte Proteine (6,5 %) und proteolytische Enzyme (2,4 %). Die restlichen ORFs (38,5 %) umfassten Proteine anderer Klassifikationen (z.B. Solutbindung, DNA-Replikation und ribosomale Proteine). Von den 8 verfügbaren Fosmiden konnten 5 archaeele Sequenzen identifiziert werden, welche alle aus der Metagenomprobe der heißen Quellen stammten und lediglich in *T. thermophilus* aufgefunden werden konnten, nicht hingegen in *E. coli*.

Die Ergebnisse der Sequenzierung nach erfolgter Assemblierung sind in Tabelle 44 (siehe Anhang) dargestellt. Im Falle des Fosmids M12-4-D9 wurde eine *shotgun*-Library in pCR-TOPO erstellt, wobei per Aktivitätsnachweis in *E. coli* zwei Klone mit gleichem *Insert* identifiziert werden konnten. Das ca. 2,6 bis 2,7 kb große *Insert* kodiert für einen 204 Aminosäuren langes Protein, das 99% Sequenzidentität zu einem annotierten hypothetischen Protein von *Caldibacillus debilis* aufweist (Accession Nummer WP\_020155129.1,  $e$ -value  $1 \times e^{-141}$ ) und 71 % Sequenzübereinstimmung zu einer charakterisierten Carboxylesterase Est30 aus *Geobacillus stearothermophilus* hat, wovon bereits eine Kristallstruktur aufgeklärt worden ist (Accession 1TQH\_A,  $e$ -value  $8 \times e^{-102}$ ). Die Analyse der Proteindomänen mittels Pfam zeigte eine charakteristische  $\alpha/\beta$ -Hydrolase der Familie 6.



**Abbildung 39:** Übersicht der GO-Klassifizierungen aller kodierten Proteine, die aus den Metagenomsequenzen vorhergesagt werden konnten.

Fosmid AZ2-4-B6, das zunächst in *E. coli* identifiziert worden war, konnte nur unvollständig sequenziert werden (*contig* 14 aus Sequenzierprojekt ANG48). ORF 8c konnte als putative Esterase identifiziert werden, da das 291 Aminosäuren lange Protein einer  $\alpha/\beta$ -Hydrolase der Familie 6 angehört. Seine Sequenz hat eine hohe Übereinstimmung zu annotierten 2-Hydroxy-6-Oxo-6-Phenylhexa-2,4-Dienoat Hydrolasen (biphenyl\_bphD Zuordnung laut *conserved domain database*, CDD) aus *Dyella ginsengisoli* LA-4 (71 % zu ACH87186.1, *e-value*  $1 \times e^{-149}$ ), bzw. *Alicyclophilus denitrificans* BC (73 % zu YP\_004125001.1, *e-value*  $2 \times e^{-148}$ ).

Von den anderen putativen Esterasen sind 454-NGS Pyrosequenzierungen mit markierten Sequenzen (*multiplexing*) durchgeführt worden (LGC-Genomics, Berlin). Dadurch waren die assemblierten Sequenzen jedem Fosmidklon eindeutig zuordenbar. Von den erhaltenen Sequenzierungen waren lediglich zwei Fosmide vollständig in ein *contig* assembliert worden (AZ3-33-C12 mit knapp 40 kb und M12-4-G10 mit über 50 kb Sequenzgröße). Unvollständige aber größtenteils assemblierte Sequenzierungen waren im Fall von AZ3-14-D2, AZ3-14-D11 und AZ3-32-E5 möglich (Abbildung 40). Bei der *in silico* Analyse der vollständig sequenzierten Fosmide konnten keine direkten Kandidatengene identifiziert werden, da keine Übereinstimmung mit bekannten annotierten Esterasen oder lipolytischen Enzymen bestanden (Tabelle 32). Die Vorhersage durch Motivsuche mittels *regular expressions* für das konservierte Pentapeptid „G-X-S-X-G“ (Arpigny und Jäger, 1999) lieferte mehrere Kandidatengene. Im zweiten Suchlauf mittels manueller *blastp*-Suche wurden die erhaltenen Genprodukte erneut auf Ähnlichkeiten zu annotierten Esterasen/Lipasen untersucht, und im Falle von AZ3-33-C12 und M12-4-G10 konnte je ein Kandidatengen identifiziert werden. Für AZ3-33-C12 kodierte ORF 2 für ein hypothetisches Protein aus 152 Aminosäuren, welches höchste Sequenzübereinstimmung zum Genprodukt von Ahos\_0948 aus dem Archaeon *Acidianus hospitalis* W1 hat (YP\_004458132.1 mit 66 % Identität, *e-value*  $1 \times e^{-59}$ ). Ferner ist eine annotierte archaeele Phospholipase A2/Esterase aus *Desulfurococcus kamchatkensis* 1221n (DKAM\_1126, YP\_002428819.1 mit 45 % Sequenzübereinstimmung, *e-value*  $8 \times e^{-17}$ ) und *Aeropyrum pernix* K1 (APE\_2325, NP\_148539.1 mit 48 % Sequenzübereinstimmung, *e-value*  $5 \times e^{-16}$ ) vorhergesagt worden, die beide wesentlich geringere Ähnlichkeiten aufweisen. Fosmid M12-4-G10 zeigte das Pentapeptid-Muster für ORF 18c, der für eine 581 aa langes Protein kodiert und aus drei Domänen besteht. Laut Pfam Klassifizierung handelt es sich bei der N-terminalen Domäne dieses Enzyms um ein Mitglied der Metallo- $\beta$ -Lactamase B-Superfamilie, welche neben  $\beta$ -Lactamasen auch Thiolesterasen, Glyoxalase II Proteinfamilie und Kompetenzproteine für die natürliche Transformation von *Neisseria gonorrhoeae* umfasst. Die beiden C-terminalen Domänen zeigen signifikante Übereinstimmung zu Alkyl-Sulfatasen (Dimerisierungs-Domäne und C-terminale Domäne). Die höchste Sequenzidentität von insgesamt 40 % war zu einem hypothetischen Protein L288\_08650 aus *Sphingobium quisquiliarum* P25 (EQB08083.1, *e-value*  $2 \times e^{-141}$ ) sowie zu Lactamasen diverser *Vibrio* Gattungen (39 %, *e-values* zwischen  $4 \times e^{-137}$  bis  $8 \times e^{-135}$ ) nachweisbar.

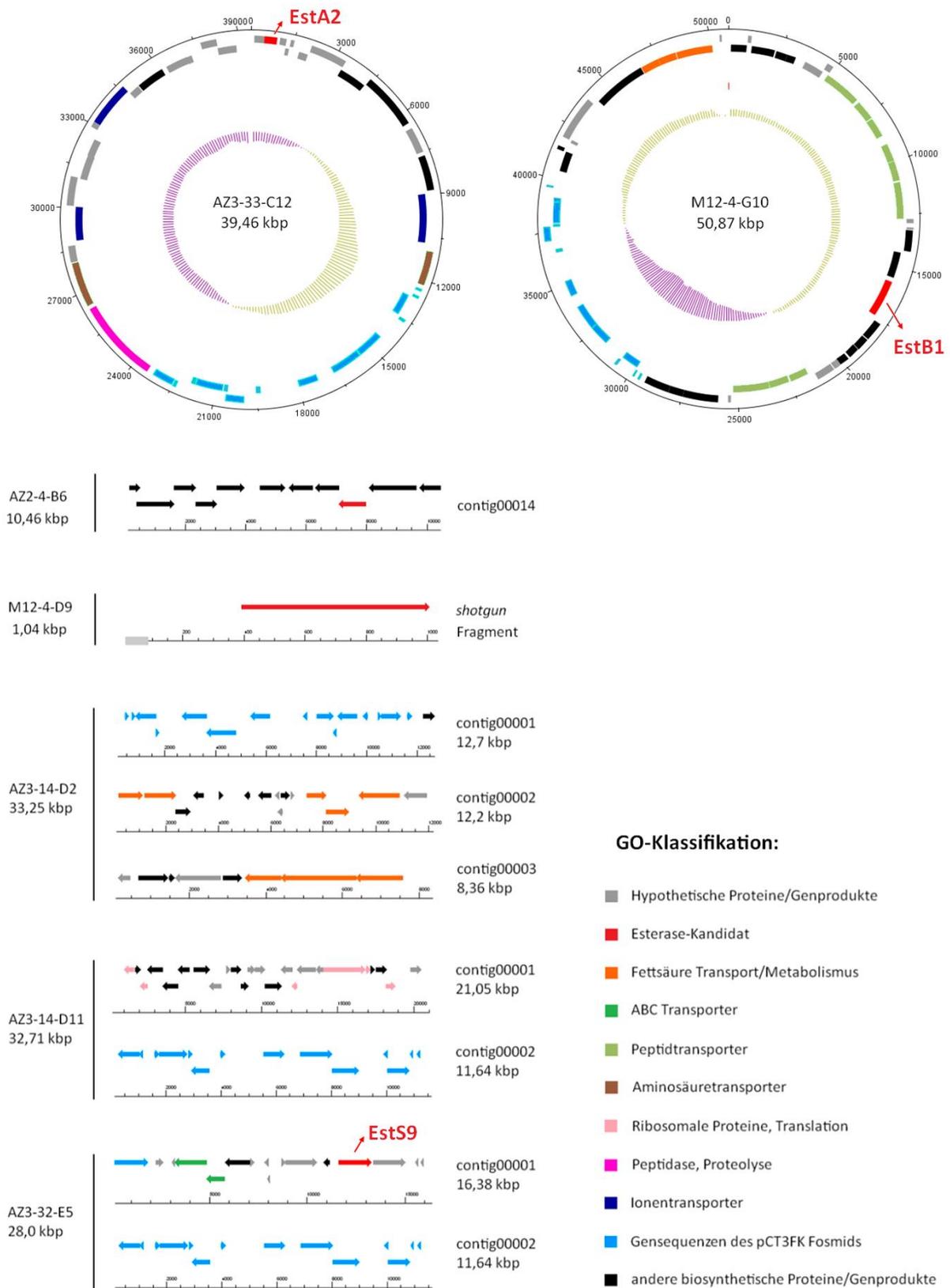


Abbildung 40: Übersicht aller sequenzierten und assemblierten Fosmide aus den vergleichenden Durchmusterungen in *E. coli* und *T. thermophilus*. Die vorhergesagten ORFs (Pfeile und Blöcke) sowie ihre kodierten Proteine wurden bezüglich ihrer funktionellen Zugehörigkeit klassifiziert (siehe Legende mit verschiedenen Farbkodierungen). Die beiden vollständig sequenzierten Fosmide AZ3-33-C12 und M12-4-G10 sind als zirkuläre Vektorkarte angegeben, nicht vollständig sequenzierte Fosmide sind linear dargestellt. Drei ausgewählte Esterasen wurden im Laufe der Arbeit näher charakterisiert (rote Pfeile). Die Fosmid-Abbildung wurde erneut mit dem DNAPlotter durchgeführt (siehe Schema in Abbildung 20).

Tabelle 32: Bioinformatische Analyse der sequenzierten und assemblierten Fosmidklone aus vergleichenden Durchmusterungen in beiden Wirten.

Fosmid	Identifiziert in	Assemblierte Sequenz	Kandidat ORF	Höchste Übereinstimmung ( <i>blastp</i> )	Annotierte Protein-domänen (Pfam)
AZ2-4-B6 <sup>1</sup>	<i>E. coli</i> <i>T. thermophilus</i>	10 kb <i>contig 14</i>	8c	meta-fission product hydrolase ( <i>Dyella ginsengisoli</i> LA-4) $\alpha/\beta$ -Hydrolase ( <i>Alicyclophilus denitrificans</i> BC)	$\alpha/\beta$ -Hydrolase 6
AZ3-14-D2	<i>T. thermophilus</i>	35 kb	-	<i>Metallosphaera sedula</i> DSM 5348	-
AZ3-14-D11	<i>T. thermophilus</i>	33 kb	-	<i>Sulfolobus</i> Spezies	-
AZ3-32-E5	<i>T. thermophilus</i>	29 kb	12	Dipeptidyl-Aminopeptidase / Acyl-Aminoacyl-Peptidase aus <i>Sulfolobus islandicus</i> LAL14/1	Peptidase_S9
AZ3-33-C12 <sup>2</sup>	<i>T. thermophilus</i>	39 kb	2	Hypothetisches Protein Ahos_0948 ( <i>Acidianus hospitalis</i> W1) Phospholipase A2 / Esterase ( <i>Desulfurococcus kamchatkensis</i> )	PaREP1 Archaeelle PaREP1/PaREP8 Familie (nicht signifikant)
M12-4-D9 <sup>3</sup>	<i>E. coli</i>	<i>shotgun</i> -Bank	1	Hypothetisches Protein ( <i>Caldibacillus debilis</i> ) Carboxyl-Esterase Est30 ( <i>Geobacillus stearothermophilus</i> )	$\alpha/\beta$ -Hydrolase 6
M12-4-G10 <sup>2</sup>	<i>T. thermophilus</i>	51 kb	18c	Hypothetisches Protein ( <i>Sphingobium quisquiliarum</i> P25)	Lactamase B Metallo- $\beta$ -Lactamase Superfamilie, Alkyl-Sulfatase mit C-terminaler und Dimerisierung-Domäne

1: Fosmid sequenziert am G<sub>2</sub>L aus Fosmidpool ANG48.

2: Fosmid vollständig sequenziert.

3: ORF aus *shotgun*-Genbank in pCR-TOPO identifiziert, keine weiteren Sequenzdaten vorhanden.

Im nicht vollständig sequenzierten Fosmid AZ3-32-E5 konnte für ORF 12 eine Prolyl-Oligopeptidase (Peptidase\_S9) vorhergesagt werden. Diese Enzymfamilie hat laut Pfam-Informationen Ähnlichkeiten zu verschiedenen  $\alpha/\beta$ -Hydrolase-Familien und CO-Esterasen. Eine Suche zeigte mit 60 % die höchste Übereinstimmung mit einer bakteriellen Dipeptidyl-Amino-peptidase/Acylaminoacyl-Peptidase aus *Sulfolobus islandicus* LAL14/1 (Accession Nummer YP\_007865031.1, *e-value* 0,0).

Für die restlichen Fosmide konnten keine plausiblen Kandidatengene identifiziert werden. Aus den Sequenzdaten des Fosmids AZ3-14-D2 konnte zwar die Präsenz mehrerer am Fettsäuremetabolismus beteiligten Enzyme gefunden werden, jedoch konnte kein eindeutiges Kandidatengen abgeleitet werden.

### 3.5. Expression und vorläufige Charakterisierung ausgewählter Esterasen

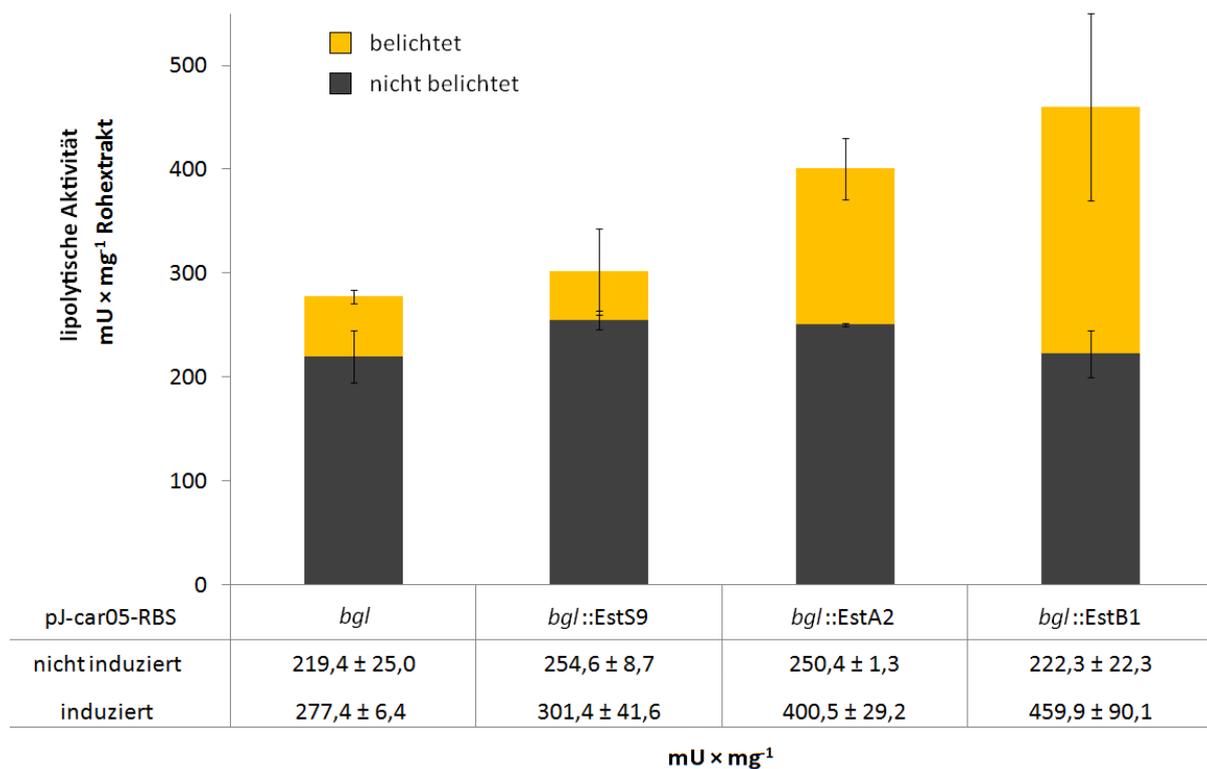
Die erhaltenen neuartigen lipolytischen Kandidatengene der Fosmide AZ3-32-E5 (im Folgenden EstS9), AZ3-33-C12 (EstA2) und M12-4-G10 (EstB1) wurden in Expressionsvektoren mit und ohne poly-Histidin-tag für *E. coli* und *Thermus* kloniert und erste Aktivitätstests zur Charakterisierung der Esterasen wurden durchgeführt. Dabei standen für die Expressionsversuche neben etablierten Wirt/Vektor-Systemen in beiden Wirten (pET21a für *E. coli* BL21, pMKE2 und pMK18 für *Thermus*) auch das neuartige, licht-induzierbare Expressionssystem auf Basis des Carotinoid-Clusters in *T. thermophilus* zur Verfügung (pJ-car05-RBS, Terzi, Lehrstuhl für Mikrobiologie).

#### 3.5.1. Expression in *T. thermophilus*

Für die induzierbare Expression der Zielproteine EstS9, EstA2 und EstB1 wurden die drei in *Thermus* verfügbaren Expressionssysteme verwendet. Ausgehend von Plasmid pJ-car05-RBS-Bgl (F. Terzi, Masterthesis, 2012) wurde eine Vektor-PCR (pJ-car05\_R und GA-car\_RBS-r) durchgeführt, und das 5,1 kb große Fragment mit den jeweiligen Esterase-Sequenzen durch Gibson Assembly erfolgreich assembliert (Primerpaare der Zielproteine: car-AZ332E5\_F und car-AZ332E5-R für EstS9, car-AZ333C12\_F und car-AZ333C12-R für EstA2, car-M124G10\_F und -R für EstB1). Die Vektoren pJ-car05-RBS-EstS9, pJ-car05-RBS-EstA2 und pJ-car05-RBS-EstB1 wurden in *T. thermophilus* Stamm BL03 übertragen und die Integration der Expressionsvektoren durch Kanamycin-Supplementation mit Endkonzentrationen von 20  $\mu\text{g}/\text{mL}$  sichergestellt. In Vortests konnten die Aktivitätsmaxima der klonierten und exprimierten Konstrukte ermittelt werden. Dabei waren die höchsten messbaren Aktivitäten im Bereich um  $85 \pm 5$  °C und pH 8,0-8,8 gemessen worden, wobei kürzerkettige Substrate

wie *pNP-C<sub>4</sub>* die höchsten Aktivitäten aufwiesen. Die Expressionsexperimente wurden schließlich unter den gewählten Idealbedingungen der metagenomischen Esterasen durchgeführt. Die Aktivitäten der Rohextrakte vor und nach Induktion mit Licht und höheren Schüttelgeschwindigkeiten zum gesteigerten Sauerstoffeintrag im Kulturmedium wurden parallel getestet und miteinander verglichen (Abbildung 41). Als Negativkontrolle ohne zusätzliche Esterase wurde der ursprüngliche Vektor *pJ-car05-RBS* verwendet, der die *Thermus*-eigene  $\beta$ -Glucosidase (*bgl*) homolog exprimiert und über keine zusätzliche Esterase verfügt.

Unter nicht-induktiven Bedingungen (Kulturen wurden zum Schutz vor Licht mit Aluminiumfolie abgedeckt) traten in allen Expressionskulturen durchschnittliche Hintergrundaktivitäten von  $236,7 \pm 18,4 \text{ mU} \times \text{mg}^{-1}$  Rohextrakt auf ( $n = 4$ ). Während bei dem Expressionskonstrukt mit *Bgl* und *EstS9* keine wesentlichen Zunahmen der lipolytischen Aktivitäten nachweisbar waren, zeigten die beiden metagenomischen Esterasen *EstA2* und *EstB1* mit etwa 150 mU und 237,6 mU signifikante Aktivitätssteigerungen gegenüber nicht-induzierender Kultivierungsbedingung in *pJ-car05-RBS*.



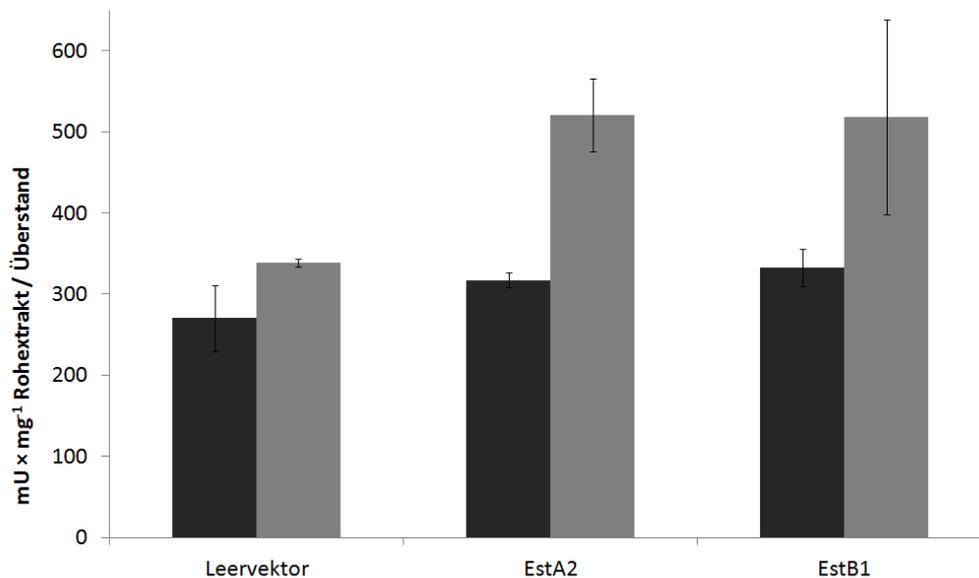
**Abbildung 41:** Aktivitätsdaten der getesteten Esterasen im *pJ-car05-RBS* Expressionsvektor. Gezeigt sind der nicht-induzierte Zustand (dunkle Balken) und nach 4 Stunden Exposition der Kulturen mit Licht und erhöhtem Sauerstoffeintrag (helle Balken). Als Negativ-Kontrolle wurde die *Thermus*-eigene  $\beta$ -Glucosidase (*bgl*, TT\_P0042) in BL03 exprimiert und mit den Expressionsprofilen der metagenomischen lipolytischen Kandidaten *EstS9*, *EstA2* und *EstB* verglichen. Die Aktivitätsparameter waren 85 °C in 50 mM Tris-HCl pH 8.8 auf *pNP-C<sub>4</sub>* Substrat. Die Mittelwerte ( $n = 3$ ) der Aktivitäten  $\pm$  Standardabweichungen sind unter den Balken angegeben.

Zum Vergleich des neuen Systems mit einem etablierten Expressionssystem für *Thermus* wurde das pMKE2 System für die Expression der aktiven Esterasen getestet. Der pMKE2-Vektor ist von der Präsenz des *nar*-Operons aus *T. thermophilus* HB8 abhängig, welches in den HB27 Expressionstamm kloniert wurde (verwendeter Stamm HB27::*nar*, Ramírez-Arcos *et al.*, 1998b). Die Klonierungen für EstA2 und EstB1 aus den bestehenden Expressionsvektoren (beispielsweise pJ-car05-RBS) in pMKE2 waren über die kompatiblen Schnittstellen *Nde*I und *Not*I möglich (resultierende Vektoren pMKE2-EstA2 und pMKE2-EstB1). Die Induktion erfolgte über 40 mM KNO<sub>3</sub> einer 35 mL Kultur ohne Schütteln für 4 Stunden bei 60 °C zur Aktivierung des Nitratreduktase-Promotors. Die Rohextrakte der induzierten bzw. nicht induzierten Kulturen der beiden Esterasen sowie der Negativkontrolle (pMKE2 Leervektor) wurden miteinander analog der Aktivitätstests mit *p*NP-Butyrat bei 85 °C und pH 8,8 getestet. Dabei konnten allerdings keine nachweisbaren Aktivitätsunterschiede zwischen den Rohextrakten der Leervektor-Kontrolle und den Esterasen EstA2 und EstB1 festgestellt werden.

Als weiteres Expressionssystem für die konstitutive Expression der Zielgene wurde der pMK18 *E. coli*-*Thermus Shuttle*-Vektor für beide Esterasen EstA2 und EstB1 verwendet. Der Vektor verfügt über einen starken *surface layer protein* Promotor P<sub>slp</sub> und eine Kanamycin-Resistenzkassette (*kat*) sowie einer weiter *downstream* lokalisierten Multiplen Klonierungsstelle (*MCS*), in welcher die Zielgene kotranskriptional gelegen hinter die *kat*-Kassette mittels Gibson Assembly kloniert wurden. Zur Amplifikation der vollständigen Zielgene wurden die Primer aus der Klonierung in pJ-car05-RBS verwendet, das Primerpaar pMK18\_RBS\_FOR/pMK18\_RBS\_REV diente für die Vektor-PCR (als Matrize wurde *Hind*III-linearisierter pMK18-Vektor verwendet). Zur verbesserten Rekrutierung der Ribosomen und gesteigerten Expression der Esterasen wurde eine zusätzliche ribosomale Bindestelle (**RBS**) vor das Startcodon (*ATG*) platziert (5'-ATTAAGGAGGTGTGAGGCATATG-3'). Die Konstrukte wurden pMK18-EstA2his und pMK18-EstB1his genannt und konnten erfolgreich in *E. coli* überprüft, in *T. thermophilus* BL03 übertragen und auf TB-Medien mit 20 µg/mL Kanamycin bei 60 °C selektiert werden. Von je 5 mL Vorkultur wurde 10 %ig in 35 mL Expressionskultur überimpft und nach sieben und 14 Stunden die Zellen geerntet, aufgeschlossen und die lipolytischen Aktivitäten der Rohextrakte und Überstände unter optimalen Reaktionsbedingungen gemessen (85 °C, 50 mM Tris-HCl) (Abbildung 42).

Nach sieben Stunden konnte kein signifikanter Unterschied zwischen den Rohextrakten gemessen werden, nach weiteren sieben Stunden fielen die Unterschiede immer noch sehr gering aus. Eine Analyse der erfolgreichen Proteinexpression mittels SDS-PAGE zeigte keine überrepräsentierten Banden in den Rohextrakten gegenüber der Leervektor-Kontrolle. Bezogen auf den gemessenen Proteingehalt der Überstände nach 14 Stunden Inkubation waren die spezifischen Aktivitäten der extrazellulären Fraktion höher ausgeprägt als bei den intrazellulären Fraktionen. Die womöglich exprimierten Proteine EstA2 und EstB1 zeigten in den extrazellulären Fraktionen mehr Aktivität als die Kontrolle ( $521,1 \pm 44,8$  und  $518,2 \pm 119,7$  mU pro mg Protein gegenüber  $338,4 \pm 4,8$  mU  $\times$  mg<sup>-1</sup>).

Die Untersuchung der Proteinexpression in den vorhandenen *Thermus*-Expressionssystemen konnte zeigen, dass mindestens zwei der drei Kandidatengene hydrolytische Aktivität auf den getesteten pNP-Substraten aufwiesen.

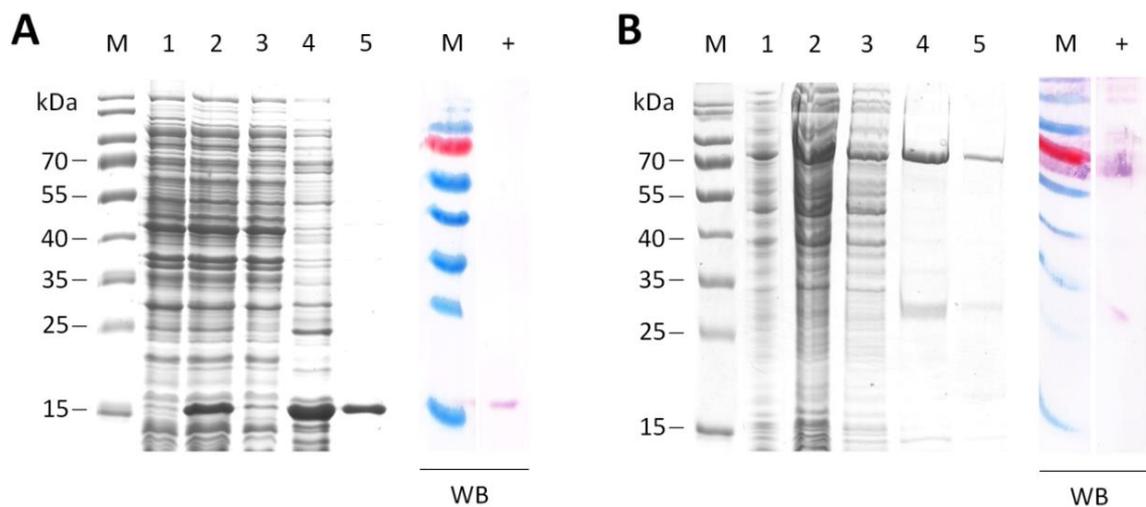


**Abbildung 42:** Expression der metagenomischen Esterasen EstA2 und EstB1 mit pMK18 Plasmid in *T. thermophilus* BL03 nach 14 Stunden konstitutiver Expression bei 60 °C. Das Balkendiagramm zeigt die Zusammenfassung der Aktivitätsdaten auf pNP-C<sub>4</sub> Substrat der intrazellulären (schwarze Balken) und extrazellulären Fraktionen (graue Balken) im direkten Vergleich.

### 3.5.2. pET21a(+)-Expression in *E. coli*

Zur verbesserten Expression der beiden Esterasen EstA2 und EstB1 wurden die vollständigen Gene in den gängigen pET21a(+)-Vektor über *Nde*I und *Not*I aus den vorhandenen pJ-car05-RBS-Konstrukten ligiert. Zum Anfügen eines poly-Histidin-tags (6xHis) wurden die bisherigen Expressionsvektoren mittels *site-directed Mutagenesis* (Affymetrix) mit M124G10\_PHO\_F (für EstB1) und AZ333C12\_PHO\_F (für EstA2) abgeändert (resultierende Vektoren wurden pET21a-EstA2his und pET21a-EstB1his genannt). Nach erfolgreicher Klonierung und Sequenzierung wurden die Konstrukte zur Expression in *E. coli* BL21 übertragen. Der C-terminal fusionierte 6xHis-tag stellte sicher, dass die metagenomischen Proteine zur Charakterisierung schließlich über immobilisierte Metallionen-Affinitätschromatographie aufgereinigt werden können. Im Gegensatz zu den verwendeten *Thermus*-Expressionsvektoren konnte die Herstellung rekombinanten Proteins in *E. coli* mittels analytischer SDS-PAGE und Western Blot nachgewiesen werden. Zur Expression der Zielproteine wurde je

250 mL Kulturvolumen verwendet, welches mit je 1,0 mM IPTG für 4 Stunden bei 37 °C induziert wurde, ehe die Zellen zur Zellernte pelletiert wurden. Die Pellets wurden eingewogen und nach Aufschluss mittels Ultraschall-Behandlung wurden die löslichen Fraktionen der Zielproteinen parallel mit dem Protino Ni-IDA 2000 Protein Purification System (Macherey-Nagel) aufgereinigt. Der Verlauf der Expression und Aufreinigungsschritte ist in Abbildung 43 dargestellt. Die erfolgreiche Expression war gegenüber der Leervektor-Kontrolle im SDS-Gel nachzuweisen, die beiden Zielproteine befanden sich kaum im Durchfluss und ließen sich mit guten Ausbeuten aufreinigen. Das rekombinante EstA2 Protein mit His-tag hatte eine vorhergesagte Größe von etwa 18,7 kDa. EstB1 ließ sich zwischen 65 und 70 kDa auftrennen und stimmte mit dem theoretischen Molekulargewicht (64,6 kDa) überein. Leichte Kontaminationsbanden traten bei der aufgereinigten Esterase EstB1 auf und konnten durch eine weitere Hitzebehandlung bei 60 °C minimiert werden (nicht gezeigt). Zur Immunodetektion mittels Western Blot wurden die beiden heterolog exprimierten Zielproteine unter denaturierenden Bedingungen gelelektrophoretisch aufgetrennt (SDS-PAGE) und durch einen primären poly-Histidin-spezifischen Antikörper (GenScript, USA) detektiert.

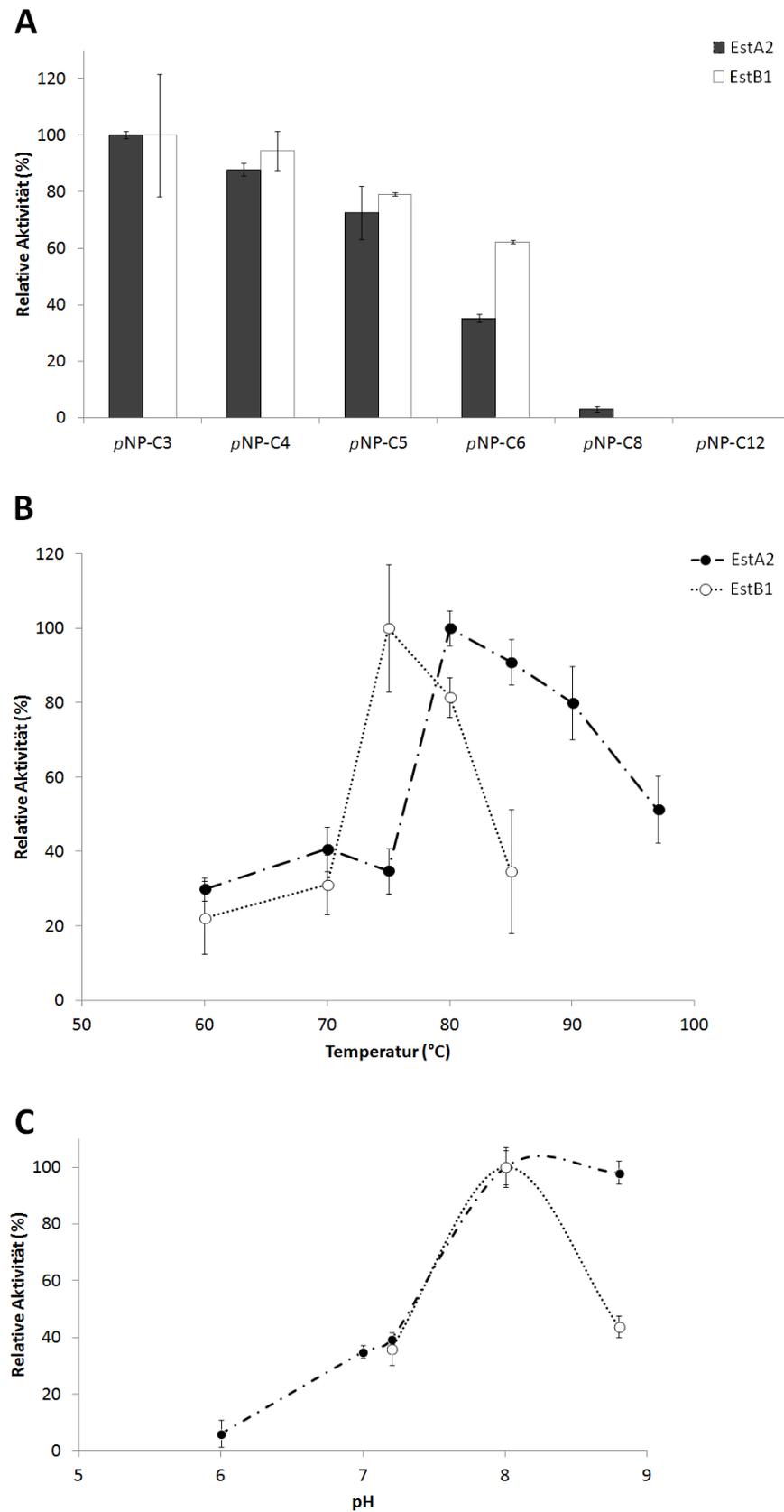


**Abbildung 43:** Visualisierung der Aufreinigungsschritte der beiden metagenomischen Esterasen EstA2 (A) sowie EstB1 (B) mittels SDS-PAGE und Western-Blot. Gezeigt sind jeweils die Rohextrakte aus pET21a(+) Leervektor-Kontrolle (Spur 1) und des Zielproteins (2), der Durchfluss der IMAC Aufreinigung (3) sowie die Fraktionen des Eluats (4 und 5). Die Immunodetektion der heterolog exprimierten und aufgereinigten Esterasen EstA2 und EstB1 (jeweils Spur: +) mit C-terminalen Poly-Histidin-tag wurde mittels Western Blot (WB) visualisiert. Hierfür wurden circa 0,02 µg EstA2 in einem höher-konzentriertem Gel (18 %) aufgetrennt, für EstB1 waren es 0,2 µg in einem 12 % Gel. Als molekularer Größenstandard diente der PageRuler Prestained Protein-Marker (Fermentas) (M). Abbildung modifiziert nach Leis *et al.*, (2015a).

Für enzymatische Charakterisierungen wurden aus 250 mL Expressionskulturen die Zielproteine außerdem mit einer Pharmacia LKB FPLC über Nickel-beladene HisTrap-Säulen (GE Healthcare) aufgereinigt. Die ermittelten Proteinkonzentrationen nach Aufkonzentrieren und Umpuffern lagen bei  $0,36 \pm 0,04 \text{ mg} \times \text{ml}^{-1}$  für EstA2, bzw. bei  $0,33 \pm 0,04 \text{ mg} \times \text{ml}^{-1}$  für EstB1. Die ermittelten spezifischen Aktivitäten unter Idealbedingungen der Enzyme waren  $44,05 \pm 2,06 \text{ U} \times \text{mg}^{-1}$  für EstA2 und  $4,13 \pm 0,71 \text{ U} \times \text{mg}^{-1}$  für EstB1 auf *p*NP-butyrat (zur Bestimmung der optimalen biochemischen Parameter, siehe folgenden Abschnitt IV.3.5.3, Abbildung 44).

### 3.5.3. Biochemische Charakterisierung

Die aufgereinigten Enzyme wurden biochemisch bezüglich ihrer Substratspezifitäten sowie Temperatur- und pH-Optima charakterisiert (Abbildung 44). Beide Enzyme spiegelten Substratpräferenzen gegenüber kurzkettigen Fettsäureestern wider, wobei die höchsten Aktivitäten bei *para*-Nitrophenyl-propionat (*p*NP-C<sub>3</sub>) beobachtet werden konnten. Die Aktivität nahm mit zunehmender Länge der Fettsäurereste graduell ab (*p*NP-C<sub>3</sub> > C<sub>4</sub> > C<sub>5</sub> > C<sub>6</sub> > C<sub>8</sub> > C<sub>12</sub>). Aktivität auf länger-kettigen Substraten wie *p*NP-Laurat konnte lediglich nach verlängerten Inkubationszeiten oder dem Einsatz von mehr Enzym erhalten werden. Die ermittelten optimalen Inkubationstemperaturen ( $T_{\text{opt}}$ ) der Enzyme liegen bei 75 °C für EstB1 und 80 °C für EstA2. Bei niedrigeren Temperaturen um 60 °C entsprach die Restaktivität nur noch 20 – 30 %. Bei Temperaturen um den Siedepunkt von Wasser war nur noch ein Bruchteil der Aktivitäten messbar. Das pH-Optimum beider Enzyme lag bei einem pH-Wert von 8,0-8,8 bei den jeweils optimalen Inkubationstemperaturen. Der Einfluss zusätzlich getesteter Salze (NaCl, KCl, CaCl<sub>2</sub>, MgCl<sub>2</sub>) und möglicher Inhibitoren (Phenylmethylsulfonylfluorid, EDTA) bei Konzentrationen von 1 mM auf die Enzymaktivität wurde ebenfalls bestimmt, jedoch zeigte keines der zugegebenen Additive Auswirkungen auf die Aktivität.



**Abbildung 44: Enzymatische Charakterisierung und Aktivitätsparameter der aufgereinigten metagenomischen Esterasen EstA2 und EstB1. Abhängigkeit der getesteten pNP-Substrate (A), Temperaturspektrum (B) und pH-Bereich (C). Die Mittelwerte  $\pm$  einfache Standardabweichungen stammen mindestens aus Zweifachbestimmungen. Abbildung modifiziert nach Leis *et al.*, (2015a).**

## V. Diskussion

### 1. Verbesserung des *Thermus* Wirt-Vektor-Systems

Die Nutzung von *T. thermophilus* HB27 als thermophilen Expressionswirt für Metagenomanalysen umfasste mehrere Teilaspekte: Die Reduktion von Hintergrundaktivitäten der getesteten Substrate, die Etablierung einer ungerichteten Transposon-Mutagenese und einer gerichteten Mutagenesestrategie zur Marker-freien Manipulation des Genoms sowie die Etablierung eines neuartigen Proteinexpressionssystems zur Herstellung rekombinanter Enzyme.

#### 1.1. Entwicklung eines licht-induzierbaren Expressionssystems auf Basis des *Car*-Genclusters

Im Gegensatz zu *E. coli*, *Bacillus* oder anderen mesophilen Bakterien gibt es für thermophile Wirte wie *T. thermophilus* nur sehr wenige etablierte Protein-Expressionssysteme. Der *E. coli*-*Thermus Shuttle*-Vektor pMK18 (de Grado *et al.*, 1999) erlaubt die Positionierung von Zielgenen hinter den konstitutiven *surface layer protein* Promotor ( $P_{slp}$ ), welcher jedoch keine gezielte Induktion der Proteinexpression ermöglicht. Ein kommerziell erhältliches Expressionssystem pMKE2 (Biotools B&M Labs, S.A., Madrid, Spanien) basiert auf der kontrollierten Induktion des Nitratreduktase (*nar*)-Operon Promotors ( $P_{nar}$ ) unter Anwesenheit von Nitrat bei anoxischen Inkubationsbedingungen (Moreno *et al.*, 2005). Die Denitrifikation als anaerober Atmungsmechanismus ist für einige anaerobe und fakultativ anaerobe Bakterien beschrieben worden (Zumft, 1997), darunter für *T. thermophilus* HB8 (Ramírez-Arcos *et al.*, 1998a). Die beteiligten und charakterisierten Gene liegen auf einer sogenannten NCE-Region (*nitrate respiratory conjugative element*) und setzen sich aus Komponenten des *nar*-Operons, des *nrc*-Operons (*nitrate respiration chain*) und einer regulatorischen Region *dnrST* zusammen (Cava und Berenguer, 2006). Unter anaeroben Bedingungen aktiviert der Sauerstoff-Sensor DnrS die Promotoren des *nar*- und *dnr*-Operons. Der Sauerstoff-unabhängige Transkriptionsfaktor DnrT wird für die Transkription aller beteiligter Operons des NCE benötigt (Cava *et al.*, 2007). Das *E. coli*-*Thermus Shuttle*-Plasmid pMKE2 enthält die multiple Klonierungsstelle (MCS) von pET28b+ (Novagen) mit 6xHis-tag sowie einen modifizierten *nar*-Promotor ( $P_{nar}$ ) und kann durch Kaliumnitrat unter sauerstoffarmen Bedingungen induziert werden (Moreno *et al.*, 2005). Hierfür wird jedoch ein modifizierter *T. thermophilus*-Stamm (HB27::*nar*) benötigt, welchem die notwendigen Elemente der NCE-Region (darunter *narGC*, *dnrST*, *nrcNFED*) von *T. thermophilus* HB8 übertragen wurden (Ramírez-Arcos *et al.*, 1998b). Obwohl das Expressionssystem auf Basis des *nar*-Operons bereits für homolog und heterolog exprimierte Proteine erfolgreich eingesetzt werden konnte

(Hidalgo *et al.*, 2004; Moreno *et al.*, 2005; Park *et al.*, 2004) wurden Plasmidinstabilitäten sowie unzureichende Expressionsausbeuten regelmäßig beobachtet (A. Angelov, persönliche Mitteilung).

Das Carotinoid-Cluster (Car-Cluster) in *T. thermophilus* bot sich daher als Grundlage für ein neuartiges Expressionssystem an, da es im Gegensatz zu bisher etablierten Systemen nicht auf die Zugabe von chemischen Induktoren angewiesen ist, sondern mittels unterschiedlicher Belichtungsstärken dynamisch titrierbar ist. Außerdem wurde die Funktion des Carotinoid-Clusters in *T. thermophilus* bereits intensiv untersucht (Ortiz-Guerrero *et al.*, 2011; Takano *et al.*, 2011). Die beteiligten Elemente des *crt*-Clusters liegen auf dem Megaplasmid pTT27 und konzentrieren sich dabei auf den genetischen Locus mit den Genen TT\_P0055 bis TT\_P0058 (siehe Abbildung 45). Der zentrale Repressor LitR wird durch ORF TT\_P0056 kodiert und ist als CarA/LitR-Regulator annotiert (*light-induced transcriptional regulator*). Dieser Liganden-abhängige Repressor gehört zur MerR-Proteinfamilie (Hobman, 2007) und inhibiert die Transkription der beiden Promotoren  $P_{litR}$  und  $P_{crtB}$ . Als Cofaktor dient 5'-Deoxyadenosylcobalamin (Ado-B<sub>12</sub>), welches sich im C-terminalen Teil des Proteins anlagert und zu Konformationsänderungen führt. Als homotetramerer Komplex ist der Repressor aktiv und inaktiviert die Expression des Phytoen-Synthasegens (*crtB*), welches für ein Schlüsselenzym der Carotinoid-Biosynthese kodiert (Hoshino *et al.*, 1993). Deletionsmutanten von LitR ( $\Delta$ TT\_P0056) sind gegenüber der Expression von *crtB* dereprimiert und zeigen auffällig dunkelorange gefärbte Kolonien selbst in der Dunkelheit (A. Michaelis, Bachelorthesis, 2011). Die Überexpression der Phytoen-Synthase führt zur Akkumulation von Thermo(bis)zeaxanthinen, welches zu einem rötlichen Phänotyp in *T. thermophilus* HB27 führen kann (Fujita *et al.*, 2013). Bei Einwirkung von Licht absorbiert das 5'-Deoxyadenosylcobalamin im blauen Wellenlängenbereich. Durch Photolyse des Repressor-gebundenen Cobalamins wird die tetramere Konformation des Repressors unterbrochen. Der positive Aktivator P55 (kodiert durch ORF TT\_P0055) ist ein annotierter CRP/FNR Transkriptionsregulator, der in der kotranskribierten Region mit *litR* liegt und im intergenischen Promotorbereich zwischen  $P_{litR}$  und  $P_{crtB}$  bindet. Die Deletion des ORFs TT\_P0055 führt zur Deaktivierung der Carotinoidbildung, was sich im weißlichen Phänotyp dieser *Thermus*-Mutanten widerspiegelt (A. Angelov, unveröffentlichte Ergebnisse).

Im Vorfeld der Expressions-Versuche wurde in *T. thermophilus* ein Reporterstamm namens TL01 konstruiert, der anstelle des Phytoen-Synthasegens (TT\_P0057) ein Gen für  $\beta$ -Glucosidase (*Bgl*) unter der Kontrolle des  $P_{crtB}$ -Promotors enthält (Genotyp  $\Delta bgl$ ,  $\Delta crtB::bgl$ ), um den Einfluss der Belichtung im Miller-Assay quantifizierbar zu machen (T. Lichev, Bachelorthesis, 2011). Das modifizierte Car-Cluster mit Reporter gen im TL01-Stamm diente als Ausgangssequenz für die Konstruktion der Expressionvektoren für heterologe Proteinexpression in *T. thermophilus* HB27. Dabei wurden zwei Strategien gewählt: Die Erstellung eines replikativen Plasmids auf Basis des pMK18-Vektors mit hoher Kopienzahl (pMK18-car01 und pMK18-car02) sowie die Klonierung eines *Suicide*-Vektors zur Integration in den *crtB*-Locus des Megaplasmid pTT27 (pCar01 und pCar02).

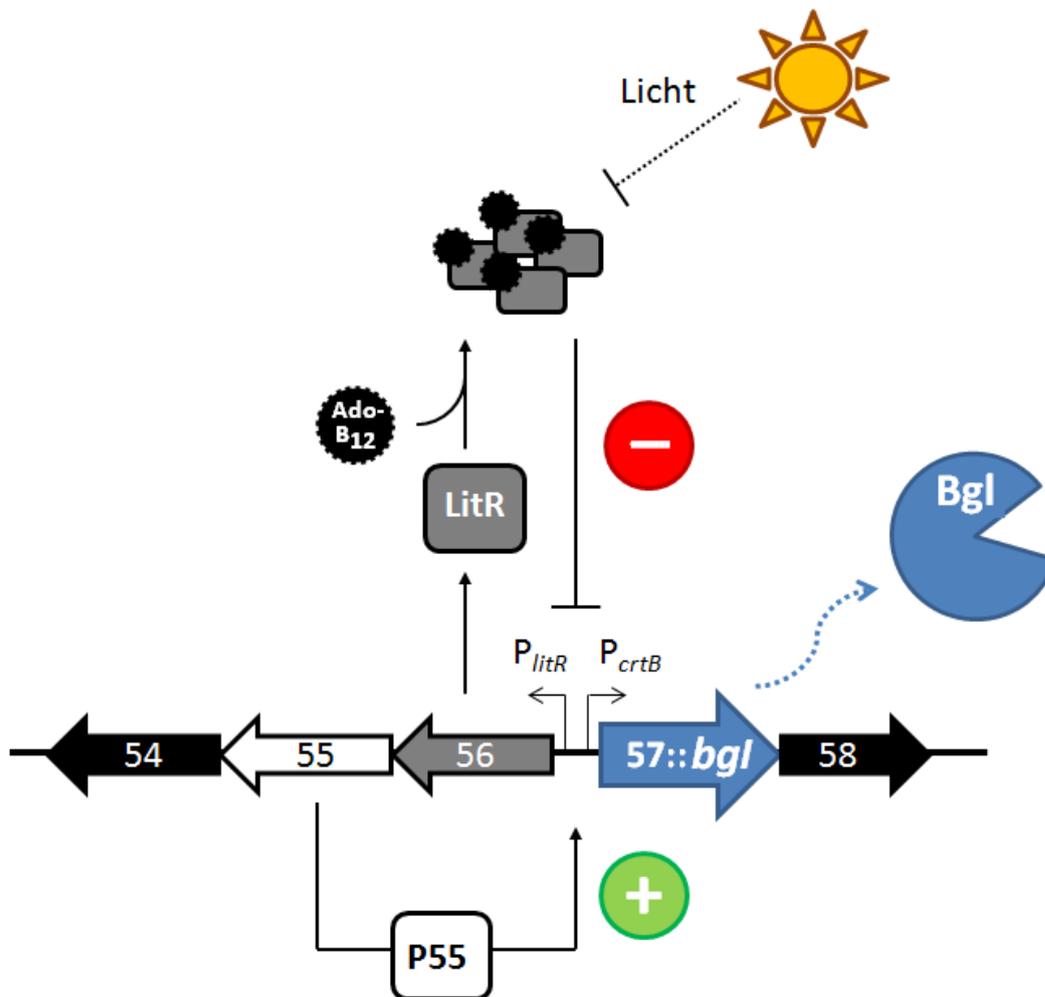


Abbildung 45: Schematische Abbildung der Regulation an der Promotorregion des *crtB*-Genclusters. TT\_P0056 kodiert für den Repressor LitR, der 5'-Deoxyadenosylcobalamin (abgekürzt Ado-B<sub>12</sub>) bindet und sich zu einem Tetramer assembliert. Als Proteinkomplex ist der Repressor in der Dunkelheit aktiv und kann die Transkription an den beiden Promotoren  $P_{litR}$  und  $P_{crtB}$  unterbinden. Belichtung führt zur Photolyse des Ado-B<sub>12</sub> und zum Zerfall des LitR Proteinkomplexes und damit zur Inaktivierung des bidirektionalen Repressors. Das Genprodukt von TT\_P0055 fungiert als Aktivator der beiden Promotoren, sodass es zur Transkription nachgeschalteter Gene kommt. In *T. thermophilus* WT entspricht TT\_P0057 der Phytoen-Synthase (CrtB), im Reporterstamm TL01 (A. Lichev, Bachelorthesis 2011) wurde der ORF durch eine  $\beta$ -Glucosidase ( $\Delta 57::bgl$ ) ersetzt (modifiziert nach Takano *et al.*, 2011 und Ortiz-Guerrero *et al.*, 2011).

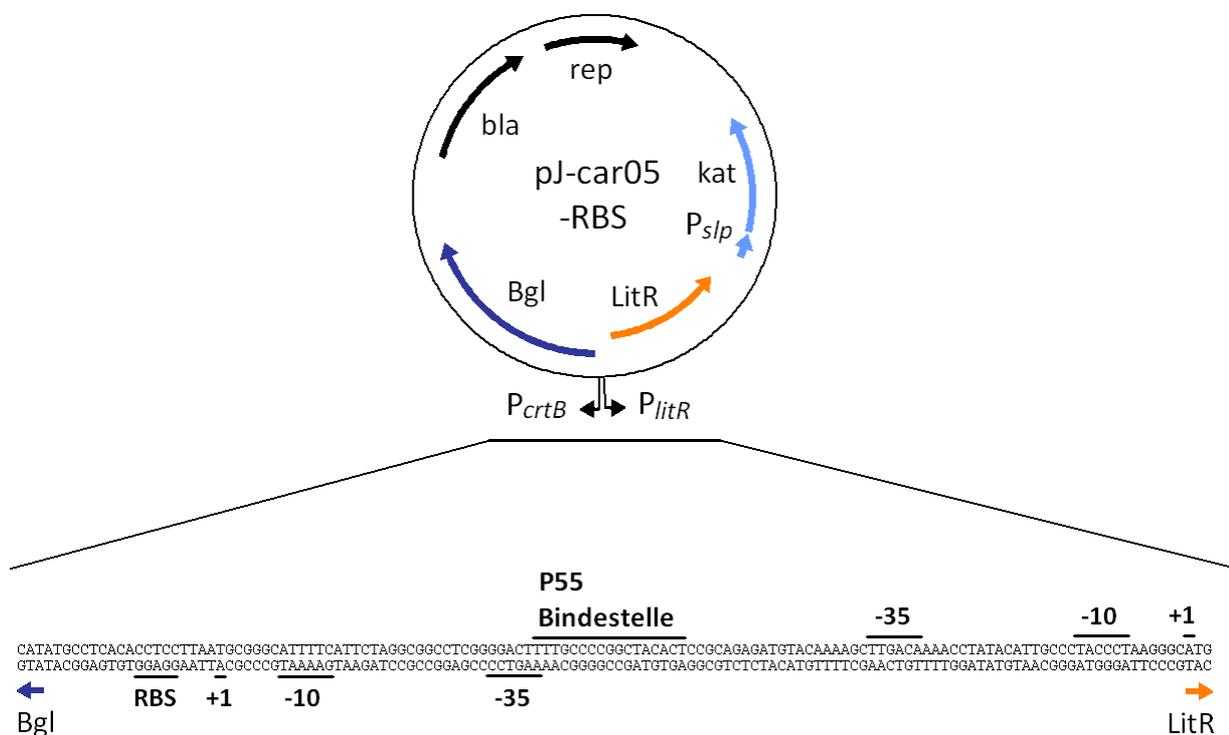
Alle klonierten Expressionsvektoren konnten in zwei Varianten vor und nach dem konstitutiven  $P_{slp}$ -Promotor kloniert werden, um den Einfluss polarer Effekte der erhaltenen Expressionskonstrukte auf die Basalaktivität (ohne Belichtung) und Induzierbarkeit hin zu untersuchen. Für die beiden pMK18-basierten replikativen Vektoren konnten nach erfolgter Transformation von *Thermus* selbst ohne Lichtexposition sehr hohe Aktivitäten gemessen werden, welche gegenüber dem Kontrollstamm  $\Delta car$  mit pMK18 Leervektor um den Faktor 8 (pMK18-car01) bis Faktor 20 (pMK18-car02) erhöht waren. Die Zunahme der Bgl-Aktivität durch gezielte Induktion mittels Belichtung bei 2-3 kLux war für beide Varianten vergleichbar niedrig. Bei den integrativen Expressionsvektoren pCar01 und pCar02 zeichnete sich ein ähnliches Bild ab, obwohl die basalen Aktivitäten ohne Lichtinduktion

insgesamt geringer ausfielen als bei den replikativen Vektoren. Das geeignetste Expressionskonstrukt war pCar02 mit einer mehr als 3-fachen Induktion der Bgl-Aktivität durch Belichtung, bei vergleichsweise niedriger Aktivität im Dunkeln. Der Reporterstamm TL01 zeigte ähnlich geringe Grundaktivität, jedoch mit nur etwa 2,1-facher Induktion durch Belichtung eine nachweislich geringere Aktivitätszunahme im durchgeführten Miller-Assay.

Bei allen getesteten Car-Expressionsvektorkonstrukten konnte eine relativ hohe Hintergrundaktivität ohne erfolgte Licht-Induktion festgestellt werden. Der Präsenz des *slp*-Promotors war in allen Fällen nachteilig und führte zu signifikant erhöhten basalen Aktivitäten der  $\beta$ -Glucosidase ohne Belichtung. Diese gehen womöglich auf polare Effekte durch den starken und konstitutiv exprimierten  $P_{slp}$ -Promotor zurück, was zu einer erhöhten Transkription der nachfolgenden (*downstream*) und kotranskribierten Gene und Gencluster auf dem Vektor führen kann. Dies könnte die notwendige Repression des  $P_{crtB}$ -Promotors unterbrechen, da die Expression des *litR*-Gens gestört ist. Die beobachtete gesteigerte  $\beta$ -Glucosidase-Aktivität ist dann auch nicht mehr länger vom positiven Regulator P55 abhängig, sondern obliegt womöglich der Präsenz eines weiteren vorgeschalteten und ausreichend starken Promotors. Das zeigt sich auch bei der Konstellation durch den replikativen Vektor pMK18-car02, welcher noch weit höhere basale Aktivitäten ohne Belichtung aufweist, obwohl das *crtB*-Cluster vor dem *slp*-Promotor lokalisiert ist. Durch die Deletion der *repA*-Sequenz (RepA ist für die Replikationsfähigkeit des pMK18-Vektors verantwortlich, de Grado *et al.*, 1998) konnten die integrativen Expressionsvektoren pCar01 und pCar02 konstruiert werden. Die Integration in den Locus des *crtB*-Clusters führt im Vergleich zum replikativen Plasmid zu deutlich niedriger Aktivität ohne Belichtung. Dies könnte durch die erhöhte Kopienzahl des Replikons erklärt werden, welche für den replikativen pMK18-Vektor höher ist (4 bis 10 Kopien pro Zelle) (de Grado *et al.*, 1999), während für das *Thermus* Chromosom und Megaplasmid eine Kopienzahl aufgrund von Polyploidie im Bereich von 4 bis 5 Kopien abgeschätzt wurde (Ohtani *et al.*, 2010). Dies könnte außerdem dazu führen, dass die notwendige Konzentration an Repressor LitR sowie des Cofaktors 5'-Deoxyadenosylcobalamin (Ado-B<sub>12</sub>) zu niedrig ist, um die Repression des  $P_{crtB}$ -Promotors im Dunkeln sicherzustellen. Versuche zur Supplementation von L-Glutamat als einer der wichtigsten Vorstufe-Moleküle zur Biosynthese des Cofaktors Ado-B<sub>12</sub> zeigten, dass ab Konzentrationen von 0,1 % (w/v) L-Glutamat bereits etwa 10 %, bei 1,0 % (w/v) etwa 50 % der Hintergrundaktivität des *T. thermophilus*  $\Delta car$ -Stammes mit pMK18-car01 Expressionsplasmid im Dunkeln abgeschwächt werden konnte. Dies weist darauf hin, dass einerseits die Bildung von Ado-B<sub>12</sub> für die Repression des Promotors im Dunkeln eine wichtige Rolle spielt, andererseits aber vom Verhältnis der Repressormoleküle zum Promotor abhängt.

Auf Basis der Daten, die sich aus dem integrativen Vektor pCar02 ableiten ließen, wurde schließlich ein neuer Expressionsvektor konstruiert (Abbildung 46). Der ebenfalls integrative Vektor pJ-car05-RBS (F. Terzi, Masterthesis, 2012) besteht aus einem pJet-Vektorrückgrat (CloneJet, Fermentas) mit thermostabiler Kanamycin-Resistenz, *slp*-Promotor und der Sequenz von *litR*

(TT\_P0056) sowie den bidirektionalen Promotoren  $P_{litR}$  und  $P_{crtB}$ .  $CrtB$  wurde durch das Reportergen  $bgl$  (TT\_P0042) ersetzt, das nun unter dem Einfluss des  $crtB$ -Promotors steht. Zusätzlich wurde eine ribosomale Bindestelle (RBS) aus dem pMK18-Vektor eingefügt (Sequenz GGAGGTGTGAGGCAT). Nach Transformation von *T. thermophilus*-Stamm  $\Delta bgl$  ist der Vektor in der Lage, in den Locus des  $crtB$ -Operons im Megaplasmid pTT27 zu integrieren und führt bei Belichtung der Zellsuspension zur Expression von  $\beta$ -Glucosidase, was unter optimalen Inkubationsbedingungen eine 6 bis 7-fache Induktion der Bgl-Aktivität zur Folge hatte (F. Terzi, Masterthesis 2012).



**Abbildung 46: Schematischer Aufbau des pJ-car05-RBS Vektors und Sequenz der beiden synthetischen Promotoren  $P_{litR}$  und  $P_{crtB}$  vor dem Zielprotein der Expression ( $Bgl$ ). Die annotierten Sequenzen zeigen die  $-10$ ,  $-35$  Region der jeweiligen Promotoren sowie die Transkriptionsstarts (*transcriptional start site*, abgekürzt als  $+1$ ), die  $P55$ -Aktivator Bindestelle und ribosomale Bindestelle (RBS) (Terzi, Masterthesis, 2012; Takano *et al.*, 2011).**

Die Anwendung des neuen Expressionssystems auf Basis des pJ-car05-RBS Vektors im Vergleich zu den etablierten Expressionssystemen in *Thermus* wurde beispielhaft für zwei metagenomische Esterasen (EstA2 und EstB1) gezeigt, die heterolog exprimiert worden sind (Tabelle 33). Im direkten Vergleich zeigt sich, dass das etablierte Expressionssystem auf Basis des HB27::*nar*-Stammes mit pMKE2-Vektor keine signifikante Zunahme der Aktivität zur Folge hatte, die selbst bei Inkubationen über Nacht keine nachweisbare Expression lieferte. Ferner war eine leichte Aktivitätssteigerung im Falle des pMK18 Vektors mit dem Zielgen nach der  $kat$ -Kassette sowie zusätzlicher RBS zu

verzeichnen, die nach 14 Stunden gemessen werden konnte. Das neu etablierte Expressionssystem auf Basis des Car-Clusters hingegen zeigte bereits nach 4 Stunden Belichtung eine bedeutsame Zunahme der lipolytischen Aktivitäten.

**Tabelle 33: Vergleich der in *T. thermophilus* verfügbaren Expressionssysteme mit pJ-car05-RBS.**

Expressions- system	Induktionsart, -Bedingungen	<i>T. thermophilus</i> - Stamm	Lipolytische Aktivitäts- zunahme ( $\text{mU} \times \text{mg}^{-1}$ )*	
			EstA2	EstB1
pMKE2	40 mM KNO <sub>3</sub> , kein Schütteln, 60 °C, 4 – 14 h Inkubation	HB27:: <i>nar</i>	-	-
pMK18_RBS	Nicht induzierbarer, konstitutiver Promotor, Schütteln bei 60 °C für 14 h	BL03	47,2 (17,5 %)	62,0 (22,9 %)
pJ-car05-RBS	Lichtinduktion 2-3 kLux, Schütteln bei 60 °C für 4 h	BL03	150,1 (59,9 %)	237,6 (106,9 %)

\*: Angegeben sind die Zunahmen der intrazellulären Esterase-Aktivität ( $\text{mU} \times \text{mg}^{-1}$  Rohextrakt, 85 °C, pH 8,8) gegenüber der Leervektor-Kontrolle auf pNP-C<sub>4</sub> unter Induktionsbedingungen (Hintergrundaktivitäten ohne Induktion wurden subtrahiert). In Klammern darunter ist die prozentuale Zunahme durch Induktion angegeben.

Die Nachteile eines anaeroben bzw. microaerophilen pMKE2-Expressionssystems in *T. thermophilus* HB27 sind zahlreich: Die Limitierung von Sauerstoff im Kulturmedium, v.a. bei größeren Volumina, verhindert optimales Wachstum von *T. thermophilus* und macht zusätzliche Begasung für höhere Zelldichten der Kulturen notwendig (Demirtas *et al.*, 2003). Zum Anaerobisieren der Kultur wurden in dieser Arbeit die Kulturröhrchen mit Schraubkappen-Verschlüssen abgedichtet und nach Zugabe des Induktors nicht weiter geschüttelt. Diese technisch einfachen Maßnahmen haben sich bei bisherigen Versuchen mit dem pMKE2-System bewährt (A. Angelov, persönliche Mitteilung). Da der P<sub>nar</sub>-Promotor lediglich unter Sauerstoff-armen Bedingungen aktivierbar ist, muss anstelle von Sauerstoff Kaliumnitrat als terminaler Elektronenakzeptor der Nitratatmung fungieren, was jedoch energetisch der aeroben Atmung gegenüber beungünstigt ist (Thauer *et al.*, 1977) und daher die effiziente Proteinexpression ebenfalls benachteiligen kann. Im Gegensatz zum anaeroben Expressionssystem war die Expression mit pMK18\_RBS Vektor unter konstitutiven Bedingungen erfolgreich. Das licht-induzierbare *car*-Expressionssystem führte jedoch zu noch höherer lipolytischer Aktivität.

Zwar bietet das neue integrative Expressionssystem auf Basis des Car-Clusters noch eine Reihe weiterer Vorteile (A. Angelov, persönliche Mitteilung), die jedoch im Zuge der vorliegenden Arbeit nicht weiter untersucht werden konnten und daher nur kurz beschrieben werden: (i) die Möglichkeit,

die Expressionsstärke über die Belichtung einzustellen, erlaubt eine Feinjustierung und damit eine gezieltere Titration der Promotorstärke. Das System würde die nachträgliche Anpassung des Expressionsgrads über die Dauer der Inkubation ermöglichen. Beispielsweise könnte eine Verringerung der Expressionsstärke zu einer Verbesserung der Ausbeuten löslichen Zielproteins führen, indem weniger *inclusion bodies* auftreten (Terpe, 2006). Bei herkömmlichen Induktorlösungen in Batch-Fermentation ist dies nicht mehr möglich, da die einmal supplementierte Menge (Konzentration) nicht mehr nachträglich verändert werden kann. (ii) Die Mobilisierung des Expressionsvektors über homologe Rekombination am Insertionsort erlaubt die gezielte Einstellung der Kopienzahl des Vektors im Chromosom, indem die Antibiotika-Konzentration variiert wird. Dabei gilt, dass höhere Antibiotika-Konzentrationen zu vermehrten Duplikationen des identischen Plasmids und damit zu einer Erhöhung der Gendosis im Wirt führen können (Tyo *et al.*, 2009). (iii) Da das Expressionssystem keine vorherige Modifikation des Wirts in Anspruch nimmt, wäre es daher in vielen *T. thermophilus*-Stämmen sofort anwendbar. Als nachteilig würde sich ein Belichtungsabhängiges Expressionssystem in einer Hochdichtefermentation erweisen, da eine Lichtquelle in Suspensionen mit hohen optischen Dichten womöglich nur einen geringen Teil der Kultur erreichen würde.

Als Alternative zum licht-induzierbaren *crtB*-Promotor existieren in *T. thermophilus* noch weitere Systeme auf Basis von regulierbaren Repressoren, welche für die Konstruktion eines neuartigen Expressionssystems in dem thermophilen Bakterium in Frage kommen könnten. Beispiele sind der transkriptionelle Repressor FadR des Fettsäureabbaus (Agari *et al.*, 2011) sowie der transkriptionelle Repressor CsoR, welcher durch die Präsenz von Kupfer gesteuert werden kann (Sakamoto *et al.*, 2010). Als Beispiel einer Zell-freien Expression konnte die Proteintranslation von *T. thermophilus* mit einem minimalen Set aus aufgereinigten Ribosomen, Transfer-Ribonukleinsäuren (tRNAs) und weniger als 3 Dutzend rekombinanten Proteinen *in vitro* rekonstruiert werden (Zhou *et al.*, 2012). Das System konnte nachweislich aktive Proteine (60 µg/mL) über einen weiten Temperaturbereich (37 bis 65 °C) synthetisieren. Interessanterweise liegt die minimale als auch die optimale Wachstumstemperatur von *T. thermophilus* deutlich über 37 °C.

## 1.2. Erstellung einer Transposonmutagenese-Bank in *T. thermophilus* HB27

Die Transposonmutagenese ermöglicht die Identifizierung von *a priori* unbekanntem genetischen Elementen, die für bestimmte Eigenschaften der Zelle verantwortlich sind. Da für *T. thermophilus* als thermophilen Modellorganismus bisher noch keine Transposonmutagenese-Bank existiert, sollte diese mit Hilfe einer *in vivo* oder einer *in vitro*-Strategie etabliert werden. Prinzipiell sollte das Transposon mit einer relativ hohen Häufigkeit transponiert werden, es sollte wenig selektiv für die Zielsequenz sein, es sollte möglichst einfach zu selektierende Gene (z.B. Antibiotikaresistenzgene) tragen und es

sollte in möglichst vielen verschiedenen Bakterienarten aktiv sein (Snyder und Champness, 2007). Eine Reihe etablierter und kommerziell erhältlicher Transposonmutagenese-Strategien sind mittlerweile verfügbar, dennoch wurde für die *in vivo* Mutagenese auf das in *Bacillus* etablierte System pMarB (Le Breton *et al.*, 2006) zurückgegriffen, welches auf dem *mariner* Transposon *Himar1* der Hornfliege *Haematobia irritans* beruht (Robertson und Lampe, 1995). Das Transposon TnYLB-1 ist bei erhöhten Temperaturen selektierbar und konnte bereits in *B. subtilis* erfolgreich angewendet werden. Nach entsprechenden Modifikationen mit einer *Thermus*-kompatiblen Kanamycin-Kassette und Ausrichtung der Transposase unter dem *slp*-Promotor, konnten mehrere Dutzend Kolonien pro µg übertragenem pTnB2-*Suicide*-Plasmid erhalten werden, während für den Kontrollvektor pTnC2 ohne Transposase keine Kolonien resultierten. Die Analyse der gewachsenen Kolonien mittels PCR konnte die Präsenz des Transposons TnB2 nachweisen, während die Transposase nicht amplifiziert werden konnte. Obwohl die PCR-basierten Nachweise auf die erfolgreiche Insertierung des Transposons und nicht des vollständigen Vektors hindeuteten, zeigte die Southern Blot Analyse an, dass das Transposon bei allen Klonen am identischen Insertionsort im Chromosom vorhanden war (Abbildung 8). Als mögliche Erklärung ließe sich die Insertion des Plasmids über die Homologien der jeweiligen *slp*-Promotorsequenzen des Vektors und des Chromosoms erklären, die über 200 Basen zu 97,5 % identisch zueinander sind. Da die *Himar1*-Transposase mit 53 bp langen invertierten Sequenzen flankiert ist, würde ihre Exzision mittels Rückrekombination erklären, warum sie nicht per PCR amplifizierbar war. Das zu erwartende Verdau-Fragment der chromosomalen DNA mittels *Bam*HI ergäbe im Southern Blot eine etwa 10,5 kb große Bande im Fall der Insertion des Transposon-tragenden Plasmids am Locus des *slp*-Promotors. Letztlich führte die *in vivo* Transposonmutagenese auf Basis von TnB2 also nicht zu einer zufälligen Insertion in das Chromosom und konnte aus diesem Grund nicht weiter verwendet werden.

Da die zufällige Insertion mittels *in vivo* Mutagenese gescheitert ist, wurde die *in vitro* Transposition mit der hyperaktiven EZ-Tn5 Transposase (Epicentre) durchgeführt. Bisherige Versuche waren wenig Erfolg versprechend, da die Ausbeute aus einer Transposonreaktion nur wenige Transposonmutanten in *Thermus* lieferte (A. Angelov, persönliche Mitteilung). Die in dieser Arbeit erwähnte Strategie zur Erstellung von zufälligen Mutanten steigerte die Ausbeuten um etwa 2 Größenordnungen gegenüber bisherigen Mutageneseversuchen. Die Ursachen hierfür werden jedoch derzeit noch untersucht (nicht veröffentlichte Daten, A. Angelov).

Nach Übertragung der Transposon-Reaktionen wurden hunderte *Thermus*-Transformanten auf TB-Festmedien mit Kanamycin selektiert, die für Transposon-Mutanten übliche phänotypische Auffälligkeiten zeigten (z.B. weißliche bis tief orange Kolonien, unterschiedliche Koloniegröße und Morphologie). Mehrere zufällig selektierte Mutanten wurden mittels Southern Blot analog zu oben analysiert (Abbildung 9). Es konnte gezeigt werden, dass einfache sowie Doppelinsertionsmutanten vorhanden waren. Die erstellte Mutagenese-Bank umfasst etwa 8.000 Einzelkolonien. Unter

Berücksichtigung der Genomgröße von 2,13 Mb und der vorhergesagten 2.218 ORFs im *T. thermophilus* HB27-Genom (Henne *et al.*, 2004) kann davon ausgegangen werden, dass theoretisch etwa 97,2 % aller vorhandenen Gene mutagenisiert werden konnten (Berechnungsgrundlage, siehe Formel in Kapitel III.6., Gleichung 3). Da diese vereinfachten Berechnungen die Mutagenese essentieller Gene ausblendet und die mögliche Beteiligung an Doppelinsertionsmutanten unterschlägt, dürfte der Anteil von Transposoninsertionen in nicht-essentiellen Genen und intergenetischen Bereichen noch höher liegen.

Die Suche nach Mutanten mit verminderter extrazellulärer Tributyraseaktivität lieferte unter allen getesteten Kolonien keinerlei phänotypisch-auffällige Kandidaten. Dieses Ergebnis weist darauf hin, dass es wahrscheinlich nicht möglich ist, eine Esterase-defiziente Mutante über ungerichtete Mutagenesen zu erzeugen. Dies kann mehrere Gründe haben: Einerseits reicht eine einzelne Mutation nicht aus, um die extrazellulären lipolytischen Aktivitäten vollständig zu deletieren. Womöglich sind mehrere Genprodukte an der gesamten lipolytischen Aktivität beteiligt (die Mutationsanalyse der Mehrfachmutante BL03 bestätigt diese Vermutung, siehe Kapitel V.1.3.). Andererseits könnten am Abbau von Triacylglyceriden beteiligte Enzyme und zentrale Regulatoren nicht deletiert werden, da sie vermutlich essentiell für die Zellproliferation sind. Darüberhinaus ist es auch möglich, dass die entsprechenden (Einzel-)Gene nicht durch die Transposonbank abgedeckt werden konnten. In einer anderen Studie konnten mit Hilfe der in dieser Arbeit erstellten Transposonmutagenese-Bank erfolgreich Mutanten isoliert werden, die nicht länger gegenüber 5-Bromo-4-Chloro-3-indoxyl (BCI)-Substraten empfindlich sind um somit den Einfluss der sonst toxischen Abbauprodukte dieser Indigosubstrate gegenüber Mikroorganismen zu untersuchen (Angelov *et al.*, 2013). Dabei zeigten drei verschiedene *Thermus*-Mutanten Resistenz gegenüber BCI- $\beta$ -D-Glucopyranosid-Spaltungsprodukten. Ihre Insertionsstellen konnten mittels Vektorette PCR (Arnold und Hodgson, 1991) überprüft werden. Alle dieser Mutanten trugen Mutationen in Genclustern der Zuckerverwertung und konnten auf mögliche Zucker-Transporter zurückgeführt werden (TT\_C0611 bis TT\_C0613, TT\_P0223 und TT\_P0039 bis TT\_P0041). Die Ergebnisse lassen darauf schließen, dass der toxische Effekt der BCI-Substrate erst durch ihren Transport in die Zelle vermittelt wird. Die Mutanten konnten offenbar die Zuckerverbindungen nicht länger aufnehmen und waren daher gegenüber den Substraten resistent geworden (Angelov *et al.*, 2013).

Diese Mutagenesebank ist der erste erfolgreiche Versuch, eine zufällige und ungerichtete Mutagenese in einem thermophilen Bakterium wie *T. thermophilus* HB27 durchzuführen. Als Modellorganismus für *structural genomics* sowie natürliche Kompetenz bietet diese Bank außerdem die Möglichkeit, neue und bisher unbekannte Gene und kodierte Proteine zu identifizieren, welche für bestimmte strukturelle oder Stoffwechselfunktionen im Organismus notwendig sind. Die Sequenzierung aller erhaltenen Transposon-Mutanten könnte ebenfalls zur Aufklärung des minimalen bzw. Kerngenoms

dienen, welche Gene für *T. thermophilus* HB27 essentiell sind. Diese Strategie wurde beispielsweise für das Genom von *Mycoplasma genitalium* angewandt (Glass *et al.*, 2006).

### 1.3. Deaktivierung der lipolytischen Hintergrundaktivität in *Thermus* und Herstellung einer Multideletionsmutante BL03 als Expressionswirt für Genbanken

Eine Vielzahl an Bakterien ist zum Abbau von Substraten mittels sekretierten lipolytischen Enzymen befähigt. *T. thermophilus* ist bekannt für die Ausbildung von extrazellulären, Zell-gebundenen und intrazellulären Esterase- und Lipase-Aktivitäten (Berger *et al.*, 1995; Domínguez *et al.*, 2004; Fuciños *et al.*, 2005a,b; Sigurgísladóttir *et al.*, 1993), welche bereits für biotechnologische Anwendungen studiert wurden. Die Expression von Esterasen und Lipasen hängt in ganz besonderem Maße von den Kultivierungsbedingungen ab (kontinuierliche Kultur, Belüftungsrate, Präsenz bestimmter Salze und mineralhaltigen Wassers sowie Lipidsupplementation zur selektiven Induktion der Expression) (Deive *et al.*, 2009; Domínguez *et al.*, 2007, 2010). Im Stamm HB27 konnten zwei extrazelluläre Proteine mit einer Größe von 34 und 62 kDa von Fuciños und Kollegen bereits beschrieben und vorläufig charakterisiert werden (Fuciños *et al.*, 2005a). ORF TT\_C0904 kodiert für eine dieser beiden Esterasen (YP\_004875) und wurde bereits ausführlich biochemisch untersucht (Fuciños *et al.*, 2011). Aus *Thermus scotoductus* SA-01 wurde eine thermostabile Esterase EstTs1 (Protein Accession Nummer ACS36170) heterolog in *E. coli* exprimiert und charakterisiert, welche ortholog zu dem Operon von ORFs TT\_C0340 bis TT\_C0341 in *T. thermophilus* HB27 ist (du Plessis *et al.*, 2010). Sieben weitere potentielle Esterase-kodierende Gene konnten mittels einer *hmmScan* Proteomanalyse aus der Pfam-A Datenbank identifiziert werden (Tabelle 22). Dabei waren die Kandidaten für typische lipolytische Proteinfamilien signifikant, wie beispielsweise etwa Carboxylesterasen,  $\alpha/\beta$ -Hydrolasen, Thioesterasen oder Phospholipasen. Alle vorhergesagten Kandidaten konnten mittels gezielter Vektorintegration unterbrochen werden, der Nachweis der erfolgten genetischen Manipulationen wurde durch analytische PCR verifiziert (Art der Integration, homozygoter Status des polyploiden Chromosoms) (siehe Anhang, Tabelle 40). Dennoch konnten nur nach Inaktivierung der beiden Kandidatengene TT\_C1787 und TT\_C0904 erhebliche Aktivitätsabnahmen verzeichnet werden, welche durch Kanamycin-Austauschmutanten ( $\Delta$ TT\_C0904::kat und  $\Delta$ TT\_C1787::kat) bestätigt werden konnten. Die Gendisruptions-Mutanten zeichneten sich durch die Bildung von kleineren hydrolytischen Höfen auf Indikatormedien mit Tributyrin als Substrat bzw. reduzierten lipolytischen Aktivitäten in den *pNP*-Tests der Überstände und intrazellulären/Zell-gebundenen Fraktionen gegenüber dem unmanipuliertem Wildtyp *T. thermophilus* HB27 aus.

Den größten Einfluss auf die Esterase-Aktivität hatte die Unterbrechung von ORF TT\_C0904, der für eine annotierte Pektinacetylerase kodiert (YP\_004875.1). Dieses extrazelluläre Enzym war vor

allem im Bereich mittellanger Fettsäuren aktiv. Die Beobachtung wird unterstützt durch die Daten von Fuciños und Kollegen (Fuciños *et al.*, 2011), die das heterolog exprimierte Enzym biochemisch charakterisierten und eine Substratpräferenz gegenüber *para*-Nitrophenyl-Caprat (*p*NP-C<sub>10</sub>) beobachten konnten.

Die Disruption des Gens für den zweiten Enzymkandidaten, eine Carboxylesterase (YP\_005756.1) führte ebenfalls zu deutlich reduzierter Hofbildung auf Substratplatten. Interessanterweise beeinflusste die Inaktivierung des ORFs TT\_C1787 weniger die extrazelluläre als die intrazelluläre/Zell-gebundene Fraktion der *Thermus*-Kulturen. Die *in silico* Analyse des Zielproteins am N-Terminus sagte eine äußerst kurze Transmembrandomäne bzw. die Präsenz einer 17 Aminosäure langen Signalpeptidsequenz vorher, die jedoch beide nicht signifikant sind. Aus den Aktivitätsdaten mit den *p*NP-Substraten zeigte sich ein Aktivitätsverlust, der spezifisch für die cytoplasmatische oder Zell-gebundene Lokalisierung war, und mit Zunahme der Acyl-Kettenlänge der Fettsäuren zunahm. Dies weist darauf hin, dass die Carboxylesterase vom Aktivitätsspektrum eher einer Lipase als einer Esterase zugeordnet werden kann, die länger-kettige Substrate mit mehr als 10 Kohlenstoffatomen bevorzugen (Jaeger *et al.*, 1999). Die beobachteten Hydrolysehöfe auf Tributyrinplatten sind daher weniger auf die Aktivität der rein sekretierten lipolytischen Enzyme zurückzuführen, sondern setzen sich aus sowohl sekretierten als auch intrazellulären Lipasen und Esterasen zusammen. Die Lyse von Zellen bei der Ausbildung von Kolonien auf der Agaroberfläche dürfte hierfür verantwortlich sein.

Die heterolog exprimierte Esterase EstTs1 aus *T. scotoeductus* SA-01 zeigte starke Ähnlichkeit zum Genprodukt von ORF TT\_C0341 (du Plessis *et al.*, 2010), einer  $\alpha/\beta$ -Hydrolase der Familie 6 (Accession YP\_004316.1). Da das Zielgen in einem Operon mit TT\_C0340 lokalisiert vorliegt, das ebenfalls zu einer  $\alpha/\beta$ -Hydrolase der Familie 6 gehört, wurden beide Gene vollständig mit einer Kanamycin-Kassette ersetzt. Dennoch zeigten sich in *Thermus* auf Indikatormedien keine phänotypischen Veränderungen der entsprechenden Einfach-Mutante ( $\Delta$ TT\_C0340-1::kat). Erst nach Deletion der beiden Gene in einem Stamm mit bereits erniedrigter extrazellulärer Esterase-Aktivität (Stammbezeichnung BL01, Genotyp  $\Delta$ TT\_P0042 und  $\Delta$ TT\_C1787) war der Einfluss der Mutation sichtbar: In dem resultierenden Stamm BL02 konnte ein beachtlicher Rückgang der Esteraseaktivität mit kürzerkettigen Fettsäure-Substraten (vor allem 5-Bromo-4-Chloro-3-indolyl-butyrat und in geringerem Ausmaß auch BCI-caprylat) festgestellt werden. Auf Tributyrin-Medien war der Rückgang des hydrolytischen Hofes gegenüber der Wildtyp-Kontrolle und des Stammes BL01 ebenfalls sichtbar. Diese Ergebnisse werden durch die Daten von du Plessis und Kollegen (du Plessis *et al.*, 2010), die die zu TT\_C0341 ähnliche Esterase aus *T. scotoeductus* SA-01 biochemisch charakterisierten, gestützt. In ihrer Arbeit zeigten sie auf, dass EstTs1 höchste Substratspezifitäten gegenüber *p*-Nitrophenylbutyrat (C<sub>4</sub>) aufwies. Dass die Inaktivierung der Einzelaktivität im Wildtyp nicht sichtbar war, lässt sich dadurch erklären, dass sie durch andere lipolytische Enzyme überlagert oder kompensiert wurde. Erst nach Inaktivierung einer der beiden lipolytischen Hauptaktivitäten in

BL01 kann die Deletion eines weiteren Esterase-kodierenden ORFs nicht mehr kompensiert werden, sodass es zu einem signifikanten Abfall der Esteraseaktivität in der Doppelmutante BL02 kommen kann.

Gendeletionen ohne verbleibende Fremd-DNA wie z.B. Selektionsmarker (sogenannte *clean deletions*) ermöglichen die sequentielle Manipulation mehrerer Loci im Genom des Zielorganismus und sind außerdem frei von polaren Effekten (Einfluss eingebrachter Promotoren auf benachbarte Gene bzw. Störung der Transkription dieser Gene). Wie oben bereits beschrieben benötigt man auch bei dieser Strategie ein Plasmid mit den flankierenden *up*- und *downstream* liegenden Sequenzen des Zielgens. Der Selektionsmarker kann auf dem Plasmid vorliegen, welcher zur Integration des Vektors an einer Flanke benötigt wird (positive Selektion). Nach erfolgter Integration des Vektors in die Zielsequenz muss dann gegen die Präsenz des integrierten Vektors selektiert werden (negative Selektion). Bei der Rückrekombination des Vektors aus dem Chromosom ergeben sich schließlich zwei mögliche Szenarien: Erstens, der Vektor wird wieder vollständig aus dem Chromosom entfernt, sodass die genetische Ausgangssituation wiederhergestellt wird (keine Deletion). Im zweiten Fall rekombinieren die beiden anderen Flanken des Vektors und des Chromosoms miteinander, sodass die Sequenz zwischen den beiden Flanken (z.B. das Zielgen) zusammen mit dem Vektor desintegriert wird. Dabei entsteht die markerlose Mutante, die für weitere Deletionen verwendet werden kann. In *Thermus* existieren bereits mehrere bekannte Deletionssysteme aufbauend auf einer Doppelmutation im ribosomalen Protein S12 (*rpsLI*-Allel) welches sowohl Streptomycin-Abhängigkeit als auch Streptomycin-Resistenz vermittelt (Blas-Galindo *et al.*, 2007). Ein weiteres System basiert auf einer Uracil-auxotrophen Mutante durch Unterbrechung des Pyrimidin-Biosynthesewegs mittels Deletion der Orotat-Phosphoribosyltransferase (kodiert durch *pyrE*) (Tamakoshi *et al.*, 1999). Die Präsenz des *pyrE*-Markers im Vektor komplementiert die Mutante bei der Verwendung von Minimalmedien ohne Uracil (positive Selektion). Im zweiten Schritt wird mit 5-Fluororotsäure (5-FOA) gegen die Präsenz von *pyrE* selektioniert, da das Substrat durch dessen Genprodukt in toxisches Fluorodeoxyuridin umgewandelt wird (negative Selektion). Der entscheidendste Nachteil beider etablierten Systeme liegt in der Bildung spontaner Mutanten unter Selektionsdruck begründet, sodass die Ausbeute von Mutanten unter Nicht-Mutanten gering ausfallen kann. In dieser Arbeit wurde zunächst versucht, Deletionen auf Basis des pS18a-Vektors mit *rpsLI*-Allel unter Streptomycin-Selektion durchzuführen (Blas-Galindo *et al.*, 2007). Nach Übertragung der *Knockout*-Vektoren für ORF TT\_C1787 und TT\_C0904 konnten zwar viele Streptomycin-resistente *Thermus*-Kolonien isoliert werden, die das Plasmid über eine der beiden homologen Flanken im Chromosom integrieren konnten. Im zweiten Schritt wurden Einzelkolonien auf Medien ohne Streptomycin ausgestrichen, da laut Blas-Galindo *et al.* (2007) etwa die Hälfte aller Kolonien, die ohne die Präsenz des Antibiotikums anwachsen können, Mutanten für die Zielgene erwarten ließen. Unter den vermeintlichen Mutanten waren jedoch keine Gendeletionen nachweisbar, sodass eine andere Mutagenesestrategie notwendig wurde.

Mit Hilfe der in dieser Arbeit beschriebenen Selektionsmarker-freien Methode zur Erstellung von markerlosen, multiplen Deletionsmutanten in *T. thermophilus* HB27 (siehe Kapitel IV.1.4.2.) konnten sequenziell vier extrazelluläre Esterasekodierende Gensequenzen deletiert werden. Nach Übertragung der entsprechenden Deletionsvektoren mit etwa 1 kbp langen Genflanken konnten unter allen selektionsfrei gewachsenen Kolonien mit einer Häufigkeit von 1,0 bis 0,33 % mindestens eine Mutante identifiziert werden. Obwohl die Ursache für dieses Phänomen noch Gegenstand aktueller Forschungsarbeiten ist (H. Li, unveröffentlichte Ergebnisse), konnte dieses neuartige, markerlose und selektionsfreie Deletionssystem zur Durchführung gezielter genetischer Manipulationen in *T. thermophilus* HB27 angewendet werden.

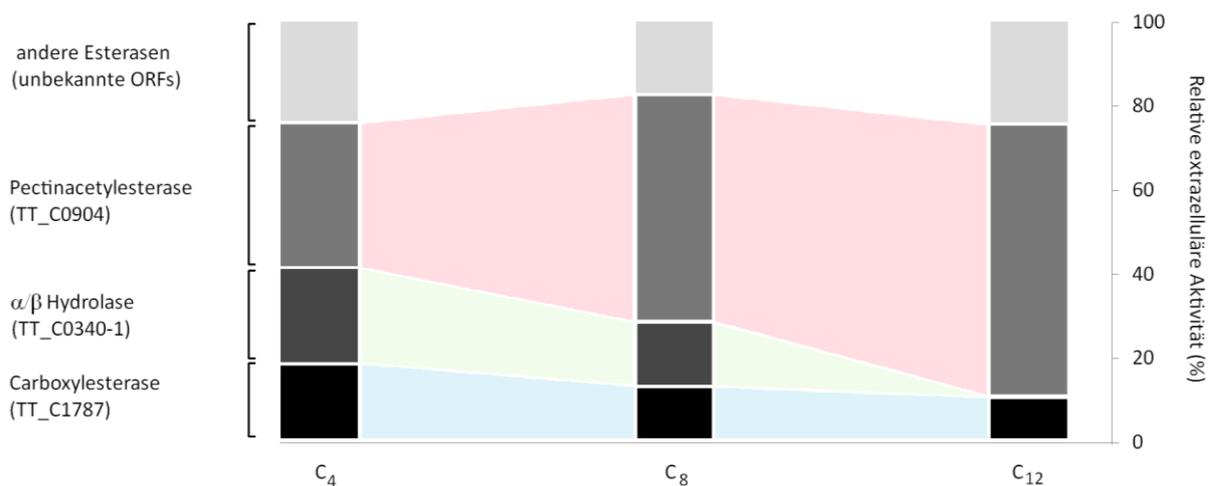
Die resultierende Vierfachmutante BL03 mit dem Genotyp  $\Delta$ TT\_P0042,  $\Delta$ TT\_C1787,  $\Delta$ TT\_0340-1 und  $\Delta$ TT\_C0904 konnte mittels analytischer PCR und Southern Blot Analyse bestätigt werden (Abbildung 14). Die Mutante zeigte gegenüber dem Wildtyp trotz der multiplen Deletionen bei reichem Nährstoffangebot (TB-Medium) keine signifikante Änderung im Wachstumsverhalten. Es konnte eine deutliche Abnahme der lipolytischen Abbaufähigkeit gegenüber *p*NP-Substraten gezeigt werden, ferner war auch eine erniedrigte Akkumulation der extrazellulären Esterasen im Medienvolumen feststellbar. Auf Tributyrin-Indikatorplatten zeigten sich graduelle Unterschiede in der Ausbildung von hydrolytischen Höfen in Abhängigkeit der eingefügten Gendeletionen (Abbildung 15).

Auf Minimalmedium (SH-Agar) mit gezielter Tributyrinsupplementation war die Vierfachmutante im Gegensatz zum Wildtyp HB27 nicht mehr in der Lage zu wachsen. Sowohl als aufgetropfte Zellsuspension, als auch aus einem Verdünnungsausstrich waren die Unterschiede deutlich sichtbar. Dies eröffnete die Möglichkeit, den Stamm BL03 erfolgreich als neuartigen Expressions- und Selektionsstamm für metagenomische Durchmusterungen im Hochdurchsatzverfahren zu verwenden, indem auf dem SH-Minimalmedium mit Tributyrin für wachstumsfähige Klone selektiert wurde, welche ein Esterase-kodierendes Gen tragen und funktionell exprimieren können.

Die Unfähigkeit des BL03-Stammes, Tributyrin abzubauen und dessen Produkte verwerten zu können liegt vermutlich an der verminderten Hydrolyse der Esterbindungen des Triacylglycerol-Moleküls. Dieser erste und grundlegende Schritt generiert freie Fettsäuren und ermöglicht deren Einspeisung in den Fettsäuremetabolismus. Durch die Freisetzung von Butyrat durch Esterase-aktive Enzyme im Medium könnte damit das Wachstum von *T. thermophilus* sichergestellt werden. Im Zuge der  $\beta$ -Oxidation werden freie Fettsäuren in Acetyl-Coenzym A (Acetyl-CoA) umgewandelt, welche in den Tricarbonsäurezyklus gespeist werden. Laut KEGG und ThYme Datenbanken verfügt *T. thermophilus* HB27 über verschiedene Fettsäure-CoA Ligasen für die Aktivierung der unterschiedlichen Fettsäuremoleküle. Dazu stehen dem Bakterium langkettige Fettsäure-CoA Ligasen (EC 6.2.1.3 TT\_C0079, TT\_C0217, TT\_C0342, TT\_C1065 und TT\_C1099) sowie entsprechende Enzyme für mittellange Fettsäuren (EC 6.2.1.- wie TT\_C0236) zur Verfügung. Gebildetes Butyryl-CoA wird

durch eine NAD<sup>+</sup>-abhängige Trans- $\Delta^2$ -Enoyl-CoA Reduktase zu Crotonyl-CoA umgesetzt, mit einer Enoyl-CoA Hydratase hydriert und zu Acetoacetyl-CoA oxidiert (3-Hydroxybutyryl-CoA Dehydrogenase), welches weiter zu Acetyl-CoA umgewandelt wird. Abhängig vom jeweiligen *Thermus*-Stamm (wie z.B. *T. aquaticus*) können Glucose und Glycerol aber auch Butyrat als Kohlenstoffquelle genutzt werden (Bergey *et al.*, 1984). In Stamm HB8 konnte die Nutzung von Palmitinsäure als alleinige Kohlenstoffquelle nachgewiesen werden (Agari *et al.*, 2011). Die Expression des Genclusters zum Fettsäureabbau (*fad*-Gene) wird dabei durch mittel- bis langkettige Acyl-CoA-Moleküle vermittelt, welche an den transkriptionellen Regulator FadR binden und seine Derepression am Promotor des Clusters herbeiführen.

Ferner konnte gezeigt werden, dass die Fähigkeit von *T. thermophilus* HB27 zum Abbau des Triacylglycerolesters Tributyrin auf der konzertierten Aktivität mehrerer lipolytischer Enzyme basiert. Die Mutationsanalyse der beteiligten Genprodukte erlaubt außerdem die Untersuchung des Einflusses eines jeden einzelnen Proteins auf die gesamte Esterase/Lipase-Aktivitäten *in vivo* (Abbildung 47).



**Abbildung 47:** Aktivitätsprofil der wichtigsten extrazellulär aktiven Esterasen und Lipasen in *T. thermophilus* HB27. Gezeigt sind drei Fettsäureester-Substrate mit verschiedenen Kettenlängen (C<sub>4</sub>: Butyrat; C<sub>8</sub>: Caprylat; C<sub>12</sub>: Laurat). Die gesamte lipolytische Aktivität setzt sich dabei aus den drei untersuchten genetischen Loci (TT\_C0904, TT\_C0340-1, TT\_C1787) sowie aus weiteren Esterasen (kodiert durch unbekannte ORFs) zusammen. Die gezeigten Werte basieren auf den experimentellen Daten von Abbildung 15. Anstelle der verbleibenden Restaktivität jeder Mutante wurde aus der Differenz der Aktivitäten zwischen den erzeugten Mutanten der Einfluss jedes Gens einzeln berechnet. Die Gesamtheit der Aktivität korrespondiert zur relativen extrazellulären Aktivität des HB27 Wildtyp-Stamms (entsprechend 100 % Gesamtaktivität). Abbildung nach Leis *et al.*, (2014).

Es konnten insgesamt drei genetische Loci festgestellt werden, welche für etwa 75 % der gesamten lipolytischen Aktivität in *T. thermophilus* verantwortlich sind. Dabei werden kurzkettige Acylester wie *p*NP-butyrat von allen kodierten Enzymen parallel hydrolysiert, während mittel- und längerkettige Esterverbindungen (C<sub>8</sub> und C<sub>12</sub>) bevorzugt von der Pektinacetylerase (kodiert durch ORF TT\_C0904) abgebaut werden. Im Vergleich zu den einzelnen Gendisruptionen ( $\Delta$ 904,  $\Delta$ 1787, siehe Abbildung 10) lassen sich gegenüber der multiplen Deletionsmutanten (BL01, BL02, BL03) deutliche

Aktivitätsunterschiede beobachten. Dies ist einerseits darauf zurückzuführen, dass sich die Bedingungen für Kultivierung und der enzymatischen Aktivitätstests (z.B. Inkubationstemperatur, Art der getesteten Substrate) zwischen den Experimenten unterscheiden. Andererseits kann aber auch der genetische Hintergrund der Mutante (An- bzw. Abwesenheit bestimmter Gene) die lipolytische Gesamtaktivität in *T. thermophilus* beeinflussen (epistatische Effekte).

Erwähnenswert ist ebenfalls die unerwartete lipolytische Aktivitätszunahme des *T. thermophilus*  $\Delta$ TT\_P0042 ( $\Delta$ *bgl*)-Stammes gegenüber dem Wildtyp, welche in etwa 25 % auf kurzketzigem BCI-butyrat ausmachte. Interessanterweise konnte in der unmittelbaren genetischen Umgebung stromabwärts des für die  $\beta$ -Glucosidase kodierenden Gens (lokalisiert auf dem Megaplasmid pTT27) ebenfalls eine  $\alpha/\beta$ -Hydrolase der Familie 6 vorhergesagt werden (TT\_P0046). Durch die Deletion der *bgl*-Sequenz könnte das für die potentielle Esterase kodierende Gen beeinflusst worden sein, sodass es zu erhöhter Expression und daher zu gesteigerter lipolytischer Aktivität gekommen sein könnte.

## **2. Identifizierung und Charakterisierung von biotechnologisch relevanten Enzymen aus diversen metagenomischen Fosmidbanken in *E. coli***

Aufgrund stetig sinkender Sequenzierkosten sowie laufender Verbesserungen im Bereich der Sequenziertechnologien der nächsten Generation (*next generation sequencing*, NGS) nimmt die Zahl der Sequenz-basierten Metagenomanalysen zu. Höhere Sequenziertiefen (*sequencing depth*) erlauben die Detektion von selbst unterrepräsentierter Spezies in mikrobiellen Konsortien, beispielsweise durch die unspezifische Amplifikation von Metagenomen (z.B. mit der  $\Phi$ 29 DNA Polymerase). Andere Sequenz-basierte Methoden ermöglichen die Identifizierung bereits bekannter Gensequenzen und funktioneller Domänen durch PCR-Amplifikation mit degenerierten Primern (Chandrasekharaiah *et al.*, 2012; Jiang *et al.*, 2006; Tekedar und Sanli-Mohamed, 2011) oder aber durch Hybridisierungsmethoden mit Sonden (z.B. Microarray, Ausschlussybridisierung) (Iwai *et al.*, 2011; Meyer *et al.*, 2007; Park *et al.*, 2008). Damit können bereits bekannte oder ähnliche Gene und Genprodukte vergleichsweise schnell aufgefunden werden, sofern sie in den zugänglichen Datenbanken hinterlegt sind. Ein großer Nachteil aller Sequenz-basierten Metagenomanalyse-Strategien liegt jedoch darin begründet, dass *a priori* unbekannte, nicht annotierte Sequenzen und Signaturen nicht identifizierbar sind, da ihre funktionelle Zuordnung in den Datenbanken fehlt (Liebl, 2011). Dieser Ansatz bleibt der funktionellen Metagenomanalyse vorbehalten. Sie zielt auf diejenigen metagenomischen Sequenzen ab, die auch einer Funktion zugeschrieben werden können. Funktionelle Strategien werden heute sehr häufig angewendet und werden auch in Zukunft zur Identifizierung neuartiger Gene und Genprodukte führen. Das liegt einerseits daran, dass immer sensitivere Detektionsmethoden verfügbar sind, die selbst geringe Mengen aktiver Enzyme

nachweisen können, andererseits ist der Großteil der in Umweltproben vorhandenen Vielfalt an Mikroorganismen bei weitem noch nicht abzuschätzen.

Bis heute ist *E. coli* der am häufigsten verwendete Wirt für die Erstellung und funktionelle Durchmusterung von (meta)genomischen Genbanken. Obwohl die DNA-Sequenzen (und die auf ihr kodierten Proteine) von phylogenetisch unterschiedlichsten Quellen entstammen, werden in Metagenomstudien dennoch regelmäßig neue, mitunter auch neuartige Enzyme und Enzymfamilien entdeckt.

## 2.1. Funktionelle Durchmusterung von Metagenombanken

In dieser Arbeit wurden bereits bestehende pCC1FOS-Fosmidbanken funktionell in *E. coli* durchmustert. Das metagenomische Material stammte dabei von psychrophilen, mesophilen bis hin zu thermophilen mikrobiellen Konsortien verschiedenster Herkunft (Heiße Quellen der Tiefsee, Fluss-Sedimente, tierische Darm- und Kotproben). Insgesamt wurden dabei zehntausende Einzelfosmidklone in *E. coli* auf unterschiedliche Hydrolase-Aktivitäten durchmustert, wie beispielsweise Cellulasen, Xylanasen, Amylasen, Proteasen sowie Lipasen und Esterasen. Je nach Bank und durchschnittlicher *Insert*-Größe entspricht dies mehreren hundert Mbp Sequenzinformation, die dabei auf ORFs und kodierte Enzyme untersucht werden konnten. Um den Anteil möglicher aktiver Klone zu erhöhen, wurden die Substrattests gleichzeitig bei vier verschiedenen Temperaturbereichen durchgeführt (8 °C, 15-20 °C, 30-37 °C und 60 bis maximal 70 °C). Diese Inkubationsbedingungen sollten die Herkunft der unterschiedlichen Banken widerspiegeln. Über die Hälfte der 62 aufgefundenen Aktivitäten waren lipolytische Klone (32 Tributyrasen und 6 Trioleinasen), gefolgt von hemicellulolytisch aktiven Fosmidklonen (22 Xylanasen) und einer CMC-ase. Amylolytische Klone wurden in einem Fall identifiziert, und traten in drei bereits gefundenen lipolytisch aktiven Fosmidklonen als zusätzliche Zweitaktivität auf. Keine aktiven Klone waren für Proteasen (Hydrolysehöfe auf Magermilchplatten) für alle getesteten Banken und Temperaturbereiche nachzuweisen. Die Ausbeuten aus den Funktions-basierten Durchmusterungen liegen im Bereich von etwa 0,2 bis 1,3 %. Insbesondere ermöglichte die gemischten Metagenombanken der Tiefsee (mesophile, thermophile und hyperthermophile Mikroorganismen) eine Durchmusterung im Hochdurchsatzformat, indem parallel mehrere hundert Kolonien auf Substratplatten ausplattiert werden konnten. Zwar wurden die Zellen durch den Aktivitätsnachweis entweder abgewaschen oder aber getötet, dennoch ermöglichten Kopien der Indikatorplatten (durch Überstempeln) das Auffinden von positiven Klonen. Für die gemischten Genbanken wurden etwa 20.000 Einzelklone pro Bank durchmustert, sodass die Ausbeute positiver Klone bei rund 0,1 % liegt. Obwohl allerdings nur ein Zehntel aller vermuteter Fosmidklone ausplattiert worden war, wurden in

acht Fällen bereits redundante Fosmidklone beobachtet. Dies könnte dadurch zustande kommen, dass die Zahl der Fosmidklone (ursprünglich auf etwa 20.000 bis 200.000 einzelne Kolonien geschätzt) niedriger ist als ursprünglich angegeben, oder aber die Diversität der mikrobiellen Konsortien wurde aufgrund von Anreicherungskulturen stark reduziert.

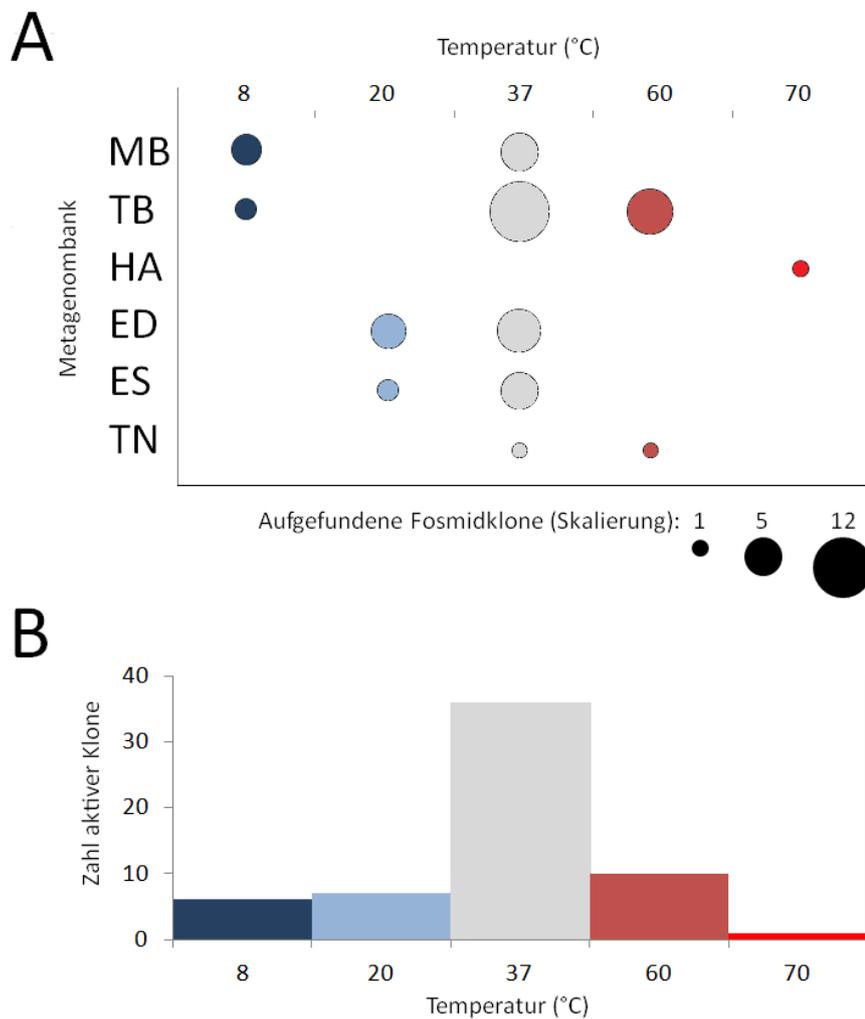
Die Ausbeuten der metagenomischen Durchmusterungen in dieser Arbeit sind dabei repräsentativ für pCC1FOS Fosmid-basierte Metagenomanalysen, zeigt jedoch gegenüber anderen Vektorsystemen, wie beispielsweise Cosmid-Banken, BACs (*bacterial artificial chromosomes*) oder Plasmid-basierte *small-insert* Metagenombanken Vorteile (Tabelle 34). Dabei wird deutlich, wie wichtig die Wahl eines geeigneten Klonierungs- und Expressionsvektors für die Metagenomanalyse ist (Kakirde *et al.*, 2010). Während BACs große Metagenomstücke von bis zu 150 kbp aufnehmen können, erlaubt eine *small-insert* Bank eines herkömmlichen *multi-copy* Vektors lediglich das Tragen von wenigen kb Sequenzinformation pro Zelle. Damit erhöht sich auch der Aufwand für funktionelle Durchmusterungen, da mehr Einzelklone auf Aktivität hin überprüft werden müssen, um vergleichbare Ausbeuten zu erhalten. In manchen Fällen ist die Extraktion von hochmolekularer DNA aus Umweltproben wie heißen oder sauren Quellen, Bodenproben und Darminhalten jedoch nicht möglich, da Tonminerale, Huminsäuren und Nukleasen freie Nukleinsäuren binden oder abbauen (Reigstad *et al.*, 2011; Kakirde *et al.*, 2010). Auch die Art der Extraktionsmethode kann erheblichen Einfluss auf den Grad der DNA-Scherung haben (Abbildung 33).

Falls keine hochmolekulare Nukleinsäure isoliert werden können, bleibt für die Klonierung eines Metagenomes keine andere Wahl als *small-insert* Banken zu erstellen. Die Verwendung kurzer Nukleinsäurefragmente ist manchmal jedoch von Vorteil: Metagenomische DNA kann für den Expressionswirt toxische Sequenzabschnitte enthalten, da kodierte und unfreiwillig stark exprimierte Proteine den Wirt in seinem Wachstum negativ beeinflussen können. Aber auch die Präsenz spezieller Promotorregionen kann die Wirt-eigenen Sigma-Faktoren rekrutieren und sie damit der für die Grundfunktionen essentiellen Genexpression (*housekeeping genes*) entziehen (Warren *et al.*, 2008). Auch intergenische Bereiche sind bekannt für ihre Toxizität gegenüber Wirten wie *E. coli* (Sorek *et al.*, 2007). Metagenomische Fragmente, die solche Sequenzbereiche tragen, können daher überhaupt nicht kloniert werden und entfallen bei den funktionellen Durchmusterungen. *Large-insert* Banken wie sie beispielsweise in BACs vorliegen können, sind reich an intergenischen Bereichen und Promotorregionen. Die Klonierbarkeit dieser Regionen entscheidet dann dabei, ob ein potentiell aktiv exprimierbares Gen gefunden werden kann oder nicht. Aus diesem Grund existieren heute meist kommerziell erhältliche Vektorsysteme, die die selektive Steuerung der Kopienzahl ermöglichen und selbst gering exprimiert Genprodukte identifizieren können.

Die Ausbeuten aktiver und identifizierter lipolytischer oder (hemi-)cellulolytischer Klone in pCC1FOS sind mit anderen Studien durchaus vergleichbar, wobei Cosmide und Fosmide mit *Insert*-Größen zwischen 20 und 40 kbp die höchsten Ausbeuten zeigen können (Tabelle 34).



Da zum Zeitpunkt dieser Studie sechs Metagenombanken verschiedenster Herkunft für funktionelle Durchmusterungen in *E. coli* zur Verfügung standen, bietet sich ein Vergleich der Ausbeuten zwischen den Banken an (Abbildung 48). Generell stimmen die Temperaturen der aktiven Klone mit der Herkunft der Mikroorganismen überein, wonach in psychrophilen und mesophilen Banken (Elbe Sediment, Elefantendung und Tiefseebank MB) lediglich Aktivitäten zwischen 8 und 37 °C identifiziert werden konnten, nicht jedoch bei höheren Temperaturen. Einzige Ausnahme bildet hier die Xylanaseaktivität aus der Metagenombank des Schiffsbohrwurms *Teredo navalis* (TN), die bei 65 °C entdeckt werden konnte. Bei den thermophilen (TB) und hyperthermophilen (HA) Mikroorganismen konnten die meisten thermostabilen Enzyme identifiziert werden.



**Abbildung 48: Zusammenfassung der Durchmusterungen verschiedener Genbanken in *E. coli*.** **A:** Die Anzahl der aktiven Fosmidklone je Metagenombank setzt sich aus allen erhaltenen Aktivitäten zusammen und werden durch Kreisflächen visualisiert. Die verschiedenen Temperaturbereiche sind durch Farbkodierung gekennzeichnet (dunkelblau: 8 °C; hellblau: 20 °C oder Raumtemperatur; grau: 30 bis 37 °C; rot: 60 °C; hellrot: 70 °C). Abkürzungen der getesteten Metagenombanken: Mesophile (MB) und Thermophile Bakterien (TB), hyperthermophile Archaeen (HA), jeweils aus Tiefseeproben; Elefantendung (ED); Elbe Sediment (ES) und Anreicherungskulturen von Isolaten des Schiffsbohrwurms *Teredo navalis* (TN). **B:** Summe aller in *E. coli* aktiven Fosmidklone in Abhängigkeit von der beobachteten Aktivitätstemperatur.

Interessanterweise ergibt sich aus allen beobachteten Gesamtaktivitäten der diversen Metagenombanken in *E. coli* ein spezifisches Temperaturspektrum. Danach wurden unter idealen (mesophilen) Wachstumsbedingungen bei Temperaturen zwischen 30 bis 37 °C in *E. coli* auch die meisten Enzymaktivitäten entdeckt. Dies liegt einerseits daran, dass viele aktive Klone in mesophilen Banken aufgefunden werden konnten, andererseits aber waren bei den thermophilen Konsortien im höheren Temperaturbereich wesentlich geringere Ausbeuten erzielt worden. Diese Daten stützen die Hypothese, dass nicht nur die phylogenetische Distanz, sondern auch physiologische Eigenschaften des Wirts für die erfolgreiche, funktionelle Expression metagenomischer Umwelt-DNA notwendig sind. Für die Metagenomanalyse bedeutet dies, dass der Erfolg der funktionsbasierten Strategien damit von physiologischen Präferenzen (z.B. Temperatur) des Expressionswirts abhängen kann.

## 2.2. Sequenzierung und bioinformatische Auswertung

Im Bereich der metagenomischen Sequenzierung haben sich bis heute viele verschiedene Sequenzierungsplattformen etabliert. Die 454-pyrosequencing Technologie bietet gegenüber klassischen Kapillarsequenzierungen nach Sanger die Möglichkeit, wesentlich höhere Durchsatzraten bei vergleichbaren *read*-Längen zu erhalten. Speziell die 454-Technologie wird in der Metagenomik sehr häufig eingesetzt, da die gegenüber anderen Sequenzierplattformen höheren *read*-Längen die Assemblierung von *a priori* unbekannter Umwelt-DNA wesentlich erleichtert. Dies gilt insbesondere für kurze, sich wiederholende Sequenzabschnitte (*repeats*) und die *de novo* Assemblierung.

### 2.2.1. Assemblierung und *mapping*

In dieser Arbeit wurden 41 pCC1FOS-Fosmide der insgesamt 48 positiven Klone (eine halbe 96-Well Mikrotiterplatte, Sequenzierprojekt *ANG48*, A. Thürmer, Göttingen Genomics Laboratory) mittels paralleler Sequenzierung ohne Identifizierungsmerkmale (*sequencing tags*) sequenziert. Dabei werden die einzelnen DNA-Moleküle äquimolar eingestellt, sodass jede Matrize mit gleicher Konzentration und Häufigkeit vorkommt. Die Zuordnung der entsprechenden Fosmidklone zu den jeweiligen assemblierten *contig*-Fragmenten war schließlich durch *mapping* möglich. Dabei wird mit den Sequenzierprimern von den Enden des Fosmids einzeln nach außen ansequenziert. Da die Fosmid- und Klonposition in der Mikrotiterplatte bekannt ist, erhält jedes übereinstimmende *contig* die dazugehörige Identifizierung.

Nach Assemblierung der *reads* konnten von den 48 Fosmiden nur die neun größten *contigs* von beiden Seiten eindeutig einem Fosmidklon zugeordnet werden (*mapping*). Für vier weitere *contigs* war eine

Flanke mit den Endsequenzierungen übereinstimmend. In sieben weiteren Fällen konnten potentielle Kandidatengene aus nicht zuordenbaren *contig*-Sequenzen identifiziert werden.

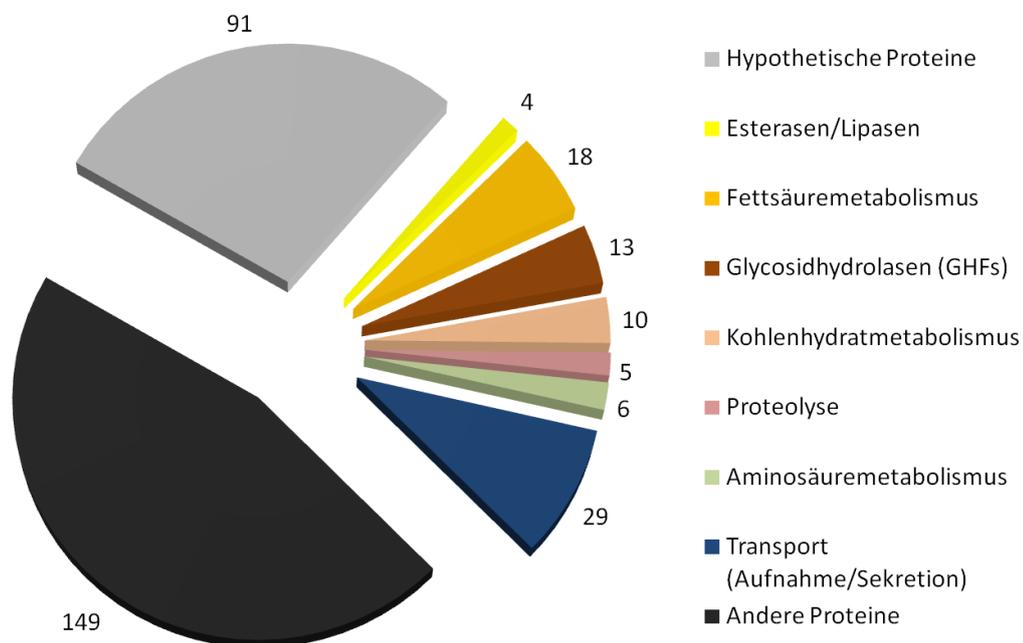
Nur 18,8 % der aufgefundenen Fosmide lagen vollständig sequenziert vor. Weitere 8,3 % konnten teilweise zugeordnet werden, weitere 14,5 % ergaben lediglich Kandidatengene ohne Zuordnungen. Auffällig viele Endsequenzierungen wurden zu identischen *contigs* zugeordnet. Dies war beispielsweise der Fall für mehrere Esterase-aktive Klone der MB-Bank. Dabei zeigten die Klone MB-est-2, 3 sowie 5 und 6 mit *contig* 10 Übereinstimmungen, MB-est-2 und 3 ebenfalls mit *contig* 12 und MB-est-7 und 8 mit *contig* 25. Dies würde erneut dafür sprechen, dass identische metagenomische Sequenzen der Tiefseeproben mehrmals aufgetreten sind (entweder durch die geringere Klonzahl oder aber durch die relativ niedrige Diversität durch Anreicherungen). Der lipolytische Klon TB-lip-3, der gegenüber Triolein aktiv war, wies nachträglich Aktivität auf CMC auf und war mit der *contig*-Sequenz von HA-cmc-1 identisch. In diesem Fall musste es sich um eine Kreuzkontamination innerhalb der Sequenzierplatte handeln, was allerdings erst nach erfolgter Sequenzierung beobachtet worden war. Alle weiteren der 153 *contig* Sequenzen über 500 bp enthielten keine partiellen oder vollständigen Sequenzen mit annotierten Glycosidhydrolase-Familien (GHF). Auch die Suche nach lipolytischen Proteinen mittels konservierten Motiven (*regular expressions*) ergaben keine weiteren Treffer.

Technische Verbesserungen der 454-Pyrosequenzierung erlauben das selektive Markieren von Fragmenten vor der Sequenzierung (*sequence-tags*), welche die spätere Zuordnung von verschiedenen Nukleinsäurematrizen in einem parallelen Sequenzierlauf ermöglichen. Das sogenannte *Amplicon-sequencing* bedient sich hierbei dem Anheften von Identifizierungssequenzen mittels PCR, sodass die zu sequenzierenden Fragmente später zuordenbar sind (*multiplexing*). Diese Strategie wäre für die Sequenzierung der 48 Fosmide von Vorteil, da sie keine nachträglichen Endsequenzierungen mehr benötigt. Selbst nicht vollständig sequenzierte *contig*-Fragmente wären dann eindeutig zuordenbar, unabhängig davon, wie groß die assemblierten Stücke sind. Selbst im ungünstigsten Fall ermöglicht *multiplexing* die Zuordnung von nicht assemblierbaren Nukleotidsequenzen, wie sie beispielsweise in *repeat*-reichen Regionen vorkommen können. Diese Bereiche enthalten lange Strecken identischer Basenabfolgen, die eine Zuordnung für Assembler-Programme erschweren. Ähnlich verhält es sich bei Genbereichen, die aus Duplikationen oder Transposon-Insertionen entstanden sind. Etwaige Lücken könnten dann schrittweise bis zu den Sequenzenden geschlossen werden (*primer-walking*).

### 2.2.2. Identifizierung von Kandidatengen

Von den assemblierten Metagenomsequenzen konnten über 300 ORF-Produkte bezüglich ihrer funktionellen Zuordnung (*Gene Ontology*, GO, Ashburner *et al.*, 2000) klassifiziert werden

(Abbildung 49). Etwa ein Drittel von diesen ORFs kodierte für hypothetische Proteine oder Proteine ohne charakterisierte Domänen. Knapp die Hälfte der Genprodukte war auf diverse, hier nicht weiter aufgeschlüsselte Funktionen zurückzuführen (beispielsweise aus diversen Stoffwechselwegen, Oxidations- und Reduktionsprozessen, Grundfunktionen der Transkription und Translation, spezifische enzymatische Reaktionen, strukturelle Proteine). Interessanterweise wurden unter allen aufgefundenen Sequenzen auch andere biotechnologisch relevante Kandidatengene vorhergesagt. Darunter sind laut Annotation mehrere proteolytisch aktive Enzyme (Peptidasen, Proteasen) und weitere Glycosidhydrolase-Proteine. In einem Fall konnte eine mögliche Fructosaminkinase in Fosmid TB-xy1-3 (ORF 19c, 90 % Sequenzidentität zu *R. marinus* DSM4252) identifiziert werden. Fructosylamin-3-Kinasen bzw. Fructoselysin-3-Kinasen (FN3K/FL3K), Fructosamin-Oxidasen (FAOX) und Fructoselysin-6-Kinase (FL6K) stellen die drei Gruppen an Enzymen dar, welche Glykierungen von reduzierenden Zuckern an  $\alpha$ - oder  $\varepsilon$ -Aminogruppen von Proteinen (z.B. Lysine) abbauen können (auch Amadoriasen genannt) (Lin und Zheng, 2010; Monnier 2005). Die Glykierungen (Amadori-Produkte) entstehen während der Maillard-Reaktion, der nicht-enzymatischen Bräunungsreaktion wie beispielsweise bei Lebensmitteln. Der Einsatz deglykierender Enzyme ist insbesondere für die Waschmittelindustrie interessant, da Amadori-Produkte zu Verfärbungen in Textilien führen (Deppe *et al.*, 2008), aber auch diagnostische Anwendungen wurden berichtet (Lin und Zheng, 2010).



**Abbildung 49: Funktionelle Klassifikation aller kodierten Proteine, die auf assemblierten Fosmidsequenzen (siehe Tabelle 24) identifiziert werden konnten. Den Daten liegen die *in silico* vorhergesagten nächsten übereinstimmenden Proteinen und Annotationen (lt. NCBI nr-Datenbank und *blastp*) sowie signifikante Pfam-A Klassifikationen zu Grunde. Insgesamt waren 375,7 kbp aus 19 *contigs* eindeutigen Kandidatengenen zuordenbar.**

Da nur ein Bruchteil der verfügbaren Sequenzdaten zu vollständigen oder teilweise sequenzierten Fosmiden führte, konzentrierte sich die bioinformatische Analyse auf diese Fosmide. Drei der pCC1FOS-Fosmide waren eindeutig ihren aktiven Klonen zuordenbar (HA-cmc-1, TB-est-7 und ES-6-B5). Deren Aktivitäten konnten durch Subklonieren oder direkter Expression bestätigt werden. Für Fosmid ES-6-B5 war eine annotierte Glycosidhydrolase der Familie 5 identifiziert worden (Xyn5A), wobei die Hauptaktivität auf Xylanen mittels kolorimetrischen DNSA-Tests von Rohextrakten bestätigt werden konnte. Die lipolytische Aktivität des Fosmids TB-est-7, vermutlich von einer annotierten Patatin-ähnlichen Phospholipase, wird momentan noch einer heterologen Expression und Charakterisierung unterzogen (V. T. T. Pham, Lehrstuhl für Mikrobiologie, TU München). Die Analyse des cellulolytischen Fosmids HA-cmc-1 wird im nächsten Abschnitt ausführlich erläutert.

Der hemicellulolytische Klon TB-xyl-3 zeigte mit ORF 9c eine Glycosidhydrolase der Familie 18 als einzige Hydrolase in der Fosmidsequenz an. Laut der Carbohydrat-Active enZyme database (CAZy-Datenbank) umfasst diese Enzymfamilie Chitinasen und Chitino oligosaccharid-Hydrolasen, Lysozym (Spaltung der  $\beta$ -1,4-glykosidische Bindungen zwischen N-Acetylmuraminsäure- und N-Acetylglucosamin, EC 3.2.1.17) und Endo- $\beta$ -N-acetylglucosaminidase-Aktivitäten (EC 3.2.1.96) sowie die Hydrolyse von Lipochitooligosacchariden (*nodulation factor hydrolase*, EC 3.2.1.-). Auch nicht-katalytisch aktive Proteine wie Xylanase Inhibitoren, Kohlenhydratbinde- und Speicherprotein (Concanavalin B, Narbonin) sind strukturell mit dieser Proteinfamilie verwandt (Durand *et al.*, 2005; Lombard *et al.*, 2013). Da für diese Proteinklasse bisher keinerlei Xylan abbauende Aktivität beschrieben worden ist, könnte es sich bei diesem Protein um eine neuartige hemicellulolytische Aktivität innerhalb der GH18-Familie handeln. Strukturelle Ähnlichkeiten zwischen Cellulose und Chitin beruhen auf der identischen  $\beta$ -1,4-verbundenen Glucose-Molekülen, wobei die Hydroxygruppen an Position 2 des  $\beta$ -D-Glucopyranose-Monomers durch Aminogruppen (N-Glucosamin) oder durch Acetamidogruppen (N-Acetylglucosamin) substituiert vorliegen können. Xylane hingegen sind strukturell abweichende Heteropolysaccharide, gehören zu den Hemicellulosen und sind aufgrund der starken Variabilität der Seitengruppen komplexer aufgebaut. Weiterhin erscheint der katalytische Mechanismus der GH18-Enzymfamilie bei der Hydrolyse des Chitins nicht geeignet für eine Hydrolyse von Xylanen. Dieser Mechanismus beinhaltet die Reaktion des Carbonyl-Sauerstoffs der 2-Acetamidogruppe des Substrats, welcher als Nukleophil zur Ausbildung des Oxazoliniumion-Intermediates fungiert. Die Carboxylgruppen eines Asparaginsäure- und Glutaminsäurerests des katalytischen Zentrums sind an der Reaktion unmittelbar beteiligt und liegen im diagnostischen D-X-X-D-X-D-X-E Motif vor, welches für Xyn18A ebenfalls vorhanden ist (Van Aalten *et al.*, 2001). Zwar verfügen Xylane ebenso über Acetyl-Seitengruppen, jedoch sind diese über die Hydroxygruppe und nicht über ein Amid wie im Falle von Chitin substituiert. Zur Analyse der beobachteten Aktivität und biochemischen Charakterisierung bedarf es daher der funktionellen Expression des Zielproteins Xyn18A. Die Expression des vorläufigen Vollängenproteins führte zu

nachweisbaren Proteinmengen im Rohextrakt, jedoch war der überwiegende Teil des Proteins in der nicht-löslichen Fraktion aufgefunden worden (als ausgefallenes Protein in sogenannten *inclusion bodies* oder aber nach Temperatureinwirkung). Als mögliche Ursache kann die N-terminale Präsenz eines Signalpeptides in der vollständigen Aminosäuresequenz genannt werden, welche die heterologe Herstellung von löslichen Proteinpräparationen empfindlich hemmen kann. Die erneute Klonierung und Expression des Xyn18A-Proteins ohne das N-terminale Signalpeptid oder als Fusion mit einem anderen löslichen Protein könnte die Ausbeuten erhöhen.

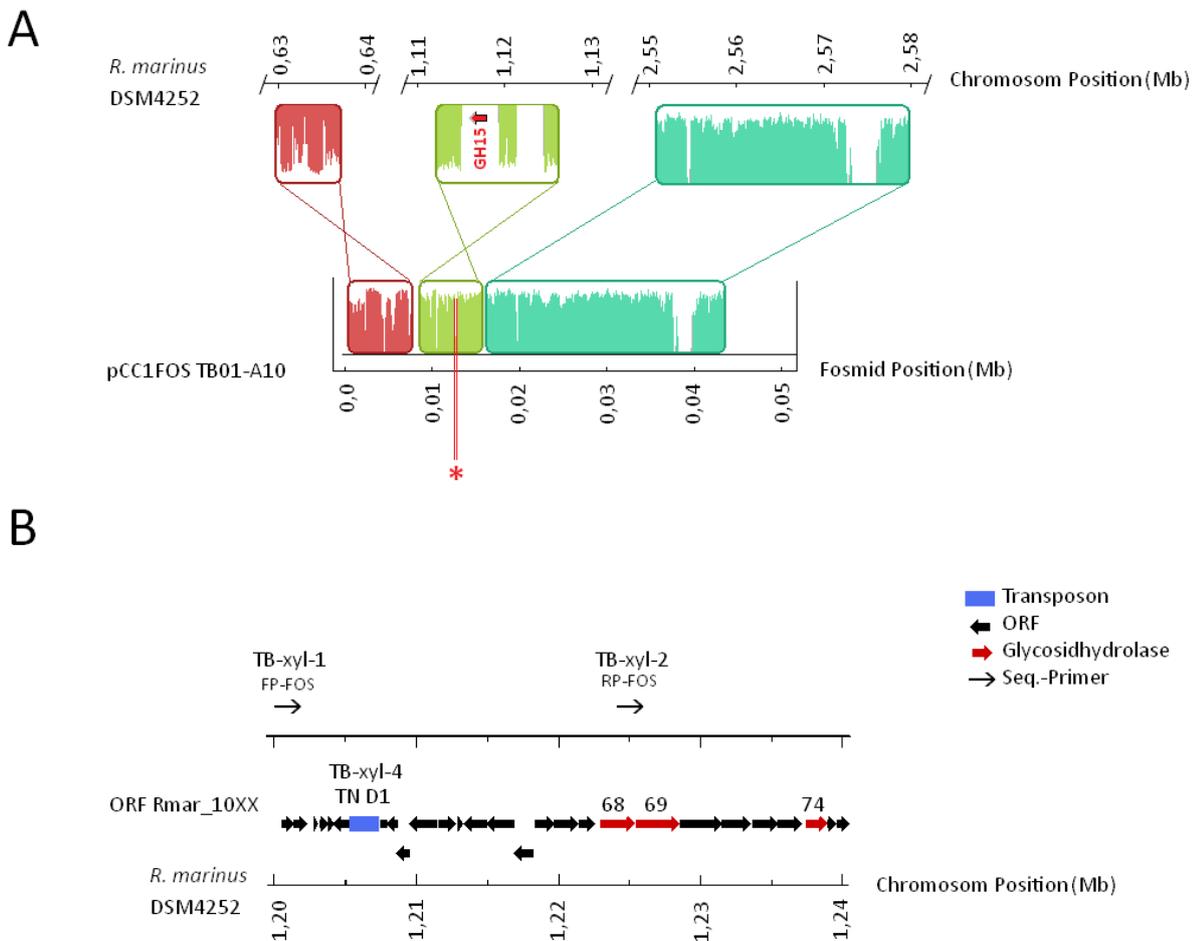
Vier weitere Fosmide waren zwar vollständig sequenziert worden, jedoch waren keine Kandidatengene vorhersagbar (TB01-A10, TB-xyl-4 und TN-9-B5). Die starke Lipase-Aktivität von Klon ES-7-E10 ist aufgrund einer verschleppten Kontamination mit einem Tributyrase-aktiven Mikroorganismus zu erklären und wurde daher nicht weiter untersucht. Sieben weitere *contigs* enthielten Gensequenzen, die für potentiell aktive Enzyme kodieren könnten. Ihnen fehlte jedoch die entsprechende Zuordnung (*mapping*) zum korrespondierenden Fosmidklon und wurden daher ebenfalls nicht weiter untersucht.

Bei den beiden Fosmiden ohne direkten Kandidaten, TB01-A10 und TB-xyl-4, wurden vergleichende Genomanalysen mit den nächsten Homologen Vertreter *Rhodothermus marinus* Stamm DSM 4252 mit dem Programm Mauve durchgeführt (Abbildung 50A).

Für den amylolytischen Fosmidklon TB01-A10 zeigten sich drei Homologieblöcke (*similarity blocks*), die höhere Übereinstimmung mit Sequenzen aus dem *R. marinus* Stamm DSM 4252 aufweisen. An Position 12,6 bis 12,8 kbp der Fosmidsequenz enden jeweils die Homologien zum *R. marinus* Chromosom an Position 1,1191 bis 1,1149 Mb. Zwischen der Lücke liegt unter anderem der ORF Rmar\_0982, welcher für ein annotiertes Glycosidhydrolase 15-ähnliches Protein (Proteinaccession YP\_003290264.1) kodiert. Diese Enzymklasse umfasst Proteine mit Aktivitäten als Glucoamylase (EC 3.2.1.3),  $\alpha$ -Glucosidase (EC 3.2.1.20) sowie mit Glucodextranase-Aktivitäten (EC 3.2.1.70). Erstere katalysieren die Freisetzung von D-Glucose vom nicht-reduzierenden Ende von Oligo- oder Polysacchariden sowie von Stärke. Es ist daher denkbar, dass die beobachtete amylolytische Aktivität des Klones TB01-A10 auf diese Gensequenz zurückgeführt werden könnte. Dies würde jedoch voraussetzen, dass die Fosmidsequenz nicht vollständig ist (z.B. durch inkorrekte Assemblierung der *reads* an dieser Stelle). Denkbar wäre ebenfalls ein Verlust der Zielsequenz durch Fosmid-Instabilität während der Kultivierungen bis zum Zeitpunkt der Sequenzierung.

Klon TB-xyl-4 zeigt laut vergleichender Genomanalyse zusammenhängende Sequenzabschnitte (*contiguous*) zur Chromosomsequenz von *R. marinus*. Da sich für diesen Sequenzabschnitt keine möglichen (hemi-)cellulolytischen Kandidatengene voraussagen ließen, wurde eine Transposonmutagenese-Strategie gewählt. In einem Fall konnte eine Mutante (Nummer 1-D4) isoliert werden, die gegenüber den anderen Mutanten eine erhöhte Xylanase-Aktivität aufwies (sichtbar durch

Hydrolysehöfe auf Substratplatten mit Xylan und anschließender Kongorot-Färbung). Die Sequenzierung der Flanken zeigte bei diesem Klon jedoch keine Übereinstimmung mit der Fosmidsequenz von TB-xyl-4, war jedoch im Chromosom von *R. marinus* aufzufinden (Abbildung 50B). Die chromosomale Region um Position 1,20 bis 1,24 Mb kodiert für mehrere Glycosidhydrolasen (Rmar\_1068, Rmar\_1069 und Rmar\_1074). Die aufgefundenen Proteine sind eine GHF 43 mit CBM6 Bindedomäne (YP\_003290348.1), eine GHF 10 mit zwei Bindemodulen der Familie 4\_9 (YP\_003290349.1) sowie einem GHF 10 Protein ohne weitere Bindemodule (YP\_003290354.1).



**Abbildung 50: Bioinformatische Analyse der beiden Fosmide der Klone TB01-A10 und TB-xyl-4, welchen keine direkten Kandidatengene vorhergesagt werden konnten. Die beiden Sequenzen waren homolog zur Genomsequenz von *Rhodothermus marinus* DSM 4255. A: Mauve-Alignment des Fosmids pCC1FOS TB01-A10 mit den entsprechenden Homologieblöcken im *R. marinus* Chromosom. In Block 2 (grüne Einfärbung) ist in einer etwa 4 kbp großen Lücke eine annotierte Glycosidhydrolase der Familie 15 (GH15, roter Pfeil) kodiert (Genprodukt des ORFs Rmar\_0982, YP\_003290264.1). Dieser Sequenzabschnitt kommt in der Fosmidsequenz nicht vor (\*, rot markiert). B: Sequenzabschnitt des *R. marinus* Chromosoms, der für das Transposon (TN, Klon 1-D4) in Fosmid TB-xyl-4 verwendet worden ist. Neben dem Transposon binden beide Sequenzierprimer der Fosmide TB-xyl-1 und TB-xyl-2 in der Zielsequenz. Rot markiert sind die Kandidatengene Rmar\_1068, Rmar\_1069 sowie Rmar\_1074, welche jeweils für drei vorhergesagte Glycosidhydrolasen kodieren (GHF 43 mit CBM6 Domäne, GHF 10 mit 2x CBM4\_9 sowie einem GHF 10 Protein).**

Die beiden erwähnten GHF Familien sind durch vielfältige Aktivitäten gegenüber (hemi-)cellulolytischen Substraten charakterisiert, haben jedoch eine Endo-1,4- $\beta$ -Xylanase Aktivität (EC 3.2.1.8) gemeinsam. Interessanterweise konnten auch die Endsequenzen von den beiden anderen Fosmidklonen TB-xyl-1 und TB-xyl-2 in diesem gemeinsamen chromosomalen Locus aufgefunden werden. Dies stützt erneut die Hypothese, dass alle Xylanase-aktiven Klone (TB-xyl-1 bis 4) womöglich auf ein *Rhodothermus*-ähnliches Bakterium einer Anreicherungskultur zurückgeführt werden können. Auch für die anderen aktiven Fosmidklone TB-est-1, TB-est-5 und TB-est-7 konnten Übereinstimmungen gegenüber diesem thermophilen Bakterium nachgewiesen werden.

Einige der oben genannten sequenzierten Fosmide zeigten mittels gängigen bioinformatischen Methoden keine Übereinstimmungen zu Kandidatengen und kodierten Proteinen. Es ist daher möglich, dass sie neuartige Sequenzsignaturen tragen, welche unbekannt Proteinfamilien angehören. Ebenso ist es jedoch möglich, dass bereits bekannte Proteine neuartige Enzymaktivitäten aufweisen, die vorher noch nicht beobachtet worden waren. Welches der Kandidatengene – oder ihre Kombination daraus – für die beobachtete Aktivität verantwortlich ist, kann durch *shotgun*-Banken, selektive PCR-Amplifikation, heterologe Expression und entsprechende Charakterisierung der Zielproteine beantwortet werden. Ob und warum die Assemblierung und Zuordnung (*mapping*) zu den korrespondierenden Klonen in Einzelfällen falsch war, bleibt in dieser Arbeit ungeklärt.

### **2.3. Identifizierung und Charakterisierung einer archaeellen Endoglucanase Cel12E**

#### **2.3.1. Durchmusterung und *in silico*-Analyse des Fosmids HA-cmc-1**

Von allen durchmusterten Fosmidklonen der Anreicherungen hyperthermophiler Archaeen konnten cellulolytische Aktivitäten auf CMC-Substratplatten nur bei erhöhten Temperaturen zwischen 60 und 70 °C identifiziert werden. Alle Einzelfosmide waren auf ein Fosmid (HA-cmc-1) zurückzuführen, welches im Restriktionsverdau immer ein identisches Bandenmuster lieferte. Es ist daher zu vermuten, dass das genomische Nukleinsäurefragment ursprünglich einer einzelnen Kolonie entstammt, welche zur Erstellung des Fosmidpools von der Agarplatte abgewaschen worden war. Alle weiteren Substrate ergaben keine aktiven Klone dieser Fosmidbank. Die Sequenzierung des betreffenden Fosmids HA-cmc-1 lieferte ein vollständig assembliertes *contig* mit 38,2 kbp Größe, welches für bioinformatische Analysen zur Verfügung stand. Der einzige für eine potentielle Glycosidhydrolase-kodierende ORF Nummer 23 liegt etwa in der Mitte des Fosmids lokalisiert. Vor diesem ORF liegen ein putativer Transkriptionsfaktor IID (TATA-Bindeprotein, ORF 21) sowie eine Histondeacetylase (ORF 22), welche möglicherweise transkriptionell regulierende Funktionen übernehmen. Stromabwärts von ORF 23 wurde ein Gen für einen Natrium-Protonen-Antiporter vorhergesagt. Das

von ORF 23 kodierte Polypeptid umfasst ein katalytisches Modul der Glycosidhydrolase Familie 12 (GHF 12), sowie zwei C-terminale Kohlenhydratbindedomänen der Familie 2 (CBM2). Glycosidhydrolasen der Familie 12 (ehemals Cellulasefamilie H) sind durch eine  $\beta$ -jelly roll Struktur charakterisiert. Der Reaktionsmechanismus umfasst eine doppelte Verlagerung (*double-displacement*) unter Beibehaltung der Konfiguration des anomeren Kohlenstoffatoms (*retaining mechanism*). Dabei kommt es zur Ausbildung eines kovalenten Enzym-Glycosyl-Intermediates. Zwei Glutamatreste dienen dem katalytischen Zentrum als katalytisches Nukleophil und als Protonendonator zur Hydrolyse des Polysaccharids an den glycosidischen Bindungen. Die beteiligten Glutamatreste (Glu171, Glu266) des katalytischen Zentrums ließen sich auf Basis eines multiplen Sequenzalignments bestimmen (Abbildung 51A). Zu bekannten Substraten dieser Enzymfamilie zählen neben rein  $\beta$ -1,4-glycosidisch verknüpfter Cellulose (Aktivität als Endoglucanase, EC 3.2.1.4) und gemischten glycosidischen Bindungen (Aktivität als  $\beta$ -1,3-1,4-Glucanase, EC 3.2.1.73) auch hemicellulolytische Aktivitäten gegenüber Xyloglucan (Hydrolase EC 3.2.1.151, Endotransglycosylase EC 2.4.1.207). Die GHF 12 Enzymfamilie tritt bei Vertretern aller Domänen des Lebens auf und wurde bereits in einer Vielzahl an cellulolytischen Bakterien und Pilzen sowie hyperthermophilen Archaeen charakterisiert. Weiterhin wurden zwei C-terminale Kohlenhydratbindemodule (*carbohydrate binding module*, CBM) der Familie 2 identifiziert, die 47,4 % Sequenzidentität zueinander haben. Die CBM-Familie 2 (ehemals Cellulose-Bindedomäne der Familie II, CBD II) weist eine dreidimensionale  $\beta$ -Sandwich Struktur auf und stellt eine der weit verbreitetsten prokaryotischen Kohlenhydratbindemodule dar. CBMs dienen der Lokalisierung der im Protein gekoppelten katalytisch aktiven Domäne (z.B. ein Glycosidhydrolase-Modul) auf unlöslichen Substraten. Mitglieder dieser CBM-Familie vermögen Cellulose aber auch Xylane und Chitin zu binden. Die Bindung zu unlöslichen Polysacchariden wird dabei durch aromatische Aminosäurereste an der Oberfläche mittels hydrophober Interaktionen ermöglicht. Auf Basis struktureller und funktioneller Übereinstimmungen wird die planare Ausrichtung der aromatischen Seitengruppen (Typ A, CBM2a) den zueinander verdrehten Konformationen gegenübergestellt (Typ B, CBM2b), welche unterschiedliche Substratspezifitäten nach sich ziehen (Simpson *et al.*, 2000). Archaeelle Vertreter dieser Bindemodule sind lediglich für Chi18A und Chi18B aus *Pyrococcus furiosus* DSM 3638 sowie Chi18A in *Thermococcus kodakarensis* KOD1 berichtet worden, jedoch ausschließlich im Zusammenhang mit Chitin-Bindung und -Hydrolyse durch GH18-Module (Nakamura *et al.*, 2008; Tanaka *et al.*, 1999). Im multiplen Sequenzalignment konnten für jede der beiden vorhandenen CBM 2-Module vier konservierte aromatische Aminosäuren (Tryptophane) vorhergesagt werden (Abbildung 51B). Binstudien zeigten, dass Cel12E an unlöslicher Cellulose (Avicel, PASC) und darüberhinaus auf reinen 1,3- oder gemischten 1,3-1,4- $\beta$ -D-Glucanverbindungen (Pachyman und Gerstenglucan) sowie Chitin binden kann, jedoch nicht auf den verschiedenen getesteten Xylanen. Die Bindefähigkeit der CBM-Familie ist dabei eher von der Struktur des Substrats als von der spezifischen Erkennung funktioneller Gruppen abhängig (Simpson *et al.*, 2000).

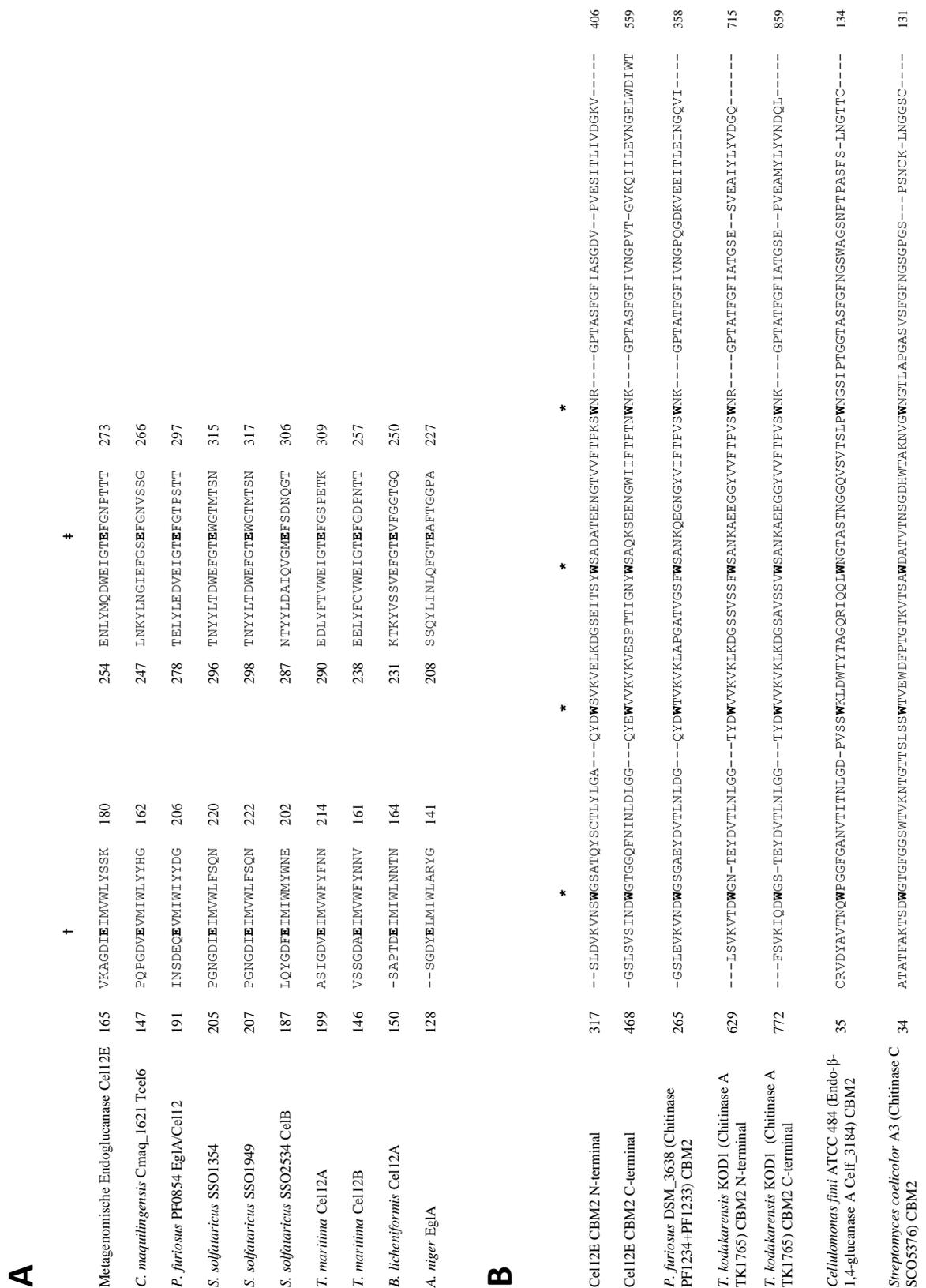


Abbildung 51: Multiples Sequenz-Alignment der katalytischen Zentren ausgewählter Glycosidhydrolasen der Familie 12 (A) und Kohlenhydratbindemodule (B) der Familie 2 (CBM 2). Die gezeigten Ausschnitte zeigen die an der katalytischen Diade beteiligten Glutamatreste, die als Nukleophil (†) sowie Protodonor (‡) fungieren. Die Bindung der CBMs an unlösliche Polysaccharide geschieht über mindestens zwei der vier konservierten Tryptophanreste (\*). Die jeweiligen Aminosäurepositionen der einzelnen Proteine sind vor und nach den Alignments angegeben.

Da die helikale Sekundärstruktur des Xylans die Bindung von CBM2b favorisiert, liegt die Einordnung der Bindemodule für Cel12E zum Typ A nahe. Die Präsenz zweier solcher CBMs könnte eine Adaptation an hohe Temperaturen sein. Ihr Einfluss auf die Bindefähigkeit sowie auf die enzymatische Aktivität müsste anhand von C-terminal verkürzten Cel12E-Varianten getestet werden.

Aus biologischer Sicht könnte es sich bei der archaeellen Endoglucanase um ein sekretiertes oder Membran-assoziiertes Protein handeln, welches aus einem Archaeon hydrothermalen Tiefseequellen stammt. Organische Polysaccharide aus abgestorbener Biomasse bzw. Speicherkohlenhydrat könnte dem Mikroorganismus durch enzymatische Hydrolyse als Kohlenstoffquelle dienen. Ähnlichkeitsstudien zeigten höchste Übereinstimmungen zu *Thermococcus* sp. AM4. Dieser archaeelle und hyperthermophile Organismus zählt ebenso wie *Pyrococcus* zum Taxon Euryarchaeota, die unter anaeroben Bedingungen organische Verbindungen als Elektronendonator und elementaren Schwefel als terminalen Elektronenakzeptor nutzen können (Madigan *et al.*, 2001).

### 2.3.2. Expression und biochemische Charakterisierung

Obwohl die Expression von Cel12E in *E. coli* BL21 in pET21a(+)-Vektor zu beachtlichen Proteinausbeuten in der löslichen Fraktion des Zelllysats führte, war die Proteinintegrität nicht problemfrei. Einzelne Degradationsfragmente traten im Laufe aller Aufreinigungsschritte bei Anionenaustausch und Gelfiltrationschromatographie auf. Wiederholte Aufreinigungen nach entsprechender Reinigung der ÄKTA FPLC-Anlage und Säulen nach Herstellerangaben (Behandlung mit hochkonzentrierter NaOH-Lösung) ergaben keine Verbesserung des Ergebnisses. Die Größe der entstandenen Fragmente ermöglichte die Rekonstruktion der vermuteten proteolytischen Schnittstellen, die offenbar relativ genau zwischen den einzelnen Modulen stattfanden (Daten nicht gezeigt). Inkubationen mit dem Säulenmaterial ohne weitere Zusätze führten zu keinerlei proteolytischen Effekten. Erst die wiederholte Zugabe von Protease-Inhibitoren konnte die fortschreitende Degradation des Zielproteins Cel12E verhindern.

Das Protein Cel12E zeigte vorwiegend auf löslichen Polysacchariden mit  $\beta$ -1,4-glycosidischen Bindungen (z.B. Carboxymethylcellulose, Hydroxyethylcellulose) sowie auf  $\beta$ -Glucan der Gerste (*Barley  $\beta$ -glucan*,  $\beta$ -1,3-1,4-Bindungen) nachweisbare Aktivität. Während mikrokristalline Cellulose nicht abgebaut werden konnte, war geringfügige Bildung reduzierender Zuckereinheiten im DNSA-Assay auf Säure-geschwollener Cellulose (PASC) und aufgrund geringer amorpher Cellulose-Anteile auch auf Avicel messbar. Mittels Dünnschichtchromatographie konnten vor allem Cellobiose und Cellotriose als Hydrolyseprodukte langkettiger  $\beta$ -1,4-verknüfter Polysaccharide und Cellopentaose festgestellt werden. Die gemachten Beobachtungen legen nahe, dass es sich bei Cel12E um eine klassische Endo- $\beta$ -1,4-Glucanase handelt. Allerdings lassen die Daten keinen Aufschluss über

Prozessivität und Hydrolyserichtung im Polysaccharid zu (Aktivität vom reduzierenden bzw. nicht-reduzierenden Zuckerende). Das Enzym wies beim Temperaturoptimum von 92 °C nur eingeschränkte Stabilität gegen Inaktivierung auf. Die Aktivität war nach weniger als zwei Stunden auf die Hälfte gefallen. Aktivitäten über dem Siedepunkt des Wassers (entsprechend ca. 97 °C) waren kaum nachweisbar. Der pH-Bereich der optimalen Aktivität lag um pH 5,5. Andere charakterisierte GHF12 Endoglucanasen archaellen Ursprungs sind CelS aus *Sulfolobus solfataricus* MT4 (Limauro *et al.*, 2001), die Zellwand-verankerte oder putativ extrazelluläre Endoglucanase Sso1354 (Girfoglio, *et al.*, 2012) sowie Sso2534 und Sso1949 (Huang *et al.*, 2005), die auf Basis der Genomsequenz von *S. solfataricus* Stamm P2 (She *et al.*, 2001) kloniert und exprimiert werden konnten. Ein weiteres Beispiel ist EglA aus *Pyrococcus furiosus* (Bauer *et al.*, 1999), deren Kristallstruktur bereits vorhanden ist (Kim *et al.*, 2012). Weitere thermostabile Vertreter der Glycosidhydrolasefamilie 12 sowie ihre biochemischen Eigenschaften sind in (Tabelle 35) zusammengefasst.

Im Zuge der biochemischen Charakterisierung konnte der Einfluss von Additiven auf die Aktivität untersucht werden. Insbesondere das Reduktionsmittel Dithiothreitol (DTT) führte zu signifikanten Aktivitätssteigerungen bei Konzentrationen von 1 bis 10 mM. Dies weist darauf hin, dass das Enzym oxidationsempfindliche Gruppen besitzt, deren Oxidation die Aktivität beeinträchtigen kann. Ebenfalls höhere Aktivitäten wurden bei erhöhten Chlorid-Konzentrationen (Kalzium- oder Natriumchlorid mit 0,3 bis 0,4 M Endkonzentration) beobachtet. Zweiwertige Metallionen mit Konzentrationen ab 1 bis 2 mM Mangan und 0,5 mM Kobalt konnten die Aktivität ebenfalls um Faktor 2 steigern. Die beobachteten Effekte waren additiv, sodass mit idealen Konzentrationen aus Chlorid-Ionen und gleichzeitiger Inkubation mit zweiwertigen Metallionen rund 430 % der relativen Ausgangsaktivität erreicht werden konnte. Die Bindung von Kobalt und Mangan-Ionen durch Chelatierung durch EDTA-Zugabe machte den positiven Effekt rückgängig. Der aktivitätsstimulierende Effekt von zweiwertigen Metallionen auf Endoglucanasen ist in der Literatur bereits mehrfach beschrieben worden (Au und Chan, 1987; Kim *et al.*, 2012; Santos *et al.*, 2012; Saxena *et al.*, 1992). Der Mechanismus der Interaktion ist jedoch nur unzureichend erforscht. Im Falle der Cellulase BsCel5A aus *Bacillus subtilis* 168 konnte mittels Strukturanalyse die Rolle von Manganionen untersucht werden (Santos *et al.*, 2012). Dabei wurde auf der gegenüberliegenden Seite der aktiven Stelle die selektive Bindung der Ionen in einem elektrostatisch negativ-geladenen Bereich in einer azidischen Vertiefung beobachtet. Die Präsenz des Manganions hatte keine konformationelle Änderungen der Vertiefung oder des aktiven Zentrums zur Folge, jedoch konnte eine beachtliche Steigerung der Temperaturstabilität des Enzyms festgestellt werden (die Schmelztemperatur stieg von 64 auf 87 °C). Andere zweiwertige Kationen wie Magnesium- und Kalziumionen zeigten ebenfalls positive Wirkungen auf die Thermostabilität von Glycosidhydrolasen (Bronnenmeier und Staudenbauer, 1988). Außerdem können thermostabilisierende Effekte manchmal auch auf hohe Salzkonzentrationen zurückgeführt werden (Liebl *et al.*, 1996).

Tabelle 35: Beispielhafte Übersicht charakterisierter bakterieller und archaeeller, thermostabiler Glycosidhydrolasen der Familie 12.

Enzym / Ursprung	Molekulargewicht (kDa)	Aufreinigungsmethode(n)	pH-Optimum	Temperatur-optimum (°C)	Thermostabilität (Halbwertszeit)	Bevorzugtes Polysaccharid (Bindung)	Spezifische Aktivität (U/mg)	Referenz
Endoglucanase EglA ( <i>Pyrococcus furiosus</i> )	35,9	Hitzebehandlung, Anionenaustauscher, Gelfiltration	6,0	100	40 h, 95 °C	Glucan der Gerste ( $\beta$ -1,3/4)	58,0	Bauer <i>et al.</i> (1999)
Endo- $\beta$ -glucanase SSO1949 ( <i>Sulfolobus solfataricus</i> n <sup>o</sup> )	37,0	IMAC mit His-tag, Hitzebehandlung, Anionenaustauscher	1,8	80	160 min, 80 °C, pH 1,0	Längerkettige Cello- oligosaccharide, Carboxymethylcellulose	1,0 – 1,2	Huang <i>et al.</i> (2005)
Endo- $\beta$ -1,4-glucanase SSO1354 ( <i>Sulfolobus solfataricus</i> P2)	35,5/45,0	Anionenaustauscher, IMAC mit His-tag	2,5/4,0	80/90	180 min, 90 °C, pH 4,5	Lichenan ( $\beta$ -1,3/4)	1,2/223,5	Girfoglio <i>et al.</i> (2012)
CelStrep ( <i>Streptomyces</i> sp. G12, isoliert aus Kompost)	37,0	Hydrophobe interaktions- chromatographie (HIC)	4,8	50	24 h, 60 °C	Carboxymethylcellulose ( $\beta$ -1,4)	600	Amore <i>et al.</i> (2012)
CelA ( <i>Thermotoga maritima</i> MSB8)	29,8	Hitzebehandlung, Anionenaustauscher, HIC	5,0	90	3 h bis >48 h mit 2,5 M NaCl, 90 °C	Glucan der Gerste ( $\beta$ -1,3/4)	1.785	Liebl <i>et al.</i> (1996)
CelB ( <i>Thermotoga maritima</i> MSB8)	29,7	Hitzebehandlung, Anionenaustauscher, HIC	7,0	85-90	12 h bis 24 h mit 2,5 M NaCl, 90 °C	Glucan der Gerste ( $\beta$ -1,3/4)	2,91	Liebl <i>et al.</i> (1996)
Metagenomische Endo- $\beta$ - Glucanase Cel12E (ähnlich zu <i>Thermococcus</i> sp. AM4)	62,3	Hitzebehandlung, Anionenaustauscher, Gelfiltration	5,5	92	120 min, 92 °C	Carboxymethylcellulose ( $\beta$ -1,4)	700,4	diese Arbeit, Leis <i>et al.</i> (2015b)

Endoglucanasen sind für eine Vielzahl an biotechnologischen Prozessen unentbehrlich, und spielen insbesondere bei der Hydrolyse von lignocellulosehaltiger Biomasse zur fermentativen Erzeugung regenerativer Biokraftstoffe oder Ausgangsverbindungen für die chemische Industrie eine immer wichtigere Rolle. Zum effizienten Abbau pflanzlicher Polysaccharide bedarf es der konzertierten und synergistischen Hydrolyse durch verschiedenartige Glycosidhydrolase-Enzyme (Morais *et al.*, 2012). Die Eigenschaften von Cel12E für die Hydrolyse von derartigen Substraten müsste in Zukunft evaluiert werden.

### **3. Vergleichende Durchmusterung thermophiler Metagenombanken in *E. coli* und *T. thermophilus***

Eines der größten Hindernisse der funktionellen Metagenomik liegt darin begründet, dass die Zahl der verfügbaren Wirte zur heterologen Expression bis heute sehr gering ist. Standardwirte wie *E. coli* werden zwar am häufigsten zu diesem Zweck verwendet, jedoch ist das Bakterium kein „Alleskönner“. Es wird daher diskutiert, dass andere Wirte als *E. coli* die Schlagzahl neuer Gene und Genprodukte erhöhen können (Angelov *et al.*, 2009; Ekkers *et al.*, 2012; Taupp *et al.*, 2011).

In dieser Arbeit wurde als neuartiger Wirt *T. thermophilus* HB27 verwendet, der ebenso wie *E. coli* einige prinzipiellen Anforderungen als heterologer Expressionswirt erfüllt (modifiziert nach Angelov *et al.*, 2009): (i) Beide Bakterien sind genetisch zugänglich und erlauben die selektive Einführung und Aktivierung von geeigneten genetischen Elementen; (ii) Die Metagenombank muss in die beiden Wirte transformierbar sein, sei es mittels artifiziell-vermittelter oder natürlicher Kompetenz; (iii) Das Vektorsystem muss stabil im Wirt propagiert werden können, dabei darf aber das metagenomische Fragment nicht verloren gehen oder ungewollt modifiziert werden; (iv) Beide Wirte sollen einfach zu kultivieren sein und können unter bestimmten Inkubationsbedingungen auf neuartige Aktivitäten hin durchmustert und selektiert werden.

Eine besondere Rolle kommt der phylogenetischen und physiologischen Distanz zwischen der Donor-DNA und dem Expressionswirt zu. Es ist davon auszugehen, dass Sequenz-basierte Signale auf der DNA in einem nah verwandten Expressionwirt mit höherer Wahrscheinlichkeit erkannt und durch dessen Genexpressionsapparat effektiv in funktionelles Protein umgesetzt werden können. Im Zuge der funktionellen Durchmusterung einer Metagenombank, wäre es wünschenswert, solche Wirte einzusetzen, die den Organismen der Ausgangsprobe, aus der die DNA isoliert wurde, möglichst nahe stehen. Eine mögliche Strategie wäre es, die phylogenetische Zusammensetzung eines Habitats vor der Durchmusterung zu bestimmen, und einen entsprechend nahe verwandten Wirt auszuwählen, welcher für die funktionelle Metagenomanalyse angewendet werden kann (Leis *et al.*, 2013). Dennoch muss beachtet werden, dass dadurch weniger stark in den Konsortien vertretene Mikroorganismen

übergangen werden würden. In dieser Arbeit wurde die Anwendung von *T. thermophilus* zur Erstellung einer Genbank für den nah verwandten Stamm *T. Brockianus* durchgeführt.

### 3.1. Analyse der Genbank von *T. Brockianus*

Als beispielhafte Studie wurde eine Genbank des *Thermus*-Stammes *T. Brockianus*, dessen Genomsequenz unbekannt war, in das *Shuttle*-Fosmid pCT3FK kloniert und ein vergleichendes funktionelles *Screening* in *E. coli* und *T. thermophilus* durchgeführt (Leis *et al.*, 2013). Wie erwartet konnten im thermophilen Bakterium mehr aktive Klone identifiziert werden als im mesophilen Wirt *E. coli* (Tabelle 26). Insgesamt wurden je zwei hemicellulolytische Klone auf Xylan und AZCL-Galactomannan, zwei proteolytische Klone auf Magermilch sowie ein Klon mit einer lipolytischen Aktivität auf Tributyrinplatten identifiziert. Nach Sequenzierung ausgewählter Fosmide reichte die Sequenzinformation nur zur Assemblierung und Vorhersage von Kandidatengenomen zweier Fosmide, T17-1-D4 und T17-1-B9, aus. Der lipolytische Klon T17-1-D4 ergab ein Gen für eine  $\alpha/\beta$ -Hydrolase der Familie 5, welches ähnlich zu ORF TT\_C1275 in *T. thermophilus* HB27 war und für die beobachtete Aktivität verantwortlich sein könnte. Die vorläufige Charakterisierung dieses Proteins auf Fosmidebene ergab eine Substratpräferenz für kurzkettige *para*-Nitrophenyl-Substrate (Daten nicht gezeigt), was auf eine Esterase anstatt einer Lipase hindeutet. Die Sequenz des Fosmids des auf AZCL-Galactomannan aktiven Klons T17-1-B9 lag nicht vollständig sequenziert vor sondern teilte sich auf die beiden *contigs* 11 und 145 auf. Das kleinere *contig*-Fragment war mittels Endsequenzierung diesem Fosmidklon zuordenbar und kodiert für einen partiellen ORF mit Übereinstimmung zu TT\_C1540 im Stamm HB27. Die Annotation der Proteindomänen zeigte eine Glycosidhydrolase der Familie 57 (GHF 57) auf. Diese GH-Familie umfasst  $\alpha$ -Amylasen, 4- $\alpha$ -Glucanotransferasen,  $\alpha$ -Galactosidasen, Amylopullulanasen und 1,4- $\alpha$ -verzweigende Enzyme. Die beobachtete Freisetzung des chromogenen AZCL-Farbstoffes geht vermutlich auf die Hydrolyse von  $\alpha$ -glykosydischen Bindungen im Galactomannan zurück und nicht auf die Hydrolyse der  $\beta$ -Verbindungen. Galactomannan ist ein aus 1,4-verknüpften- $\beta$ -D-Mannopyranose-Molekülen bestehendes Polysaccharid, welches durch  $\alpha$ -D-Galactopyranose als Seitengruppen substituiert ist. Da nicht die vollständige Fosmidsequenz für weitere bioinformatische Analysen zur Verfügung steht, könnte der für die Aktivität verantwortliche ORF aber eventuell auch nicht identifiziert worden sein.

Die Verwendung von *T. thermophilus* HB27 als Expressionswirt für Genbanken thermophiler Mikroorganismen erlaubt die rasche Identifizierung biotechnologisch relevanter Enzyme, ohne dass die vollständige Genomsequenz zur Verfügung stehen muss. Ständig werden neue *Thermus*-Stämme isoliert, jedoch liegt bis heute nur ein Teil der Genome sequenziert vor. Da *Thermus* bereits als Quelle für anwendungsorientierte Enzyme genutzt wird (prominente Beispiele sind thermostabile DNA-

Polymerasen wie die *Taq*-Polymerase aus *T. aquaticus*), könnten sich auch weitere Anwendungsfelder mit biotechnologischem Nutzen ergeben (Liebl, 2004).

### 3.2. Konstruktion einer metagenomischen Fosmidbank aus heißen Quellen in *E. coli*

Die Konstruktion einer Metagenombank hängt in ganz besonderem Maße von der Qualität der isolierten Umwelt-DNA ab (Leis *et al.*, 2013). Dabei sollen die aufgereinigten Nukleinsäuren für spätere molekularbiologische Arbeiten (beispielsweise enzymatische Modifikationen, Klonierungen und Transformation) geeignet sein. Kriterien hierfür sind: (i) Die Reinheit der DNA-Präparation; (ii) die Fragmentengröße und der Grad der Scherung; (iii) Endmenge und Konzentration; (iv) die Diversität des genetischen Materials und Vermeidung von Methoden, die bestimmte Sequenzen bevorzugen (beispielsweise durch Amplifikation der Nukleinsäuren) und die Diversität der Probe stören (*bias*). Für die Erstellung von Fosmidbanken mit *Insert*-Größen zwischen 35 und 40 kb Länge ist die Isolierung hochmolekularer DNA unerlässlich. Dabei hängt der Erfolg für die Konstruktion einer Fosmidbank von der verwendeten Extraktionsmethode ab. Je nach Art des Probenmaterials (flüssig, fest) existieren physikalische, chemische und enzymatische Methoden, die einzeln oder in Kombination angewendet werden können. Meist kann *a priori* nicht die ideale Methode abgeschätzt werden, sodass die optimale(n) Extraktionsmethode(n) empirisch bestimmt werden müssen.

Für die Konstruktion der Fosmidbank aus Sedimenten, Biofilmen und Wasser der heißen Quellen wurden mehrere DNA Extraktionskits verwendet, welchen verschiedenen Extraktionsmethoden zugrunde liegen (Tabelle 7). Unabhängig von der Extraktionsmethode waren die beiden Proben AZ2 (Sedimente) und AZ4 (Biofilme) durch sehr geringe DNA-Ausbeuten und einem hohen Grad an gescherten Nukleinsäuren charakterisiert. Bei Probe AZ1 waren keine Nukleinsäuren in der Gelkontrolle und mittels DNA-Quantifizierung nachweisbar. Lediglich im Falle der Probe AZ3 (Wasser und Sediment) waren beachtliche Ausbeuten an reiner, hochmolekularer metagenomischer DNA erhalten worden.

Zur phylogenetischen Analyse des Probenmaterials wurden 16S rDNA-Sequenzen der größten Probe, AZ3, untersucht. Mit Temperaturen um 60 °C und äußerst saurem Milieu (pH 2,0) war eine äußerst geringe Diversität der mikrobiellen Gemeinschaften erwartet worden. Bei den bakteriellen Sequenzen konnten überwiegend Homologien zu mesophilen Gattungen festgestellt werden (Tabelle 31), darunter auch erstaunlicherweise häufig auftretende human- und tierpathogene Vertreter (*E. coli* und *Shigella* machten etwa die Hälfte der bakteriellen 16S-Zuordnungen aus). Ihr Nachweis ist womöglich der ungewollten Amplifikation mittels PCR geschuldet, die selbst geringe Kopienzahlen dieser 16S rDNA nachweisen können. Aufgrund der extremen Acidität und den hohen Temperaturen der heißen Quelle, kann davon ausgegangen werden, dass die Mehrzahl der nachgewiesenen bakteriellen Spezies in

diesem Habitat nicht lebensfähig gewesen sein können. Grund für die Präsenz der Mikroorganismen dürfte hier unter anderem die Nutzung der heißen Quelle für das Garen von Gemüse sein. Eine möglicherweise niedrige Diversität zeigte sich bei den archaeellen Vertretern, denn es wurde lediglich eine Sequenz eines Crenarchaeons mit 99 % Sequenzhomologie zu *Stygiolobus azoricus* DSM 6296 gefunden. Dieser Mikroorganismus zählt zu der Gruppe der Sulfolobales, welche häufig in stark bis schwach sauren, schwefelhaltigen und äußerst heißen Quellen vorkommen können (Madigan *et al.*, 2001). Die Gattung *Stygiolobus* ist wie viele andere dieser Gruppe durch einen relativ niedrigen GC-Gehalt (38 %) charakterisiert, die optimalen Wachstumsbedingungen liegen bei etwa 80 °C und einem pH-Wert von 3,0. Mehrere Isolate wurden ebenfalls aus heißen Quellen von der Azoren-Insel São Miguel (Portugal) isoliert (Segerer *et al.*, 1991). Die mikrobielle Diversität insbesondere in heißen Habitaten kann extrem unterschiedlich stark ausgeprägt sein (Sturner und Liebl, 2001). Trotz der eingeschränkten Probenzahl zur phylogenetischen Analyse der Klone lässt sich eine niedrige archaeelle Diversität vermuten. Dennoch konnten aus den Umweltproben (insbesondere Probe AZ3) genug hochmolekulare Nukleinsäuren extrahiert werden, um eine *large-insert* Fosmidbank für metagenomische Durchmusterungen zu konstruieren.

### **3.3. Funktionelle Durchmusterung thermophiler Metagenombanken in beiden Wirten**

Funktionelle Durchmusterungen metagenomischer Klonbanken erlaubten bereits wiederholt die Identifizierung neuer lipolytische Enzyme aus Umwelt-DNA, wobei thermophile Habitate interessante Quellen für neuartige Esterasen sind (Tabelle 36). Meist werden sogenannte direkte Detektionsmethoden verwendet, wie im Falle von Substratplatten mit trüben Emulsionen von Tributyrin in wässrigem Medium. Der Nachweis lipolytischer Enzyme geschieht über die Bildung von hydrolytischen Höfen um die betreffende Kolonie. Diese Technik setzt die Nutzung eines Wirts voraus, der keine oder vernachlässigbare eigene Aktivität besitzt, wie beispielsweise *E. coli*.

Da *Thermus*-Arten starke eigene lipolytische Aktivitäten ausbilden, war die Entwicklung einer Mutante mit verminderter Hintergrundaktivität notwendig. Die Herstellung der Vierfachmutante BL03 stellt damit eine wichtige Verbesserung des Wirts *T. thermophilus* HB27 zur Durchmusterung von Metagenombanken nach Genen für Esterasen/Lipasen dar. Nur durch diese Mutante wurde eine Komplementations- und Selektionsstrategie ermöglicht.

**Tabelle 36: Beispiele für thermostabile Esterasen und Lipasen, die aus metagenomischen Durchmusterungen gewonnen werden konnten, im Vergleich mit dieser Arbeit.**

Herkunft der Metagenomprobe	Vektor	Insert-Größe (kbp)	Größe der Bank (Klone)	Positive Klone (charakt. Hydrolasen)	Nächstes Homolog	Referenz
Heiße Quellen, Sileri (Indonesien)	Fosmid	30-50	2.000	1	<i>Pyrobaculum calidifontis</i>	Rhee <i>et al.</i> (2005)
Heiße Quellen, Jae Sawn (Thailand)	Plasmid	1-10	36.000	2 (2)	<i>Pseudomonas mendocina</i> , <i>Chloroflexus aurantiacus</i>	Tirawongsoroj <i>et al.</i> (2008)
Marines Sediment, Süd-chinesisches Meer	Plasmid	1-8,5	60.000	15 (1)	Unkultivierter Organismus, <i>Clostridium beijerinckii</i>	Peng <i>et al.</i> (2011)
Anreicherungen aus Heizwasser, Hamburg (Deutschland)	Cosmid	27,5	576	4 (1)	<i>Thermus scotoductus</i>	Chow <i>et al.</i> (2012)
Anreicherungen aus Boden- und Wasserproben (Deutschland)	Cosmid	27,5	6.500	6 (1)	<i>Symbiobacterium thermophilum</i>	Chow <i>et al.</i> (2012)
Kompost aus Holzabfällen, Göttingen (Deutschland)	Fosmid	28-40	2.000	2 (1)	<i>Sphingobium quisquiliarum</i>	diese Arbeit, Leis <i>et al.</i> (2015a)
Heiße Quellen, Azoren (Portugal)	Fosmid	28-40	6.000	6 (1)	<i>Acidianus hospitalis</i>	diese Arbeit, Leis <i>et al.</i> (2015a)

Die heterologe Komplementation nutzt dabei das Konzept, einen genetischen Defekt unter definierten und selektiven Bedingungen wiederherzustellen. Dabei bedarf es der funktionellen Expression von Genen für entsprechende Enzyme (im hier gezeigten Fall Esterasen und Lipasen) aus den metagenomischen Fragmenten, um das vorgegebene Substrat (in diesem Fall Tributyrin) zu verstoffwechseln. Letztlich können nur diejenigen Fosmidklone Kolonien ausbilden, die auch in der Lage sind, Tributyrin mittels Tributyrasen abzubauen. In der hier vorliegenden Arbeit konnte gezeigt werden, dass die multiple Mutante *T. thermophilus* BL03 gegenüber dem Wildtyp nicht mehr in der Lage war, auf dem Substrat zu wachsen (siehe Kapitel IV.1.4.3.). Ferner zeigte dieser Stamm auch eine weniger stark ausgeprägte Hofbildung auf Substratplatten mit Tributyrin, sodass eine klassische Durchmusterung wie bei *E. coli* ebenfalls möglich war.

Von insgesamt etwa 8.000 Fosmidklonen wurden bei 60 °C Inkubation 7 positive Kandidaten identifiziert, was etwa einer Ausbeute von 0,1 % entspricht (ein Positiver unter 1.000 durchmusterten Klonen). In dem mesophilen Wirt *E. coli* konnten davon lediglich zwei Klone klassisch über die Ausbildung von Hydrolysehöfen identifiziert werden: Je einen aus der Bank der heißen Quellen (AZ2-4-B6) und einen aus der Kompostprobe (M12-4-D9).

In *T. thermophilus* wurden nicht Einzelklone sondern gemischte Fosmidpools aus je 96 Einzelfosmiden (gegliedert nach der Identifikationsnummer der Mikrotiterplatten) verwendet, um Zeit- und Material-intensive Einzeltransformationen im Plattenformat zu umgehen. Da aus jeder Fosmidpool-Transformationsreaktion mehrere Hundert Kolonien auf den Selektivmedien beobachtet wurden, konnte davon ausgegangen werden, dass der Großteil aller Einzelfosmide je Fosmidgemisch statistisch abgedeckt sein musste. Die gewachsenen Kolonien wurden von den Agarplatten abgewaschen und mehrmals in Puffer gewaschen, um übrige Medienbestandteile des Vollmediums und lipolytische Enzyme des Überstandes zu entfernen. Dieser Arbeitsschritt war äußerst wichtig, denn ohne die Waschschrte vermochte die Vierfachmutante BL03 ebenso wie der Wildtyp durch Vollmediumreste auf den Minimalmedien mit Tributyrin zu wachsen. Die beobachteten Unterschiede zwischen den Stämmen zeigten sich erst durch wiederholtes Abzentrifugieren und Resuspendieren der Zellen zum Abwaschen möglicher Medienbestandteile (Abbildung 15). Durch die oben beschriebene Methode konnten zunächst 21 Pools identifiziert werden, welche sichtbares Wachstum auf den Minimalmedien zeigten. Nach mehrmaligen Passagieren der Klone auf den Minimalmedien mit Tributyrin reduzierte sich die Zahl der Kandidaten auf nun mehr etwa die Hälfte. Auch dieser Kontrollschritt war notwendig, um etwaige falsch-positive Fosmide auszuschließen und stabile Phänotypen zu selektieren (z.B. wie im Fall von Fosmid AZ4-2-H10, welches nachweislich ein Leerfosmid war). Von diesen Kandidaten wurden schließlich einzelne Transformationen im Blockformat durchgeführt, um die positiven Fosmide ihren entsprechenden Positionen in den Mikrotiterplatten zuordnen zu können. Die Selektion auf den Minimalmedien mit Tributyrin lieferte hiervon 6 Fosmide (benannt nach den Kennzeichnungen auf den Mikrotiterplatten sowie ihre

korrespondieren Positionen: AZ2-4-B6, AZ3-14-D2, AZ3-14-D11, AZ3-32-E5, AZ3-33-C12, und M12-4-G10). Rohextrakte aus diesen Fosmidkandidaten zeigten gegenüber der Kontrolle BL03 mit Leerfosmid gesteigerte lipolytische Aktivitäten. Dabei war die Substratspezifität gegenüber kurzkettigen Acylestern bevorzugt. Einzige Ausnahme war Klon AZ3-14-D2, welcher über alle getesteten Substrate hinweg die höchsten Aktivitäten zeigte (Abbildung 38). Das lediglich in *E. coli* identifizierte Fosmid M12-4-D9 führte zwar nicht zu einem Wachstum in *T. thermophilus*, jedoch waren signifikante Aktivitäten gegenüber *pNP-C<sub>8</sub>* Substrat nachweisbar. Dies kann darin begründet liegen, dass dieses Fosmid für ein Enzym kodiert, welches derivatisierte Fettsäureester wie *pNP* besonders effizient hydrolysieren kann, jedoch gegenüber Triglyceriden keine ausreichende Aktivität besitzt und damit für die Komplementationsanalyse nicht angewendet werden kann.

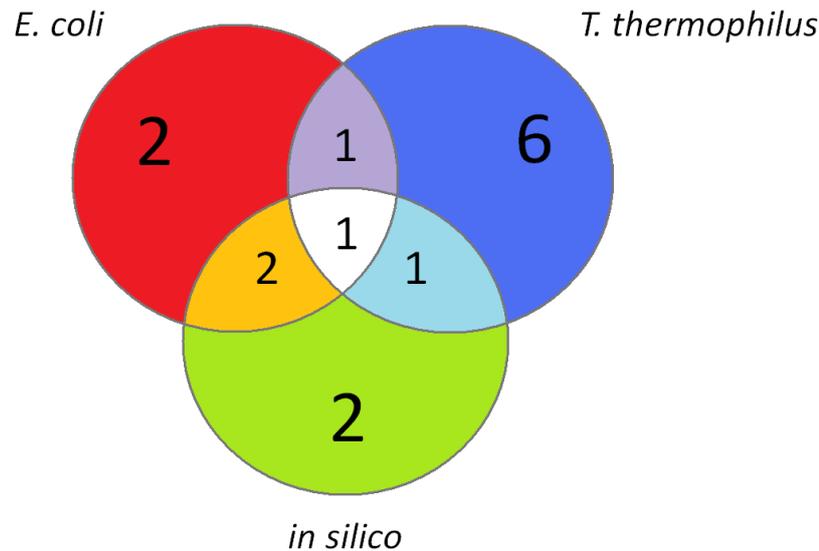
Die Kombination aus direkter Detektion und heterologer Komplementation führte zur erfolgreichen Durchmusterung der beiden Metagenombanken in *E. coli* und *Thermus*. Der Vorteil der heterologen Komplementation über die direkten Detektionsmethoden wurde bereits in anderen metagenomischen Studien gezeigt. Dies betrifft Proteine, die nicht direkt nachgewiesen werden können oder nur in geringen Mengen exprimiert werden müssen, wie beispielsweise Metabolit-Transporter (Damon *et al.*, 2011), zentrale Stoffwechsellenzyme (Kellner *et al.*, 2011; Parachin und Gorwa-Grauslund, 2011), sowie konditionell lethale Genprodukte wie die thermosensitive DNA Polymerase I (Simon *et al.*, 2009) oder Proteine, die Resistenzen gegenüber Antibiotika oder toxische Verbindungen bewirken (Uyaguari *et al.*, 2011; Sommer *et al.*, 2010). Nachteilig ist jedoch, dass der Wirtsorganismus die gesuchte Eigenschaft nicht besitzen darf, bzw. entsprechende Mutantenstämme (in dieser Arbeit der multiple Deletionsstamm *T. thermophilus* BL03) hergestellt werden müssen, ehe es zur Anwendung kommen kann.

### 3.4. Sequenzierung und *in silico* Analyse der Fosmide aus positiv getesteten Klonen

Im Vergleich zur Sequenzierung eines Gemisches an Fosmiden, wie es im Projekt *ANG48* der Fall war, zeigte sich der Vorteil des *Multiplexings* als Indizierungsmethode für 454-Sequenzierungen (Projekt *EST6*, Zusammenfassung der Assemblierung in Tabelle 44). Durch das Anfügen eines Fosmid-spezifischen *tags* an jedes Amplicon, konnte die assemblierte Sequenzinformation direkt den jeweiligen Fosmidklonen zugeordnet werden. Selbst nicht vollständig geschlossene Fosmide ließen sich eindeutig zuordnen.

Die Identifizierung von neuartigen lipolytischen Enzymen und Genprodukten mittels gängigen bioinformatischen Methoden wie beispielsweise Ähnlichkeitssuchen (*blastp*) und Klassifikationen von Proteindomänen und Familien (laut Pfam) zeigte geringere Ausbeuten als durch funktionsbasierte Durchmusterungsstrategien (Abbildung 52). Darüberhinaus wurden in *E. coli* lediglich ORFs

aufgefunden, die für Proteine mit hohen Ähnlichkeiten zu bereits in den öffentlich zugänglichen Datenbanken abgelegten bzw. vollständig charakterisierten Enzymen kodieren. Im Gegensatz dazu waren die in *T. thermophilus* erhaltenen Kandidatensequenzen größtenteils nicht annotiert (hypothetische Proteine) und wiesen wesentlich geringere Ähnlichkeiten zu bekannten Enzymen auf.



**Abbildung 52:** Graphische Repräsentation der vergleichenden Durchmusterungsergebnisse für lipolytisch-aktive Fosmide zwischen *E. coli*, *T. thermophilus* BL03 und bioinformatischen Methoden (schematisches Mengendiagramm). Den aufgeführten Daten (roter Kreis: *E. coli*; blauer Kreis: *T. thermophilus*; grüner Kreis: *in silico* Analyse) liegen die Auswertungen der Durchmusterungen und Aktivitätsnachweise zu Grunde, sich überschneidende Kreisflächen deuten die identischen Fosmide an. Bei den bioinformatischen Ergebnissen wurden nur sequenzierten Fosmide und *shotgun*-Fragmente als positiv gewertet, wenn die Kandidatengene *blastp*-Homologien über 50 % zu charakterisierten Proteinen besitzen und charakteristische Domänen laut Pfam-Suche identifiziert werden konnten (z.B.  $\alpha/\beta$ -Hydrolasen).

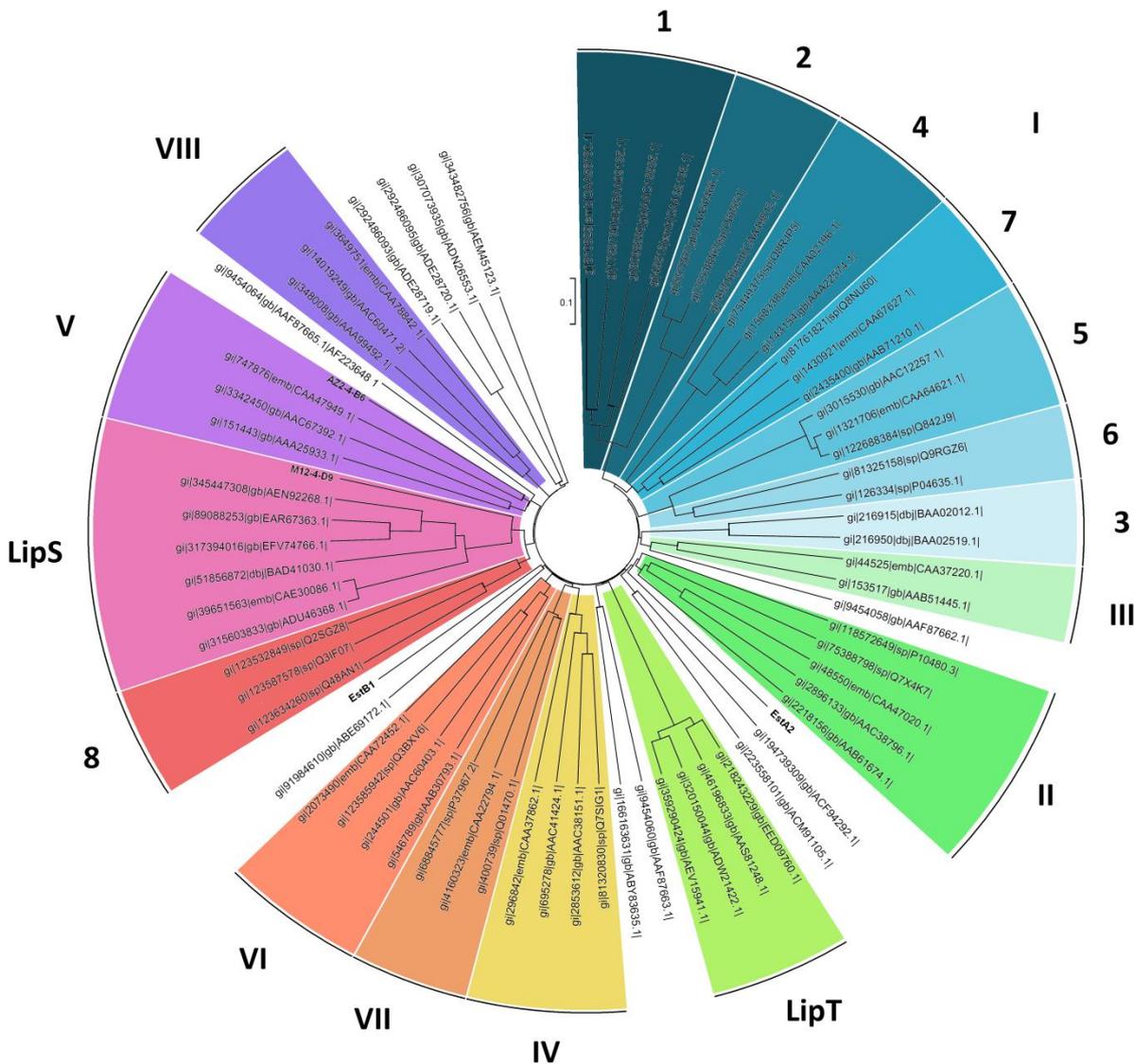
Bakterielle lipolytische Enzyme lassen sich aufgrund ihrer Aminosäuresequenz und Faltung unterscheiden. Die sogenannte  $\alpha/\beta$ -Hydrolase-Faltung findet man bei hydrolytischen Enzymen, die sowohl phylogenetisch als auch in ihrer katalytischen Funktion unterschiedlich sind. Die katalytische Domäne kommt in verschiedenartigen Hydrolasen vor, wie beispielsweise in Proteasen, Dehalogenasen sowie Esterasen und Lipasen. Ihnen gemeinsam ist eine Grundstruktur bestehend aus einem  $\alpha/\beta$ -Faltblatt (8 Stränge, die jeweils über Helices miteinander verbunden sind). Allen Vertretern dieser Enzymklasse ist auch die Präsenz einer konservierten katalytischen Triade gemein. Das charakteristische Merkmal ist ein Serin, welches als Nukleophil für die Hydrolyse des Substrats notwendig ist. Der Reaktionsmechanismus umfasst die Bildung eines Enzym-Acyl-Intermediates am Serinrest, die Spaltung der Esterbindung im Intermediat durch Wasser/Hydroxidion und die abschließende Regeneration des Enzyms. Die Aktivierung des Serins umfasst die Präsenz eines sauren

Aminosäurerests (beispielsweise Aspartat oder Glutamat) sowie eine basische Aminosäure, meist Histidin. Das katalytische Serin tritt dabei normalerweise im Zusammenhang mit einem konservierten Pentapeptidmotiv G-X-S-X-G auf. Dieses Motiv war ebenfalls in allen vorhergesagten Proteinkandidaten vorhanden (Abbildung 53).

Fosmid (ORF)	Pfam Proteinfamilie	Sequenz	Aktivität
AZ2-4-B6 (8c)	$\alpha/\beta$ -Hydrolase	98 AQAHLV <b>GNS</b> FGGALALA 114	vorhergesagt ( <i>screening</i> )
M12-4-D9 (1)	$\alpha/\beta$ -Hydrolase	91 EEI <b>VAGLS</b> LGGVFALK 107	bestätigt ( <i>shotgun</i> -Bank)
M12-4-G10 (18c)	Bla, Alk-Sulf	493 FHCTDT <b>GASY</b> GLAIRSD 509	bestätigt (EstB1)
AZ3-33-C12 (2)	unklassifiziert	59 EKILNS <b>GLSD</b> GRKRFIL 75	bestätigt (EstA2)
AZ3-32-E5 (12)	Prolyl-Oligopeptidase	404 KKVA <b>VTGAS</b> YGGYLTMM 420	vorhergesagt (EstS9)
AZ3-14-D2 (2)	Thiolase	233 VWIV <b>SQGVSS</b> GFANLSR 249	vorhergesagt ( <i>screening</i> )
* * *			

**Abbildung 53: Sequenzvergleich der putativen Kandidatenproteine sowie experimentell bestätigter Esterase-aktiver Enzyme. Gezeigt sind die für lipolytische Proteine konservierte Pentapeptidmotive G-X-S-X-G, markiert als dicke Buchstaben mit Sternen unter dem 16 Aminosäuren langen Alignment. Abkürzungen: Bla = Metallo- $\beta$ -Lactamase; Alk-Sulf = Alkyl-Sulfatase.**

Die Identifizierung von lipolytischen Kandidatenproteinen war nur für zwei Fosmidsequenzen möglich (AZ2-4-B6 und M12-4-D9), welche jeweils für klassische  $\alpha/\beta$ -Hydrolasen der Familie 6 kodierten und jeweils bakteriellen Ursprungs sind (Tabelle 57 und Tabelle 58, siehe Anhang). Beide lipolytischen Enzyme sind beim *Screening* in *E. coli* aufgefunden worden und waren jeweils der Carboxylesterase-Familie V und LipS (Arpigny und Jaeger, 1999; Hausmann und Jaeger, 2010; Chow *et al.*, 2012) zuordenbar (Abbildung 54). Für die restlichen Kandidaten gestaltete sich die Suche als schwierig, da einerseits nicht alle Fosmide vollständig sequenziert vorlagen, andererseits jedoch selbst in geschlossenen *contigs* keine Übereinstimmungen zu bekannten Enzymfamilien ersichtlich waren. Mittels Motivsuchen über reguläre Ausdrücke (*regular expressions*) konnte in den translatierten Sequenzdaten in allen drei Leserahmen nach dem für Serin-Hydrolasen charakteristischen und konservierten Pentapeptid-Motiv G-X-S-X-G gesucht werden (Abbildung 53). Die dadurch identifizierten Kandidatensequenzen wurden ausgiebig mittels *blastp* und der Pfam-Datenbank überprüft und lieferten mehrere Anhaltspunkte.



**Abbildung 54: Eingruppierung klassifizierter Esterase- und Lipase-Familien aufgrund Sequenzähnlichkeiten (*similarity tree*). Neben den klassischen Protein-Familien I bis VIII (inkl. Subfamilien 1-8) (Arpigny und Jaeger, 1999; Hausmann und Jaeger, 2010) existieren zusätzlich postulierte Familien LipS und LipT sowie fünf nicht zuordenbare Familien (UF, im Baumdiagramm ohne farblichen Hintergrund) (Chow *et al.*, 2012). Die Proteinsequenzen aus den vergleichenden Metagenomdurchmusterungen M12-4-D9, AZ2-4-B6 sowie die beiden charakterisierten Esterase-aktiven Enzyme der Fosmide EstA2 und EstB1 sind im Baumdiagramm markiert (dicke Schrift). Das multiple Sequenzalignment von Vertretern aller bekannter Familien (Tabelle 41) und der neuen, metagenomischen Sequenzen wurden mit ClustalW (T-REX web server, Boc *et al.*, 2012) erstellt. Der Baum wurde mittels Neighbor-Joining berechnet und im Newick-Format mit MEGA (Version 5.2, Tamura *et al.*, 2011) visualisiert. Die Länge der Verzweigungen gibt die Unterschiede der Proteinsequenzen zueinander wieder. Modifizierte Abbildung nach Leis *et al.*, (2015a).**

Die DNA-Insertion in Fosmid AZ3-33-C12 war archaeellen Ursprungs und zeigte höchste Homologien zu Vertretern der Ordnung der Sulfolobales, welche ebenfalls aus 16S rDNA-Analysen der Proben aus den heißen Quellen beobachtet worden waren. Die *in silico*-Suche nach potentiellen Kandidatengenen war für ORF Nummer 2 möglich, welcher für das kleine Protein EstA2 kodiert (Tabelle 62). Neben größtenteils hypothetischen Proteinen (höchste Übereinstimmung mit Ahos\_0948 aus dem Archaeon *Acidianus hospitalis* W1) konnten ebenfalls Sequenzübereinstimmungen mit einer

charakterisierten Phospholipase A<sub>2</sub>/Esterase aus *Aeropyrum pernix* K1 gezeigt werden (Wang *et al.*, 2004). EstA2 zeigte das G-X-S-X-G Motiv, ein Signalpeptid konnte nicht vorhergesagt werden. Ihre Funktion konnte außerdem noch nicht zugeordnet werden. Die Klassifizierung von EstA2 auf Basis bekannter Esterasefamilien I bis VIII (Arpigny und Jaeger, 1999) war ebenfalls nicht möglich. Selbst neue Familien LipS und LipT sowie unbekannte Familien UF1 bis UF5 (Chow *et al.*, 2012) lieferten im multiplen Sequenzalignment keine ausreichende Übereinstimmung. Selbst die aufgrund der Sequenz-Überstimmungen naheliegende Klassifikation als Phospholipase A<sub>2</sub> muss abgelehnt werden, da EstA2 kein konserviertes C-C-X-X-H-X-X-C Motiv besitzt (Wang *et al.*, 2004). Die Insertionen der drei weiteren Fosmide AZ3-14-D2, AZ3-14-D11 und AZ3-32-E5 waren ebenfalls archaeellen Ursprungs und zeigten Ähnlichkeit zu Vertretern der Sulfolobales (Tabelle 59 bis Tabelle 61, Anhang). Obwohl die Fosmide nicht vollständig sequenziert vorlagen, wurde versucht mittels Motiv- und Homologiesuchen Kandidatengene abzuleiten, die für die beobachteten Aktivitäten verantwortlich sein konnten.

Ein auf Fosmid AZ3-32-E5 gefundener ORF kodiert für EstS9, das höchste Ähnlichkeit gegenüber einer Dipeptidyl-Amino-peptidase / Acyl-Aminoacyl-Peptidase aus *Sulfolobus islandicus* LAL14/1 aufwies (Tabelle 61). Laut Pfam-Datenbank gehört dieses Protein zur Familie der Prolyl-Oligopeptidasen (Peptidase\_S9), welche strukturelle und funktionelle Ähnlichkeit mit mehreren  $\alpha/\beta$ -Hydrolasen, Carboxyl-Esterasen und anderen Esterasen besitzt. Insbesondere die Präsenz eines konservierten Pentapeptidmotifs G-X-S-X-G spricht für eine Serin-Hydrolase. Die katalytische Triade war für die Aminosäurereste Serin S412, Aspartat D491 und Histidin H523 vorhersagbar. Dass diese Enzymfamilie sowohl über Peptidase als auch Esterasefunktionen verfügt konnte anhand der Acyl-Aminoacyl-Peptidase aus *Aeropyrum pernix* K1 gezeigt werden (Selvarani *et al.*, 2004; Wang *et al.*, 2006).

Fosmid AZ3-14-D2 wurde nicht vollständig sequenziert. Die bioinformatische Untersuchung der assemblierten *contigs* zeigte mit 53,3 % einen hohen Anteil an Proteinen, die laut *gene-ontology* (GO)-Klassifikation am Fettsäuremetabolismus beteiligt sein dürften. Die gefundenen ORFs mit Koordinaten, Homologien, Annotationen und Pfam-Klassifizierungen sind in Tabelle 59 aufgelistet. Da *contig* 1 zum größten Teil aus pCT3FK-Fosmidsequenz assembliert worden war, konnte nur ein Protein aus der metagenomischen *Insert*-DNA vorhergesagt werden. *contig* 2 und *contig* 3 hingegen weisen mehrere Gene und kodierende Genprodukte auf, die mit Fettstoffwechsel und dem Abbau von Fettsäuren mittels  $\beta$ -Oxidation in Zusammenhang gebracht werden können. Beispiele sind Acyl-CoA-Dehydrogenasen, weitere FAD-bindende Elektronentransfer-Flavoproteine (ETFs), Fettsäure-CoA-Ligasen (*long-chain fatty acid CoA-ligase*, ACSL), Thiolase bzw. Acetyl-CoA-Acetyltransferase (ACAT) und 3-Hydroxyacyl-CoA-Dehydrogenase. Durch Motivsuche konnte eine weitere konservierte Pentapeptidsequenz für eine Thiolase (ORF Nummer 2) aus *contig* 2 aufgefunden werden, welche als weiterer plausibler Kandidat vorgeschlagen werden könnte (Motiv **GVSSG**).

Von Fosmid AZ3-14-D11 konnte ebenfalls keine vollständige Sequenz erhalten werden. Von der verfügbaren metagenomischen Sequenzinformation wurden weder konservierte Motive noch Übereinstimmungen zu annotierten Esterasen aufgefunden werden (Tabelle 60).

Ein weiterer aussichtsreicher ORF in Fosmid M12-4-G10 kodierte für ein Protein EstB1, welches höchste Übereinstimmung zu einem hypothetischen Protein aus dem Bakterium *Sphingobium quisquiliarum* P25 zeigte (Tabelle 63). Die Ähnlichkeit auf Sequenzbasis fiel gering aus, ferner waren Sequenzen zu mehreren Lactamasen verschiedener *Vibrio*-Spezies feststellbar, die teilweise mit Infektionskrankheiten an marinen Organismen in Verbindung stehen (z.B. *V. breoganii* oder *V. splendidus*) (Beaz *et al.*, 2009; Jensen *et al.*, 2003). Erwähnenswert ist in diesem Zusammenhang auch, dass die Mehrheit der vorhergesagten ORFs des Fosmids M12-4-G10 für Proteine kodieren, die sehr starke Übereinstimmungen (zwischen 90 und 99 %) zu *Sphaerobacter thermophilus* DSM 20745 zeigen. Dieses Bakterium gehört zum Phylum der *Chloroflexi* und wurde ursprünglich aus aeroben und thermophilen Schlamm isoliert (Hugenholz und Stackebrandt, 2004). Aus den phylogenetischen Analysen der Kompostprobe waren Vertreter der *Chloroflexi* zwar nachweisbar, jedoch innerhalb der mittels 16S rDNA nachweisbaren mikrobiellen Konsortien in der Probe stark unterrepräsentiert (< 1,0 %, M. Mientus, unveröffentlichte Daten).

EstB1 enthält die eine Domäne der B Metallo- $\beta$ -Lactamase Superfamilie, zusätzlich waren C-terminal je eine Alkylsulfatase mit C-terminaler und Dimerisierungs-Domäne mittels Pfam vorhergesagt worden. Die Faltung der Metallo- $\beta$ -Lactamasen Proteinfamilie kommt neben klassischen Zink-abhängigen  $\beta$ -Lactamasen auch in Thioester-aktiven Hydrolasen und anderen Proteinen vor (Carfi *et al.*, 1995). Darunter fallen beispielsweise Thiolesterasen (EC 3.1.2.7) und Glyoxalase II (EC 3.1.2.6). Im Zuge der Klassifizierung von mikrobiellen Carboxylesterasen wurden  $\beta$ -Lactamasen der Familie C bereits zur Familie VIII zugeordnet (Arpigny und Jaeger, 1999). Sie können aufgrund des S-X-X-K Sequenzmotivs identifiziert werden, wobei ein G-X-S-X-G des nukleophilen Serinrests nicht vorhanden ist. Bereits in mehreren metagenomischen Studien konnte gezeigt werden, dass Carboxyl-Esterasen der Familie VIII in der Lage sind, neben Tributyrin und anderen Fettsäure-Estern auch  $\beta$ -Lactam Antibiotica zu hydrolysieren (Jeon *et al.*, 2011; Mokoena *et al.*, 2013). EstB1 zeigte weder das S-X-X-K Motiv, noch konnte das Protein auf Basis bekannter Esterasefamilien I bis VIII klassifiziert werden (Arpigny und Jaeger, 1999).

Die weiteren vorhergesagten Domänen von EstB1 gehören zur Enzymfamilie der Alkylsulfatasen. Diese spalten aliphatische Sulfatesterbindungen und vermögen ebenfalls Natriumdodecylsulfat (SDS), ein Protein-denaturierendes Tensid abzubauen. Neben der N-terminalen katalytisch aktiven Domäne dient der C-terminale Teil dabei als eine hydrophobe und Schacht-förmige Struktur, welche zur Rekrutierung von hydrophoben Substraten verwendet wird (Hagelueken *et al.*, 2006). Das ursprünglich untersuchte Protein SdsA1 (Kristallstruktur PDB Nummer 2CG3\_A) stammte aus *Pseudomonas aeruginosa*, welches zu EstB1 32 % Sequenzidentität hat (*e-value*:  $4 \times e^{-73}$ ).

### 3.5. Expression und vorläufige biochemische Charakterisierung ausgewählter Esterasen

Aufgrund der bioinformatisch vorhersagbaren Kandidatengene wurden drei archaelle Proteine, EstS9, EstB1 und EstA2 parallel in *E. coli* und *T. thermophilus* für vergleichende Expressionsstudien untersucht. Da die für *Thermus* etablierten Expressionsvektoren bereits in Kapitel V.1.1. vergleichend getestet wurden, beschränkt sich dieser Abschnitt lediglich auf die heterologe Expression im pET21a(+)-Expressionsvektor für *E. coli* BL21.

Die beiden Proteine EstA2 und EstB1 waren jeweils unter Standardexpressionbedingungen in ausreichenden Mengen mittels SDS-PAGE nachweisbar. Die Aufreinigung über immobilisierte Metallionen Affinitätschromatographie (IMAC) konnten sowohl mittels Gravitationsdurchfluss (Protino Nickel-Iminodiessigsäure) als auch mit einer FPLC (HisTrap Nickel-Sepharose) erfolgreich durchgeführt werden. Im Falle von EstB1 waren je nach Fraktion leichte Verunreinigungen mit nicht-spezifischen Proteinen aufgetreten.

Die Unterscheidung von Carboxylesterasen (EC 3.1.1.1) und Lipasen (EC 3.1.1.3) basiert auf einigen postulierten Kriterien: (i) Die katalytische Aktivierung in der Anwesenheit von Öl-Wasser-Grenzflächen mittels sogenannter Grenzflächenaktivierung (*interfacial activation*), (ii) das Vorhandensein einer Abdeckung, welcher die katalytisch aktive Stelle im geschlossenen Konformationszustand verdeckt, und (iii) die Substratspezifitäten des Enzyms (Verger, 1997; Jaeger *et al.*, 1999). Letzterer Punkt ist ein anerkanntes Unterscheidungsmerkmal, insbesondere, da für die ersten beiden Merkmale (i) und (ii) zahlreiche Ausnahmefälle beobachtet worden sind. Lipasen bevorzugen langkettige Fettsäureester-Verbindungen mit mindestens 10 Kohlenstoffatomen während Esterasen kurzkettige Acylester hydrolysieren können. Dennoch ist die Definition für „lang- oder kürzerkettige“ Acylester nicht genauer definiert (Jaeger *et al.*, 1999). Die Charakterisierung der lipolytischen Aktivitäten von EstA2 und EstB1 zeigte ähnliche Substratpräferenzen gegenüber kurzkettigen Acyl-Estern (*p*NP-C<sub>3</sub> bis *p*NP-C<sub>4</sub>). Im Gegensatz dazu waren die Enzyme auf längerkettigen Fettsäureester-Substraten über 10 Kohlenstoffatome verhältnismäßig gering aktiv. Diese Beobachtungen stützen die Vermutung, dass es sich bei beiden Proteinen eher um Esterasen als Lipasen handelt. Das Temperaturoptimum lag bei 75 °C (EstB1) und 80 °C (EstA2), bei optimalen pH-Werten um pH 8,0 (EstB1) bis 8,8 (EstA2). Der markanteste Unterschied zwischen beiden Proteinen zeigte sich auf Ebene der Gesamtaktivität ( $44,05 \pm 2,06 \text{ U} \times \text{mg}^{-1}$  für EstA2 und  $4,13 \pm 0,71 \text{ U} \times \text{mg}^{-1}$  für EstB1 auf *p*NP-butyrat als Substrat). Der Einfluss einiger Salze wie NaCl, KCl, MgCl<sub>2</sub> und CaCl<sub>2</sub> in Endkonzentrationen zeigte bei beiden Enzymen keine signifikanten Aktivitätsänderungen. Inkubationen mit dem Chelator EDTA konnten die Aktivitäten ebenfalls nicht beeinträchtigen. Der Einsatz von Phenylmethylsulfonylfluorid (PMSF), einem effektiven Protease-

Inhibitor zeigte keine Wirkung auf die Aktivität von EstB1. Da PMSF irreversibel an das Serin des aktiven Zentrums bindet und damit Serin-haltige Hydrolasen selektiv hemmt, erlaubt seine Verwendung Rückschlüsse über den Reaktionsmechanismus des Enzyms. Dies bedeutet für EstB1 erwartungsgemäß, dass das Enzym keine Serin-abhängige Hydrolase ist. Für EstA2 konnte gezeigt werden, dass erhöhte Zinkchlorid-Zugaben (10 mM Endkonzentration) nur geringen Einfluss auf die Enzymaktivität ausübten (Rückgang auf 85,2 % der Ausgangsaktivität), wohingegen das Enzym APE2325 aus *Aeropyrum pernix* K1 (das charakterisierte und nächst-ähnliche Enzym zu EstA2) bei selbst geringen ZnCl<sub>2</sub>-Supplementationen (1 mM) vollständig inaktiviert wird (Wang *et al.*, 2004). Die vollständige Charakterisierung der Enzyme steht noch aus. Geklärt ist momentan ebenfalls noch nicht, auf welchen Substraten EstA2 und EstB1 ihre Hauptaktivität entfalten. Da beide Enzyme zu überwiegend hypothetischen Proteinen ähnlich sind, könnten die beobachteten Aktivitäten auf Tributyrin und artifiziellen *p*NP-Estern lediglich Nebenaktivitäten sein. Zur vollständigen Charakterisierung wären weitere Forschungsarbeiten notwendig, die im Zuge dieser Arbeit nicht beendet werden konnten. Tabelle 37 fasst die grundlegenden enzymatischen Eigenschaften von EstA2 und EstB1 im Vergleich zu einer Auswahl an thermostabilen und lipolytischen Proteinen aus archaeellen, bakteriellen und metagenomischen Quellen zusammen.

Die mögliche Anwendbarkeit der aufgefundenen Esterasen für biotechnologische Zwecke (Jaeger und Eggert, 2002) sollten in Zukunft evaluiert werden.

Tabelle 37: Auswahl (meta)genomischer, thermostabiler Esterasen und Lipasen mit biochemischen Daten im Vergleich zu EstA2 und EstB1.

Ursprung	Enzym	Molekulargewicht (kDa)	pH-Optimum	Temperatur-optimum (°C)	Thermostabilität (Halbwertszeit)	Bevorzugtes Substrat	Spezifische Aktivität (U/mg)	Referenz
<i>Sulfolobus solfataricus</i> P2	Esterase	100	8,0	75	15 min, 100 °C	-	29,5	Chung <i>et al.</i> (2000)
<i>Archaeoglobus fulgidus</i>	Esterase	27,5	7,0-8,0	70	10 % Aktivität nach 3 h, 90 °C	pNP-C <sub>4</sub>	-	Kim <i>et al.</i> (2008)
Metagenombank aus thermaler Umweltprobe	Esterase	34	6,0	95	20 min, 90 °C	pNP-C <sub>6</sub>	-	Rhee <i>et al.</i> (2005)
Anreicherungen aus Heizwasser, Hamburg (Deutschland)	Lipase	36 (dimerisiert)	8,0	75	3 h, 70 °C	pNP-C <sub>10</sub>	0,6	Chow <i>et al.</i> (2012)
Anreicherungen aus Boden- und Wasserproben (Deutschland)	Lipase	30,2	8,0	70	48 h, 70 °C	pNP-C <sub>8</sub>	12	Chow <i>et al.</i> (2012)
Hydrothermales Gebiet der Tiefsee	Esterase	51,7	8,0	60	80 % Aktivität nach 6 h, 90 °C	pNP-C <sub>4</sub>	-	Zhu <i>et al.</i> (2013)
<i>Pyrobaculum caldifontis</i>	Esterase	34,4 (trimerisiert)	7,0	90	56 min, 110 °C	pNP-C <sub>6</sub>	6.410	Hotta <i>et al.</i> (2002)
Metagenom aus heißen Quellen (Azoren)	Esterase (EstA2)	18	8,0-8,8	80	-	pNP-C <sub>3</sub>	44,1	diese Arbeit
Metagenom Kompost aus Holzabfällen	Esterase (EstB1)	70	8,0	75	-	pNP-C <sub>3</sub>	4,1	diese Arbeit

#### 4. *T. thermophilus* als neuartiger Expressionswirt für Metagenomanalysen

Die Idee zur Verwendung anderer Expressionswirte als *E. coli* für metagenomische Studien ist nicht neu (Tabelle 38). Bereits im Jahr 2003 suchten Courtois und Kollegen biosynthetische Gene für die Produktion von Metaboliten des Sekundärstoffwechsels in *E. coli* und dem Aktinomyceten *Streptomyces lividans* als Wirte für die Durchmusterung und Expression von Umweltgenombanken (Courtois *et al.*, 2003). Die Autoren durchmusterten dabei eine metagenomische *shotgun*-Bank bezüglich der Präsenz von Polyketid-Synthase-Genen mittels PCR, außerdem führten sie funktionelle Antibiose-Tests durch. Als Ergebnis konnte mindestens ein Klon identifiziert werden, welcher Kanamycin-Resistenz in *E. coli* vermittelte, außerdem konnten mittels HPLC (*high-performance liquid chromatography*) und LC-NMR (*liquid chromatography and nuclear magnetic resonance*) zwei bisher unbekannte Metaboliten identifiziert werden, welche lediglich in *S. lividans* gebildet werden konnten. Dieses Gram-positive Bakterium verfügt über einen hohen GC-Gehalt und ist aufgrund seiner Fähigkeit zur Bildung sekundärer Stoffwechselmetaboliten interessant. Weitere Studien mit diesem Aktinomyceten als Expressionswirt wurden für BACs berichtet (Martinez *et al.*, 2004), für die Durchmusterung lipolytischer und proteolytischer Aktivitäten (Lussier *et al.*, 2011) sowie anderer phänotypischer Auffälligkeiten (Morphologie oder Pigmentierung) (McMahon *et al.*, 2012). Neben *Pseudomonas putida* (Martinez *et al.*, 2004) ist eine Vielzahl an weiteren Proteobakterien mittels *broad host range*-Vektoren für metagenomische Studien verfügbar (Craig *et al.*, 2010). Weitere Systeme wurden beispielsweise als konjugative Vektoren für Fosmide (Aakvik *et al.*, 2009), BACs (Kakirde *et al.*, 2011) und Cosmide (Craig *et al.*, 2010) beschrieben. In extremophilen Organismen wie thermophilen Bakterien (z.B. *T. thermophilus*) und Archaeen (wie *Sulfolobus*) ist die Zahl der etablierten *Shuttle*-Vektor Systeme begrenzt (Berkner *et al.*, 2007; de Grado *et al.*, 1999), dennoch konnte im Vorfeld der vorliegenden Arbeit bereits das Fosmid pCT3FK als *E. coli-T. thermophilus Shuttle*-Vektor für große *Inserts* erfolgreich angewendet werden (Angelov *et al.*, 2009). In dieser Studie konnte genomische DNA des freilebenden und saccharolytischen Spirochaeten *Spirochaeta thermophila* bereits vergleichend in *E. coli* und *T. thermophilus* HB27 auf hemicellulolytische Aktivitäten durchmustert werden. Dabei wurden in 192 Fosmiden mehr Xylanaseaktivitäten in dem thermophilen Bakterium identifiziert, als im mesophilen Standardwirt. Ähnliche Ergebnisse konnten aus dem funktionellen Vergleich beider Expressionswirte bei der Durchmusterung nach diversen Hydrolasen aus einer *T. Brockianus*-Genbank erreicht werden (Leis *et al.*, 2013).

In dieser Arbeit konnte erneut gezeigt werden, dass die Anwendung von *T. thermophilus* als heterologer Expressionswirt für funktionelle Metagenomanalysen gegenüber dem Standardwirt *E. coli* erfolgreich war: (i) Es konnten in *T. thermophilus* zahlenmäßig mehr lipolytisch aktive Klone aufgefunden werden als in *E. coli*. (ii) In *T. thermophilus* HB27, einem Vertreter der „phylogenetisch

Tabelle 38: Zusammenfassung vergleichender funktioneller Metagenomanalysen in *E. coli* und anderen Bakterien.

Quelle der Nukleinsäuren	Aktivität, Phänotyp	Klone	Expressionswirte	Positive	Weitere Informationen	Referenz
Metagenombank aus Boden, Frankreich	Antimikrobielle Wirkstoffe, Polyketid-Synthasen (PKS I)	5.000	<i>E. coli</i> <i>Streptomyces lividans</i>	2 2	In <i>E. coli</i> konnten zwei antimikrobielle Klone identifiziert werden, in <i>S. lividans</i> waren neuartige Metaboliten aufgefunden worden	Courtois <i>et al.</i> (2003)
BAC, Gencluster, Bakterienisolate	Antimikrobielle Wirkstoffe	3	<i>E. coli</i> <i>S. lividans</i> <i>Pseudomonas putida</i>	1 1 1	Bekannte Gene/Gencluster/BAC-Sequenz wurden auf antimikrobielle Wirkstoffe in drei Wirten getestet	Martinez <i>et al.</i> (2004)
Genbank von Isolaten aus Boden (Alaska)/ Metagenombank aus Boden (Alaska)	Hämolytische Wirkstoffe, Pigmentierung	7.200/ 16.400	<i>E. coli</i> <i>S. lividans</i>	0 14	Durchmusterung nach phänotypischen Auffälligkeiten, Hamölyse, braune Pigmentierung, Sporenpigmentierung	McMahon <i>et al.</i> (2012)
Genomische DNA aus thermophiler, alkaliner Anreicherungskultur	Lipolytische und proteolytische Enzyme	2.000	<i>E. coli</i> <i>S. lividans</i>	17	Gleiche lipolytische Klone waren in beiden Wirten identifiziert worden, keine Proteasen identifiziert (vergleichend)	Lussier <i>et al.</i> (2011)
Metagenombanken von Waldboden, Flussbett Sediment und Wüstenboden	Pigmentierung, Antibiose, Morphologie	170.000 450.000 130.000	<i>E. coli</i> , <i>P. putida</i> , <i>R. metallidurans</i> , <i>B. graminis</i> , <i>A. tumefaciens</i> , <i>C. vibrioides</i>	2 >10 >22 8 3 0	Mittels <i>broad-host</i> Vektorsystem für Proteobakterien konnten verschiedenste Morphologien und Pigmentierungen identifiziert werden (vergleichend)	Craig <i>et al.</i> (2010)
Genbank aus <i>Spirochaeta thermophila</i>	Xylanasen	192	<i>E. coli</i> , <i>T. thermophilus</i>	12 20	Klassische Detektion auf Xylan-abbauende Fosmidklone in beiden Wirten (vergleichend)	Angelov <i>et al.</i> (2009)
Genbank aus Isolat <i>Thermus Brockianus</i>	Esterasen, Proteasen, Hemicellulasen	384	<i>E. coli</i> , <i>T. thermophilus</i>	2 7	Klassische Detektionsmethoden für verschiedene Hydrolasen in beiden Wirten (vergleichend)	diese Arbeit, Leis <i>et al.</i> (2013)
Metagenombank aus heißen Quellen / Kompost	Esterasen	8.000	<i>E. coli</i> , <i>T. thermophilus</i>	2 6	Hoch-Durchsatz Komplementationsstrategie, Klassische Detektion auf Tributyrin-Agarplatten in beiden Wirten (vergleichend)	diese Arbeit, Leis <i>et al.</i> (2015a)

alten“ und tief-verzweigten Entwicklungslinie *Deinococcus-Thermus* konnten überwiegend archaeelle Sequenzen aus Metagenombanken thermophiler Habitats aufgefunden werden. (iii) Das thermophile Bakterium exprimiert mehr thermostabile Enzyme von unterschiedlichen extremophilen Donor-Organismen mit speziellen physiologischen Eigenschaften (beispielsweise thermophile und acidophile Mikroorganismen mit Wachstumstemperaturen über 60 °C). (iv) Im Falle von EstB1 konnte das Protein einer selbst unterrepräsentierten Spezies innerhalb einer komplexen Lebensgemeinschaft nur in *T. thermophilus* aufgefunden werden (das Phylum *Chloroflexi* wies in der Kompostprobe weniger als 1 % aller 16S rDNA-Sequenzen auf; das Fosmid-insert von M12-4-G10 hatte dabei höchste Sequenzähnlichkeit mit *Sphaerobacter thermophilus* DSM 20745 als Vertreter dieser Abteilung). (v) Die identifizierten und charakterisierten Esterasen sind neuartig, da sie einerseits keinen klassischen Carboxylesterasen/Lipase-Familien zuordenbar sind und andererseits mittels gängigen bioinformatischen Methoden (mangels eindeutiger Funktionszuordnung oder Annotationen) nicht zwangsläufig identifizierbar gewesen wären.

Im Vergleich aller funktionellen Metagenomanalysen diverser Fosmidbanken in *E. coli* und in *T. thermophilus* zeigte sich ein weiterer Trend zwischen beiden Wirten: Von einem Dutzend verfügbaren Fosmidsequenzen wurden alle vorhersagbaren ORFs automatisiert in Proteinsequenzen translatiert und die ähnlichsten Proteinsequenzen aus der NCBI nr-Datenbank gesucht. Die erhaltenen Treffer (*hits*) wurden bezüglich ihrer Sequenzübereinstimmung (% Aminosäureidentität, *identity*) sortiert und als Histogramm gemäß ihrer Verteilungshäufigkeit in Abbildung 55 zusammengefasst. Auffällig ist dabei, dass in *T. thermophilus* vermehrt Fosmide mit Proteinsequenzen aufgefunden wurden, die generell eine geringere Übereinstimmung zu Datenbankeinträgen aufweisen als die von in *E. coli* identifizierten Fosmidsequenzen.

Da *E. coli* als standardisierter Expressionwirt in Metagenomstudien eingesetzt wird, mag es nicht verwundern, dass es in Zukunft immer schwieriger werden dürfte, mit diesem Expressionswirt alleine wahrhaft neuartige Gene und Genprodukte zu entdecken. Dies verdeutlicht das zentrale Problem von *E. coli* in der aktuellen Metagenomanalyse. Bedenke man die unglaubliche Vielfalt an mikrobiellem Leben in den verschiedensten natürlichen Habitats, kann es nur ein kleiner Anteil der umfangreichen Information sein, die bis heute aus Metagenomen entschlüsselbar ist. Neuartige Expressionswirte haben bereits in der Vergangenheit bewiesen, dass sie in Einzelfällen überlegene Expressionsfähigkeiten haben können. Die daran beteiligten und im Einleitungsteil bereits erwähnten Faktoren sind allerdings weitestgehend unverstanden. Die geringe Anzahl der in dieser Arbeit untersuchten Fosmidsequenzen limitiert eine systematische Analyse solcher intrinsischer, Sequenz-basierter Faktoren. Einige auffällige Beobachtungen sollen jedoch in den nächsten Abschnitten Erwähnung finden.

Sequenzidentität vorhergesagter Proteine zum nächsten Homolog (lt. blastP-Suche), Häufigkeitsverteilung in %:

Fosmid	Sequenzidentität vorhergesagter Proteine zum nächsten Homolog (lt. blastP-Suche), Häufigkeitsverteilung in %:																ORFs	Ursprung	Identifiziert in Wirt					
	0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70	71-75	76-80			81-85	86-90	91-95	96-100	<i>E. coli</i>	<i>T. thermophilus</i>
HA-cmc-1*	0	0	0	0	0	0	0	0	1	1	2	2	1	2	0	3	4	14	13	1	44	archaeell	+	-
TB-est-7*	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	12	11	1	27	bakteriell	+	+
ES-6-B5	0	0	0	0	0	0	1	1	3	0	2	2	2	0	2	2	7	0	1	0	21	bakteriell	+	n.a.
TB-xyI-3	0	0	0	0	1	0	0	0	0	0	0	0	0	1	4	3	4	9	6	3	31	bakteriell	+	n.a.
TB-xyI-4	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	8	3	8	5	28	bakteriell	+	n.a.
TB01-A10	0	0	0	0	0	0	1	0	0	0	1	0	0	2	1	1	5	10	10	3	34	bakteriell	+	n.a.
AZ2-4-B6	0	0	0	0	0	0	0	0	0	0	1	0	0	2	2	4	0	1	0	1	11	bakteriell	+	+
AZ3-33-C12	0	0	0	0	0	0	0	0	0	4	1	4	4	4	4	2	3	0	0	0	26	archaeell	-	+
AZ3-32-E5	0	0	0	0	0	0	0	0	0	0	0	1	2	3	2	2	1	0	0	0	11	archaeell	-	+
AZ3-14-D11	0	0	0	0	0	0	1	0	1	1	0	3	3	2	5	4	2	0	0	1	23	archaeell	-	+
AZ3-14-D2	0	0	0	0	0	0	0	0	1	0	2	2	4	4	4	0	1	0	2	1	17	archaeell	-	+
M12-4-G10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	10	18	32	bakteriell	-	+

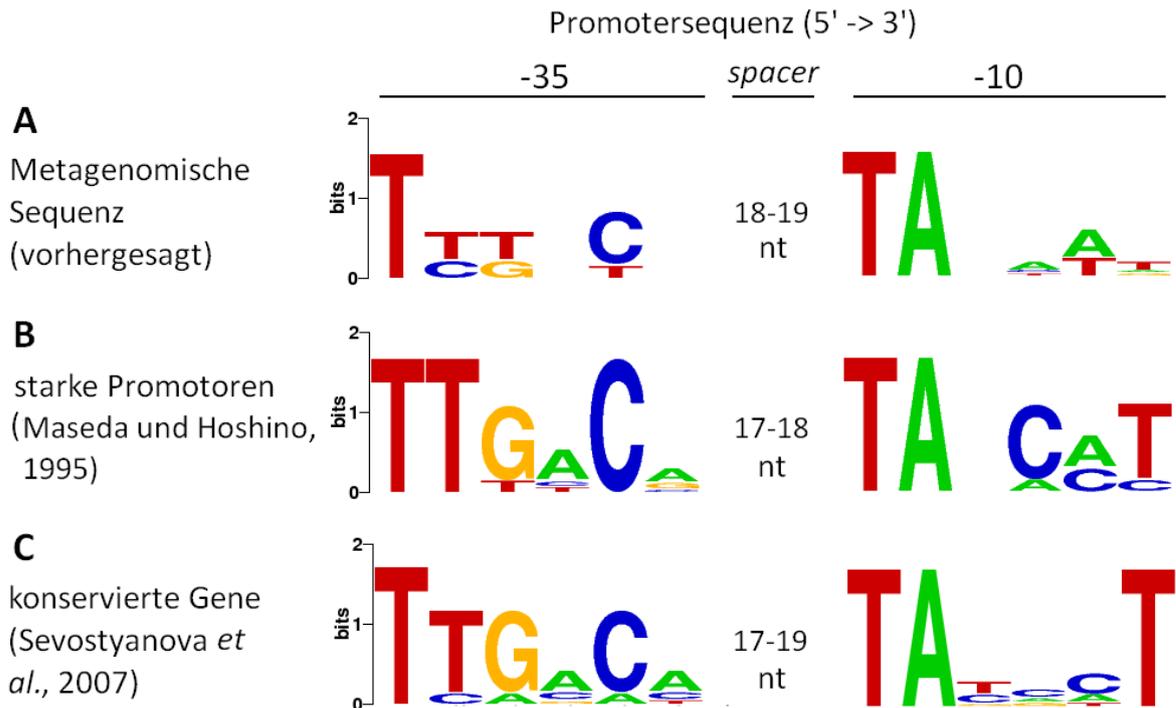
**Abbildung 55: Homologieverteilung von Proteinsequenzen metagenomischer Fosmide, die in *E. coli* und *T. thermophilus* identifiziert werden konnten. Die Übersicht zeigt alle vollständigen oder unvollständigen Fosmidsequenzen aus Metagenomen, welche eindeutig zuzuordnen waren und deren Aktivität mittels Substratplatten oder geeigneter enzymatischer Tests (z.B. pNP oder DNSA) bestätigt werden konnten. Die aus den Sequenzdaten vorhergesagten ORFs wurden in 5 %-Intervalle nach deren Häufigkeit eingeordnet. Die relativen Häufigkeit jedes Intervalls wurde graphisch als *Heatmap* visualisiert (weiß: 0 % Häufigkeit, rot: häufigstes Intervall). Falls keine vorhergesagte Proteinsequenz aus dem Fosmid in einem speziellen Intervall (z.B. 5-10 % Sequenzhomologie) auftritt, so ist ihre relative Häufigkeit 0 %, die Farbkodierung weiß. Je häufiger Proteinsequenzen einem 5 %-Intervall zugeordnet werden konnten, desto intensiver ist die rote Farbkodierung.**

\*: Die beiden Fosmide HA-cmc-1 und TB-est-7 wurden von pCC1FOS in pCT3FK-Fosmide umklontiert, sodass sie für vergleichende Analysen in beiden Wirten zur Verfügung standen (Daten nicht gezeigt, V. T. T. Pham, unveröffentlichte Ergebnisse).

Die in *E. coli* identifizierte metagenomische DNA, welche zu aktiven Fosmidklonen führte, hatte einen GC-Gehalt zwischen etwa 55 und 65 % (bei einem eigenen chromosomalen GC-Gehalt von 50,8 % in *E. coli* Stamm K-12). In *T. thermophilus* aufgefundene Metagenomsequenzen, welche in dem thermophilen Wirt Aktivität vermittelten, wiesen hingegen wesentlich breitere GC-Gehalte im Bereich von nur 32,1 bis maximal 67,6 % auf (das Genom des Stammes HB27 selbst enthält 69 % GC). Die vorliegenden Daten aus Sequenzierung und Aktivitäts-basierter *Screening* beider Wirte lassen daher vermuten, dass der GC-Gehalt einen Einfluss auf die Expressionsfähigkeit der metagenomischen DNA im Wirt hat. Dabei scheint *T. thermophilus* vor allem bei Nukleinsäurefragmenten mit niedrigerem GC-Gehalt Vorteile gegenüber *E. coli* zu besitzen, da mehr aktive Klone in dem thermophilen Bakterium aufgefunden wurden. Die Kenntnis der verschiedenen GC-Gehalte von Umweltorganismen führte zur Einführung sogenannter GC-Profile metagenomischer Nukleinsäuren (Holben und Harris, 1995), welche auch für funktionelle *Screenings* verwendet werden können. Als mögliche Strategie könnte metagenomische DNA mittels Dichtegradientenzentrifugation aufgetrennt (Holben *et al.*, 2004) und je nach GC-Gehalt in den entsprechenden Wirt kloniert werden, um die Ausbeuten aus der funktionellen Metagenomanalyse zu steigern (Ekkers *et al.*, 2012).

Bis heute ist es ebenfalls unbekannt, wie bakterielle Wirte Promotoren von metagenomischer Fremd-DNA erkennen. Daher wurden mögliche Promotorsequenzen zwischen 40 bis 200 Basen vor den Startcodon gelegenen Esterase-Genen aus den Durchmusterungen in *T. thermophilus* manuell analysiert und vorhergesagt. Interessanterweise konnten vergleichbare Sequenzmuster der -35 und -10 Region der vorhergesagten Promotoren aufgefunden werden (Abbildung 56A). Das Motiv Tttncn-(18-19nt)-TAnaAT hatte dabei beachtliche Ähnlichkeit zu experimentell bestätigten, starken Promotorsequenzen in *T. thermophilus* (Maseda und Hoshino, 1995, Abbildung 56B) sowie zu relevanten Promotoren wichtiger *housekeeping*-Gene (Sevostyanova *et al.*, 2007, Abbildung 56C). Die vorhergesagten Promotorsequenzen haben ebenfalls Ähnlichkeit zu dem bekannten *E. coli* RNA-Polymerase *Consensus*-Motiv TTGACA-(18nt)-TATAAT (Harley und Reynolds, 1987).

Analysiert man nun die in *E. coli* aktiven Klone AZ2-4-B6 und M12-4-D9, so zeigt sich für die vorhergesagten Promotorsequenzen an den -10 und -35 Regionen vor den Esterase-kodierenden ORFs der beiden Fosmide ein abweichendes Muster TCnGCn-(18-19nt)-TAnnn (Abbildung 57). Umgekehrt betrachtet waren keine Esterase-aktiven Klone mit Fosmiden archaeller Herkunft nachweisbar, obwohl deren Promotorsequenzen in *E. coli* eigentlich ideal sein müssten. Es wird diskutiert, dass suboptimale Promotorsequenzen für die Expression von Genen der Fremd-DNA geeigneter sind, da sie weniger das Transkriptionsprofil des Wirts beeinflussen können (Warren *et al.*, 2008). Wirts-eigene Sigma-Faktoren sollten durch die eingebrachten Nukleinsäuren des Metagenoms weniger stark rekrutiert werden, damit sie nicht in direkter Konkurrenz zu eigenen Promotoren überlebensnotwendiger (*housekeeping*) Gene stehen. Unter anderem können solche Sequenzen toxisch und damit für den Wirt unklonierbar sein (Sorek *et al.*, 2007).



**Abbildung 56: Vergleich vorhergesagter und konservierter Promotorsequenzen in *T. thermophilus*.** Die Abbildung zeigt die -35 und -10 Regionen vorhergesagter metagenomischer Promotorsequenzen im Vergleich zu experimentell ermittelten Promotorsequenzen in *T. thermophilus*. Ein Sequenzalignment der jeweiligen *Consensus*-Sequenzen wurde als WebLogo (<http://weblogo.berkeley.edu/>) visualisiert.

Zur Klärung dieser grundlegenden Frage, warum die o.g. Fosmidgene nicht funktionell exprimierbar sind, bedarf es systematischer Transkriptions- und Expressionsstudien. Dazu werden momentan zwei Modellfosmide TB-est-7 und HA-cmc-1 vergleichend in *E. coli* und *T. thermophilus* untersucht (V. T. T. Pham, unveröffentlichte Ergebnisse).

Ein weiterer Hinweis ergibt sich aus der vergleichenden Analyse der ribosomalen Bindestellen vor putativen Kandidaten-ORFs. Obwohl Archaeen sich gegenüber Bakterien und Eukaryoten phylogenetisch abgrenzen lassen, teilen sie einige Mechanismen der Translationsinitiation mit denen der Bakterien, wie polycistronische mRNA, Shine-Dalgarno (S-D)-Sequenzen und das Fehlen eines Poly(A)-Schwanzes (Keeling und Doolittle, 1995). Studien an *S. solfataricus* haben jedoch gezeigt, dass es verschiedene Transkriptionssignale in Archaeen geben kann (Tolstrup *et al.*, 2000). Dabei fiel den Autoren auf, dass die klassische Shine-Dalgarno-Sequenz nicht für das erste Gen in kotranskribierten Operons sowie in einzelnen, isolierten Genen aufgefunden werden konnte. Stattdessen identifizierte man eine sogenannte A-Box als Transkriptionssignal, welche 27 Basen vor dem Startcodon liegt. Diese Sequenzen verfügen über keine konservierten Sequenzmotive und existieren auch nicht in Bakterien. In Übereinstimmung zu den Autoren waren für die drei durch *T. thermophilus* entdeckten Sequenzen (AZ3-32-E5, AZ3-33-C12 und M12-4-G10) keine ribosomalen Bindestellen (RBS) in den Transkriptionsinitiationsregionen (TIRs) der Kandidatengene vorhersagbar.

Fosmid	Identifiziert/ aktiv in	Promotorsequenz (5' → 3')			27 nt Translationsinitiations- region (TIR) vor Startcodon der Kandidaten-ORFs
		-35	spacer	-10	
AZ2-4-B6	<i>E. coli</i> / <i>T. thermophilus</i>	CTCAAGAATG	TCTGCA TCAAGCTCTGACTGCGTTC	TACTAT	GCGATGCGTG 98 nt TTCCCTGAAGAAA <u>AGGAGC</u> AAAAAGACCCGAA
AZ3-32-E5	<i>T. thermophilus</i>	AATGAAAGAT	TTTACG AATCTTAATTTGATTTCTT	TAACAT	TCATAGGTTT 161 nt TACGATTTTTACATAACTCACAAAACCC
AZ3-33-C12 <sup>1</sup>	<i>T. thermophilus</i>	ACAATATGTT	TTTACA GGGTACTTTAATGGGATA	TAAATA	GCCTGGCTAG 79 nt ATAAAAATTTCTACAAAATACTACAATTA
M12-4-D9 <sup>2</sup>	<i>E. coli</i> / <i>T. thermophilus</i>	CCGGGATAAA	TCGGCC CGGACGGGCAAAACCTTA	TAGATG	AATGAAAAGGA 8 nt TATAGATGAATGAAA <u>AGGATG</u> CCATGCGGC
M12-4-G10	<i>T. thermophilus</i>	CGGAGATCCA	TTGCTG CGTCGTCGGACTTCCAAT	TAGAAAT	GCATGTTCTG 35 nt GCACGTGTTTGTGGATGGGAGAACAGG
<i>consensus</i>	<i>T. thermophilus</i>	Ttttncn	18-19 nt	TAnaat	
<i>consensus</i>	<i>E. coli</i>	TcncGcn	18-19 nt	TAnnn	

**Abbildung 57: Promotorsequenzen, welche vor den für Esterasen kodierenden Kandidatengen vorhergesagt werden konnten.**

<sup>1</sup>: Die Promotorsequenz von Fosmid AZ3-33-C12 stammt aus den ersten ORF des Operons, die *upstream*-Sequenz des Startcodons stammt vom Zielgen selbst, welcher der zweite ORF im Operon ist.

<sup>2</sup>: Obwohl das Fosmid ursprünglich in *E. coli* aufgefunden worden ist, zeigte sich in *T. thermophilus* ebenfalls signifikante lipolytische Aktivität in *pNP*-Assays. Die vorhergesagte Shine-Dalgarno-Sequenz (unterstrichene und dick markierte Nukleotidsequenz) wurde nur angezeigt, wenn mindestens 5 von 6 Basen zur idealen S-D-Sequenz (5'-AGGAGG-3') identisch sind.

Für die in *E. coli* identifizierten Fosmide konnten vor den beiden Genen jeweils zur typischen bakteriellen Shine-Dalgarno-Sequenz (5'-AGGAGG-3') zumindest ähnliche Sequenzen aufgefunden werden (Abbildung 57, markierte und unterstrichene Sequenzen) (Snyder und Champness, 2007). Die Beobachtungen könnten einen Hinweis darauf geben, ob die Präsenz der S-D Sequenz für die korrekte Lokalisierung der Ribosomen zur Translationsinitiation in *T. thermophilus* weniger notwendig ist als für *E. coli*, bzw. alternative Mechanismen der Translationsinitiation existieren. Weitere Experimente mittels gezielter Mutationen wären nötig, um die Rolle der Shine-Dalgarno Sequenz für die aktive Expression in diesen und anderen Wirten ebenfalls systematisch zu untersuchen.

## VI. Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Verbesserung des *Thermus thermophilus* Wirt/Vektor-Systems und der Nutzung von *T. thermophilus* als alternatives Wirtsbakterium für das *Screening* und die Expression von metagenomischen Genen.

In dieser Arbeit konnte durch *in vitro* Transposon-Mutagenese eine Transposonmutagenesebank in *T. thermophilus* HB27 mit knapp 8.000 Einzelmutanten erstellt werden. Ihre Durchmusterung lieferte verschiedene Mutanten, welche gegenüber toxischen BCI-Substraten resistent waren. Die Transposonmutanten-Sammlung stellt ein wertvolles Werkzeug dar, um weitere Eigenschaften dieses extrem thermophilen Modellorganismus studieren zu können.

Auf Basis des licht-induzierbaren *crtB*-Promotors des Carotinoid-Genclusters wurden replikative und insertions-befähigte Expressionsvektoren für *T. thermophilus* HB27 hergestellt und damit ein homologes  $\beta$ -Glucosidasegen exprimiert. Mit Hilfe eines weiterentwickelten, integrativen Vektors mit einem synthetischen *crtB*-Promotor und einer zusätzlichen ribosomalen Bindestelle (pJ-car05-RBS) konnten zwei metagenomische Esterasen EstA2 und EstB1 überexprimiert werden.

Mittels einer neuen, markerlosen Gendeletionsmethode konnten in *T. thermophilus* HB27 ein Stamm (BL03) erzeugt werden, in dem die vier ORFs TT\_C0340, TT\_C0341, TT\_C0904 und TT\_C1787 deletiert sind, die für die Hydrolyse von Tributyrin in dem thermophilen Wirt verantwortlich sind. Die sequentielle Deletion dieser Gene erlaubte eine Mutationsanalyse der lipolytischen Eigenschaften des Bakteriums. Die erhobenen Daten lassen darauf schließen, dass die Tributyrin-Esteraseaktivität von *T. thermophilus* auf mehrere Genprodukte zurückzuführen ist, die durch unterschiedlich hohe Aktivitäten, Lokalisierung und Substratpräferenzen charakterisiert sind. Es konnte außerdem gezeigt werden, dass die Vierfachmutante BL03 nicht mehr in der Lage war, auf Minimalmedien mit Tributyrin als definierte Kohlenstoffquelle zu wachsen.

Verschiedene pCC1FOS-Fosmidbanken aus metagenomischer Umwelt-DNA und aus Anreicherungskulturen (aus Elefantendung, Elbe Fluss-Sediment, Schiffsbohrwurm, Tiefseeproben) wurden mittels funktioneller Metagenomanalyse in *E. coli* auf neuartige, biotechnologisch relevante Enzyme hin untersucht (Esterasen, Lipasen, Cellulasen, Hemicellulasen, Amylasen, Proteasen). Insgesamt wurden 37 aktive Fosmidklone in *E. coli* aufgefunden. Die Sequenzanalyse 454-Pyrosequenzierter Fosmide erlaubte die Identifizierung mehrerer Kandidaten-ORFs. Eine archaeelle und thermostabile Glycosidhydrolase der Familie 12 wurde biochemisch charakterisiert. Das Enzym Cel12E zeichnete sich durch hohe Aktivitäten auf löslichen Cellulosesubstraten wie Carboxymethylcellulose (CMC) aus und zeigte im verwendeten Assaysystem bei 92 °C und pH 5,5 seine höchste Aktivität. Das Temperaturmaximum von Cel12E lag im Bereich des Siedepunktes von Wasser. Die Analyse der Hydrolyseprodukte (Cellotriase und Cellotetraose) auf verschiedenen

Substraten lässt auf eine Endo- $\beta$ -1,4-Glucanase-Aktivität des Enzyms schließen. Hohe Chloridkonzentrationen und zweiwertige Mangan- und Cobaltionen steigerten die Aktivität signifikant.

Aus dem Probenmaterial heißer und teils stark azidischer Quellen von den Azoren konnte mit Hilfe verschiedener Extraktionsmethoden hoch-molekulare metagenomische DNA isoliert werden. Die Konstruktion einer pCT3FK *Shuttle*-Fosmidbank (AZ) resultierte in etwa 14.500 Klonen in *E. coli*. Aus diesen und aus Klonen einer weiteren Fosmidbank aus einer natürlich erhitzten Kompostprobe (M12) mit 2.000 Klonen konnten 8.000 Einzelfosmide in *T. thermophilus* HB27 übertragen werden, was die funktionelle und parallele Durchmusterung der Klone auf lipolytische Aktivitäten in beiden Wirten bei 60 °C ermöglichte. In *E. coli* konnten zwei aktive Klone identifiziert werden. Im thermophilen Bakterium konnten mittels einer Komplementationsstrategie der BL03-Mutante sechs Fosmide identifiziert werden, die dem Stamm durch lipolytische Aktivitäten das Wachstum auf Minimalmedium mit Tributyrin ermöglichten.

Die Sequenzierung und *in silico* Analyse der in *E. coli* identifizierten Fosmide lieferte zwei klassische  $\alpha/\beta$ -Hydrolasen der Familie 6 (AZ2-4-B6, M12-4-D9) welche deutliche Sequenzähnlichkeit mit bereits bekannten lipolytischen Enzymen aufwiesen. In *T. thermophilus* BL03 wurden ein bereits in *E. coli* aufgefundenes Fosmid (AZ2-4-B6) sowie fünf weitere Fosmide (AZ3-14-D2, AZ3-14-D11, AZ3-32-E5, AZ3-33-C12, M12-4-G10) identifiziert und auf Kandidatengene hin analysiert. Es konnten drei potentielle Esterasen (EstA2, EstB1 und EstS9) festgestellt werden, von welchen EstA2 und EstB1 heterolog in *E. coli* BL21 exprimiert und vorläufig charakterisiert wurden. Die beiden Enzyme zeigten Substratpräferenzen für kurzkettige *pNP*-Acylester ( $C_3$ - $C_4$ ) und wiesen mit den verwendeten Assayverfahren höchste Aktivitäten bei Temperaturen von 80 °C bzw. 75 °C und pH-Werten von 8,0 bis 8,8 auf. Bei den beiden Proteinen, die offenbar archaeellen bzw. bakteriellen Ursprungs sind, handelt es sich um neuartige Esterasen, da sie weder zu bekannten lipolytischen Enzymfamilien zugeordnet werden können, noch zeigen sie deutliche Sequenzähnlichkeiten zu annotierten oder charakterisierten Esterasen und Lipasen.

Da *E. coli* bis heute der am meisten eingesetzte Expressionswirt für funktionelle Metagenomanalysen ist, unterstreicht die Auffindung der beiden neuen Esterasegene in *T. thermophilus* den Nutzen der Entwicklung und Anwendung alternativer Expressionswirte. Diese Arbeit zeigt, dass *T. thermophilus* ein geeigneter Wirt für die funktionelle Durchmusterung insbesondere von Metagenombanken thermophiler Habitats ist und zur Erhöhung der Ausbeuten aus funktionellen Durchmusterungen, zum Auffinden archaeeller und bakterieller Sequenzen von selbst unterrepräsentierten Spezies, und zum Auffinden neuartiger Proteine ohne Sequenzähnlichkeit zu bekannten Enzymen dienen kann.

## VII. Im Rahmen der Dissertation entstandene Veröffentlichungen

1. **Leis, Benedikt**; Angelov, Angel; Liebl, Wolfgang (2013): Screening and expression of genes from metagenomes. *Advances in Applied Microbiology* 83, S. 1-68.
2. Angelov, Angel; Li, Haijuan; Geissler, Andreas; **Leis, Benedikt**; Liebl, Wolfgang (2013): Toxicity of indoxyl derivative accumulation in bacteria and its use as a new counterselection principle. *Systematic and Applied Microbiology* 36 (8), S. 585-592.
3. **Leis, Benedikt**; Angelov, Angel; Li, Haijuan; Liebl, Wolfgang (2014): Genetic analysis of lipolytic activities in *T. thermophilus* HB27. *Journal of Biotechnology* 191, S. 150-157.
4. **Leis, Benedikt**; Angelov, Angel; Mientus, Markus; Li, Haijuan; Pham, V. T. Trang; Lauinger, Benjamin; Bongen, Patrick; Pietruszka, Jörg; Gonçalves, Luís G.; Santos, Helena; Liebl, Wolfgang (2015a): Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Frontiers in Microbiology* 6, 275.
5. **Leis, Benedikt**; Heinze, Simon; Angelov, Angel; Pham, V. T. Trang; Thürmer, Andrea; Jebbar, Mohamed; Golyshin, Peter N.; Streit, Wolfgang R.; Daniel, Rolf; Liebl, Wolfgang (2015b): Functional Screening of Hydrolytic Activities Reveals an Extremely Thermostable Cellulase from a Deep-Sea Archaeon. *Frontiers in Bioengineering and Biotechnology* 3, 95.

### Beiträge zu Fachtagungen:

1. **Leis, Benedikt**; Angelov, Angel; Liebl, Wolfgang (2011): *In vivo* mobilization of fosmid (meta)genomic libraries. Posterpräsentation, Jahrestagung der Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM), Karlsruhe, 2011.
2. **Leis, Benedikt**; Angelov, Angel; Golyshin, Peter; Liebl, Wolfgang (2011): *In vivo* mobilization of fosmid (meta)genomic libraries. Posterpräsentation, 5<sup>th</sup> European Conference on Prokaryotic and Fungal Genomics (ProkaGENOMICS), Göttingen, 2011.
3. **Leis, Benedikt**; Angelov, Angel; Golyshin, Peter; Liebl, Wolfgang (2014): Identification and characterization of a thermostable archaeal endoglucanase uncovered by functional metagenomics. Posterpräsentation, MIE BIOFORUM 2014 Lignocellulose Degradation and Biorefinery, Ise-Shima, Japan, 2014.

## VIII. Literaturangaben

- Aakvik, T.; Degnes, K.F.; Dahlsrud, R.; Schmidt, F.; Dam, R.; Yu, L.; Völker, U.; Ellingsen, T.E.; Valla, S. (2009): A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species, In: *FEMS microbiology letters* 296 (2), S. 149–158.
- Agari, Y.; Agari, K.; Sakamoto, K.; Kuramitsu, S.; Shinkai, A. (2011): TetR-family transcriptional repressor *Thermus thermophilus* FadR controls fatty acid degradation. In: *Microbiology (Reading, Engl.)* 157 (Pt 6), S. 1589–1601.
- Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. (1990): Basic local alignment search tool. In: *J. Mol. Biol.* 215 (3), S. 403–410.
- Amann, R. I.; Ludwig, W.; Schleifer, K. H. (1995): Phylogenetic identification and in situ detection of individual microbial cells without cultivation. In: *Microbiol. Rev.* 59 (1), S. 143–169.
- Amore, A.; Pepe, O.; Ventorino, V.; Birolo, L.; Giangrande, C.; Faraco, V. (2012): Cloning and recombinant expression of a cellulase from the cellulolytic strain *Streptomyces* sp. G12 isolated from compost. In: *Microb. Cell Fact.* 11, S. 164.
- Angelov, A.; Li, H.; Geissler, A.; Leis, B.; Liebl, W. (2013): Toxicity of indoxyl derivative accumulation in bacteria and its use as a new counterselection principle. In: *Syst. Appl. Microbiol.* 36 (8), S. 585–592.
- Angelov, A.; Mientus, M.; Liebl, S.; Liebl, W. (2009): A two-host fosmid system for functional screening of (meta)genomic libraries from extreme thermophiles. In: *Syst. Appl. Microbiol.* 32 (3), S. 177–185.
- Arnold, C.; Hodgson, I. J. (1991): Vectorette PCR: a novel approach to genomic walking. In: *PCR Methods Appl.* 1 (1), S. 39–42.
- Arpigny, J. L.; Jaeger, K. E. (1999): Bacterial lipolytic enzymes: classification and properties. In: *Biochem. J.* 343 Pt 1, S. 177–183.
- Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M. *et al.* (2000): Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. In: *Nat. Genet.* 25 (1), S. 25–29.
- Au, K.-S.; Chan, K.-Y. (1987): Purification and Properties of the Endo-1,4- $\beta$ -glucanase from *Bacillus subtilis*. In: *Microbiology* 133 (8), S. 2155–2162.
- Bauer, M. W.; Driskill, L. E.; Callen, W.; Snead, M. A.; Mathur, E. J.; Kelly, R. M. (1999): An endoglucanase, EglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1-3),(1-4)-beta-D-glucans and cellulose. In: *J. Bacteriol.* 181 (1), S. 284–290.
- Beaz H., R.; Cleenwerck, I.; Balboa, S.; Prado, S.; Vos, P. de; Romalde, J. L. (2009): *Vibrio breoganii* sp. nov., a non-motile, alginolytic, marine bacterium within the *Vibrio halioticoli* clade. In: *Int. J. Syst. Evol. Microbiol.* 59 (Pt 7), S. 1589–1594.
- Berger, J. L.; Lee, B. H.; Lacroix, C. (1995): Identification of new enzyme activities of several strains of *Thermus* species. In: *Appl. Microbiol. Biotechnol.* 44 (1-2), S. 81–87.
- Bergey, D. H.; Krieg, Noel R.; Holt, John G. (1984-c1989): *Bergey's manual of systematic bacteriology*. Baltimore, MD: Williams & Wilkins.
- Berkner, S.; Grogan, D.; Albers, S.-V.; Lipps, G. (2007): Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren)archaea. In: *Nucleic Acids Res.* 35 (12), S. e88.
- Berry, A. E.; Chiocchini, C.; Selby, T.; Sosio, M.; Wellington, E. M. H. (2003): Isolation of high molecular weight DNA from soil for cloning into BAC vectors. In: *FEMS Microbiol. Lett.* 223 (1), S. 15–20.
- Bijtenhoorn, P.; Mayerhofer, H.; Müller-Dieckmann, J.; Utpatel, C.; Schipper, C.; Hornung, C. *et al.* (2011): A novel metagenomic short-chain dehydrogenase/reductase attenuates *Pseudomonas aeruginosa* biofilm formation and virulence on *Caenorhabditis elegans*. In: *PLoS ONE* 6 (10), S. e26278.
- Binga, E. K.; Lasken, R. S.; Neufeld, J. D. (2008): Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. In: *ISME J* 2 (3), S. 233–241.
- Birnboim, H. C.; Doly, J. (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. In: *Nucleic Acids Res.* 7 (6), S. 1513–1523.

- Blas-Galindo, E.; Cava, F.; López-Viñas, E.; Mendieta, J.; Berenguer, J. (2007): Use of a dominant rpsL allele conferring streptomycin dependence for positive and negative selection in *Thermus thermophilus*. In: *Appl. Environ. Microbiol.* 73 (16), S. 5138–5145.
- Boc, A.; Diallo, A. B.; Makarenkov, V. (2012): T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. In: *Nucleic Acids Research* 40 (W1), S. W573.
- Bradford, M. M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. In: *Analytical Biochemistry* 72 (1-2), S. 248–254.
- Bronnenmeier, K.; Staudenbauer, W. L. (1988): Purification and properties of an extracellular  $\beta$ -glucosidase from the cellulolytic thermophile *Clostridium stercorarium*. In: *Appl Microbiol Biotechnol* 28 (4-5), S. 380–386.
- Carfi, A.; Pares, S.; Duée, E.; Galleni, M.; Duez, C.; Frère, J. M.; Dideberg, O. (1995): The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. In: *EMBO J.* 14 (20), S. 4914–4921.
- Carver, T.; Thomson, N.; Bleasby, A.; Berriman, M.; Parkhill, J. (2009): DNAPlotter: circular and linear interactive genome visualization. In: *Bioinformatics* 25 (1), S. 119–120.
- Caspi, R.; Altman, T.; Dreher, K.; Fulcher, C. A.; Subhraveti, P.; Keseler, I. M. *et al.* (2012): The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. In: *Nucleic Acids Res.* 40 (Database issue), S. D742-53.
- Cava, F.; Berenguer, J. (2006): Biochemical and regulatory properties of a respiratory island encoded by a conjugative plasmid in the extreme thermophile *Thermus thermophilus*. In: *Biochem. Soc. Trans.* 34 (Pt 1), S. 97–100.
- Cava, F.; Hidalgo, A.; Berenguer, J. (2009): *Thermus thermophilus* as biological model. In: *Extremophiles* 13 (2), S. 213–231.
- Cava, F.; Laptenko, O.; Borukhov, S.; Chahlafai, Z.; Blas-Galindo, E.; Gómez-Puertas, P.; Berenguer, J. (2007): Control of the respiratory metabolism of *Thermus thermophilus* by the nitrate respiration conjugative element NCE. In: *Mol. Microbiol.* 64 (3), S. 630–646.
- Chandrasekharaiah, M.; Thulasi, A.; Vijayarani, K.; Kumar, D. P.; Santosh, S. S.; Palanivel, C. *et al.* (2012): Expression and biochemical characterization of two novel feruloyl esterases derived from fecal samples of *Rusa unicolor* and *Equus burchelli*. In: *Gene* 500 (1), S. 134–139.
- Chiu, J.; March, P. E.; Lee, R.; Tillett, D. (2004): Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. In: *Nucleic Acids Res.* 32 (21), S. e174.
- Chow, J.; Kovacic, F.; Dall Antonia, Y.; Krauss, U.; Fersini, F.; Schmeisser, C. *et al.* (2012): The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. In: *PLoS ONE* 7 (10), S. e47665.
- Chung, Y. M.; Park, C. B.; Lee, S. B. (2000): Partial purification and characterization of thermostable esterase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. In: *Biotechnol. Bioprocess Eng.* 5 (1), S. 53-56. Online verfügbar unter <http://dx.doi.org/10.1007/BF02932354>.
- Cole, J. R.; Chai, B.; Marsh, T. L.; Farris, R. J.; Wang, Q.; Kulam, S. A. *et al.* (2003): The Ribosomal Database Project (RDP-II): previewing a new aligner that allows regular updates and the new prokaryotic taxonomy. In: *Nucleic Acids Res.* 31 (1), S. 442–443.
- Courtois, S.; Cappellano, C. M.; Ball, M.; Francou, F.-X.; Normand, P.; Helynck, G. *et al.* (2003): Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. In: *Appl. Environ. Microbiol.* 69 (1), S. 49–55.
- Craig, J. W.; Chang, F.-Y.; Kim, J. H.; Obiajulu, S. C.; Brady, S. F. (2010): Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. In: *Appl. Environ. Microbiol.* 76 (5), S. 1633–1641.
- Damon, C.; Vallon, L.; Zimmermann, S.; Haider, M. Z.; Galeote, V.; Dequin, S. *et al.* (2011): A novel fungal family of oligopeptide transporters identified by functional metatranscriptomics of soil eukaryotes. In: *ISME J* 5 (12), S. 1871–1880.
- Daniel, R. (2005): The metagenomics of soil. In: *Nat Rev Micro* 3 (6), S. 470–478.
- Darling, A. E.; Mau, B.; Perna, N. T.; Stajich, J. E. (2010): progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. In: *PLoS ONE* 5 (6), S. e11147.

- Deive, F. J.; Carvalho, E.; Pastrana, L.; Rúa, M. L.; Longo, M. A.; Sanroman, M. A. (2009): Assessment of Relevant Factors Influencing Lipolytic Enzyme Production by *Thermus thermophilus* HB27 in Laboratory-Scale Bioreactors. In: *Chem. Eng. Technol.* 32 (4), S. 606–612.
- Demirtas, M. U.; Kolhatkar, A.; Kilbane, J. J. (2003): Effect of aeration and agitation on growth rate of *Thermus thermophilus* in batch mode. In: *J. Biosci. Bioeng.* 95 (2), S. 113–117.
- Deppe, V.; Hoven, N.; Huchel, U.D; Maurer, K.H.D; O'connell, T.D; Prüser, I. (2008): Use of superoxide dismutases to cleave and/or remove Amadori and/or Maillard products, especially as components of detergent, cosmetic or pharmaceutical products: Google Patents. Online verfügbar unter <http://www.google.com/patents/DE102007010785A1?cl=en>.
- Domínguez, A.; Sanromán, A.; Fuciños, P.; Rúa, M. L.; Pastrana, L.; Longo, M. A. (2004): Quantification of intra- and extra-cellular thermophilic lipase/esterase production by *Thermus* sp. In: *Biotechnol. Lett.* 26 (9), S. 705–708.
- Domínguez, A.; Deive, F. J.; Pastrana, L.; Rúa, M. L.; Longo, M. A.; Sanroman, M. A. (2010): Thermostable lipolytic enzymes production in batch and continuous cultures of *Thermus thermophilus* HB27. In: *Bioprocess Biosyst Eng* 33 (3), S. 347–354.
- Domínguez, A.; Fuciños, P.; Rúa, M. L.; Pastrana, L.; Longo, M. A.; Sanromán, M. A. (2007): Stimulation of novel thermostable extracellular lipolytic enzyme in cultures of *Thermus* sp. In: *Enzyme and Microbial Technology* 40 (2), S. 187–194. Online verfügbar unter <http://www.sciencedirect.com/science/article/pii/S0141022906004674>.
- Du Plessis, E. M.; Berger, E.; Stark, T.; Louw, M. E.; Visser, D. (2010): Characterization of a novel thermostable esterase from *Thermus scotoductus* SA-01: evidence of a new family of lipolytic esterases. In: *Curr. Microbiol.* 60 (4), S. 248–253.
- Dumont, M. G.; Murrell, J. C. (2005): Stable isotope probing - linking microbial identity to function. In: *Nat. Rev. Microbiol.* 3 (6), S. 499–504.
- Durand, A.; Hughes, R.; Roussel, A.; Flatman, R.; Henrissat, B.; Juge, N. (2005): Emergence of a subfamily of xylanase inhibitors within glycoside hydrolase family 18. In: *FEBS J.* 272 (7), S. 1745–1755.
- Eddy, S. R. (2011): Accelerated Profile HMM Searches. In: *PLoS Comput. Biol.* 7 (10), S. e1002195.
- Ekkers, D. M.; Cretoiu, M. S.; Kielak, A. M.; van Elsas, J. D. (2012): The great screen anomaly--a new frontier in product discovery through functional metagenomics. In: *Appl. Microbiol. Biotechnol.* 93 (3), S. 1005–1020.
- Elend, C.; Schmeisser, C.; Hoebenreich, H.; Steele, H. L.; Streit, W. R. (2007): Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. In: *J. Biotechnol.* 130 (4), S. 370–377.
- Ferrer, M.; Ghazi, A.; Beloqui, A.; Vieites, J. M.; López-Cortés, N.; Marín-Navarro, J. *et al.* (2012): Functional Metagenomics Unveils a Multifunctional Glycosyl Hydrolase from the Family 43 Catalysing the Breakdown of Plant Polymers in the Calf Rumen. In: *PLoS ONE* 7 (6), S. e38134.
- Fu, J.; Leiros, H.-K. S.; Pascale, D. de; Johnson, K. A.; Blencke, H.-M.; Landfald, B. (2013): Functional and structural studies of a novel cold-adapted esterase from an Arctic intertidal metagenomic library. In: *Appl. Microbiol. Biotechnol.* 97 (9), S. 3965–3978.
- Fuciños, P.; Abadín, C. M.; Sanromán, A.; Longo, M. A.; Pastrana, L.; Rúa, M. L. (2005a): Identification of extracellular lipases/esterases produced by *Thermus thermophilus* HB27: partial purification and preliminary biochemical characterisation. In: *J. Biotechnol.* 117 (3), S. 233–241.
- Fuciños, P.; Atanes, E.; López-López, O.; Cerdán, E. M.; González-Siso, M. I.; Pastrana, L.; Rúa, M. L. (2011): Production and characterization of two N-terminal truncated esterases from *Thermus thermophilus* HB27 in a mesophilic yeast: effect of N-terminus in thermal activity and stability. In: *Protein Expr. Purif.* 78 (2), S. 120–130.
- Fuciños, P.; Domínguez, A.; Sanromán, M. A.; Longo, M. A.; Rúa, M. L.; Pastrana, L. (2005b): Production of thermostable lipolytic activity by *Thermus* species. In: *Biotechnol. Prog.* 21 (4), S. 1198–1205.
- Fujita, A.; Misumi, Y.; Honda, S.; Sato, T.; Koyama, Y. (2013): Construction of new cloning vectors that employ the phytoene synthase encoding gene for color screening of cloned DNA inserts in *Thermus thermophilus*. In: *Gene* 527 (2), S. 655–662.
- Gabor, E. M.; Alkema, W. B. L.; Janssen, D. B. (2004): Quantifying the accessibility of the metagenome by random expression cloning techniques. In: *Environ. Microbiol.* 6 (9), S. 879–886.

- Geissler, A. (2011): Untersuchungen zur Regulation des lichtabhängigen Promotors PcrB in *Thermus thermophilus* HB27. Masterthesis. Technische Universität München, Lehrstuhl für Mikrobiologie.
- Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. (2009): Enzymatic assembly of DNA molecules up to several hundred kilobases. In: *Nat. Methods* 6 (5), S. 343–345.
- Girfoglio, M.; Rossi, M.; Cannio, R. (2012): Cellulose degradation by *Sulfolobus solfataricus* requires a cell-anchored endo- $\beta$ -1-4-glucanase. In: *J. Bacteriol.* 194 (18), S. 5091–5100.
- Glass, J. I.; Assad-Garcia, N.; Alperovich, N.; Yooseph, S.; Lewis, M. R.; Maruf, M. *et al.* (2006): Essential genes of a minimal bacterium. In: *Proc. Natl. Acad. Sci. U.S.A.* 103 (2), S. 425–430.
- Gomes, J.; Steiner, W. (2004): The Biocatalytic Potential of Extremophiles and Extremozymes. In: *Food Technol. Biotechnol.* 42, S. 223–235.
- Gong, X.; Gruninger, R. J.; Qi, M.; Paterson, L.; Forster, R. J.; Teather, R. M.; McAllister, T. A. (2012): Cloning and identification of novel hydrolase genes from a dairy cow rumen metagenomic library and characterization of a cellulase gene. In: *BMC Res Notes* 5, S. 566.
- Grado, M. de; Castán, P.; Berenguer, J. (1999): A high-transformation-efficiency cloning vector for *Thermus thermophilus*. In: *Plasmid* 42 (3), S. 241–245.
- Grado, M. de; Lasa, I.; Berenguer, J. (1998): Characterization of a plasmid replicative origin from an extreme thermophile. In: *FEMS Microbiol. Lett.* 165 (1), S. 51–57.
- Graham, J. E.; Clark, M. E.; Nadler, D. C.; Huffer, S.; Chokhawala, H. A.; Rowland, S. E. *et al.* (2011): Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment. In: *Nat Comms* 2, S. 375.
- Hagelueken, G.; Adams, T. M.; Wiehlmann, L.; Widow, U.; Kolmar, H.; Tümmler, B. *et al.* (2006): The crystal structure of SdsA1, an alkylsulfatase from *Pseudomonas aeruginosa*, defines a third class of sulfatases. In: *Proc. Natl. Acad. Sci. U.S.A.* 103 (20), S. 7631–7636.
- Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. (1998): Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. In: *Chem. Biol.* 5 (10), S. R245-9.
- Harley, C. B.; Reynolds, R. P. (1987): Analysis of *E. coli* promoter sequences. In: *Nucleic Acids Res.* 15 (5), S. 2343–2361.
- Hausmann, S.; Jaeger, K.-E (2010): Lipolytic Enzymes from Bacteria. In: Kenneth N Timmis (Hg.): *Handbook of Hydrocarbon and Lipid Microbiology*: Springer Berlin Heidelberg, S. 1099-1126. Online verfügbar unter [http://dx.doi.org/10.1007/978-3-540-77587-4\\_77](http://dx.doi.org/10.1007/978-3-540-77587-4_77).
- Healy, F. G.; Ray, R. M.; Aldrich, H. C.; Wilkie, A. C.; Ingram, L. O.; Shanmugam, K. T. (1995): Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. In: *Appl. Microbiol. Biotechnol.* 43 (4), S. 667–674.
- Heckman, K. L.; Pease, L. R. (2007): Gene splicing and mutagenesis by PCR-driven overlap extension. In: *Nat Protoc* 2 (4), S. 924–932.
- Henne, A.; Brüggemann, H.; Raasch, C.; Wiezer, A.; Hartsch, T.; Liesegang, H. *et al.* (2004): The genome sequence of the extreme thermophile *Thermus thermophilus*. In: *Nat. Biotechnol.* 22 (5), S. 547–553.
- Hidalgo, A.; Betancor, L.; Moreno, R.; Zafra, O.; Cava, F.; Fernandez-Lafuente, R. *et al.* (2004): *Thermus thermophilus* as a Cell Factory for the Production of a Thermophilic Mn-Dependent Catalase Which Fails To Be Synthesized in an Active Form in *Escherichia coli*. In: *Applied and Environmental Microbiology* 70 (7), S. 3839–3844.
- Hobman, J. L. (2007): MerR family transcription activators: similar designs, different specificities. In: *Mol. Microbiol.* 63 (5), S. 1275–1278.
- Holben, W. E.; Harris, D. (1995): DNA-based monitoring of total bacterial community structure in environmental samples. In: *Mol. Ecol.* 4 (5), S. 627–631.
- Holben, W. E.; Feris, K. P.; Kettunen, A.; Apajalahti, J. H. A. (2004): GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis. In: *Appl. Environ. Microbiol.* 70 (4), S. 2263–2270.
- Hoshino, T.; Fujii, R.; Nakahara, T. (1993): Molecular cloning and sequence analysis of the crtB gene of *Thermus thermophilus* HB27, an extreme thermophile producing carotenoid pigments. In: *Appl. Environ. Microbiol.* 59 (9), S. 3150–3153.

- Hotta, Y.; Ezaki, S.; Atomi, H.; Imanaka, T. (2002): Extremely stable and versatile carboxylesterase from a hyperthermophilic archaeon. In: *Appl. Environ. Microbiol.* 68 (8), S. 3925–3931.
- Huang, Y.; Krauss, G.; Cottaz, S.; Driguez, H.; Lipps, G. (2005): A highly acid-stable and thermostable endo-beta-glucanase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. In: *Biochem. J.* 385 (Pt 2), S. 581–588.
- Huber, R.; Burggraf, S.; Mayer, T.; Barns, S. M.; Rossnagel, P.; Stetter, K. O. (1995): Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. In: *Nature* 376 (6535), S. 57–58.
- Hugenholtz, P.; Stackebrandt, E. (2004): Reclassification of *Sphaerobacter thermophilus* from the subclass Sphaerobacteridae in the phylum Actinobacteria to the class Thermomicrobia (emended description) in the phylum Chloroflexi (emended description). In: *Int. J. Syst. Evol. Microbiol.* 54 (Pt 6), S. 2049–2051.
- Huson, D. H.; Mitra, S. (2012): Introduction to the analysis of environmental sequences: metagenomics with MEGAN. In: *Methods Mol. Biol.* 856, S. 415–429.
- Inoue, H.; Nojima, H.; Okayama, H. (1990): High efficiency transformation of *Escherichia coli* with plasmids. In: *Gene* 96 (1), S. 23–28.
- Iwai, S.; Chai, B.; Jesus, E. da C.; Penton, C. R.; Lee, T. K.; Cole, J. R.; Tiedje, J. M. (2011): Gene-Targeted Metagenomics (GT Metagenomics) to Explore the Extensive Diversity of Genes of Interest in Microbial Communities. In: Frans J. de Bruijn (Hg.): *Handbook of Molecular Microbial Ecology I*. Hoboken, NJ, USA: John Wiley & Sons, Inc, S. 235–243.
- Jacobsen, C. S.; Rasmussen, O. F. (1992): Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. In: *Appl. Environ. Microbiol.* 58 (8), S. 2458–2462.
- Jaeger, K.-E.; Dijkstra, B. W.; Reetz, M. T. (1999): Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases. In: *Annu. Rev. Microbiol.* 53, S. 315–351.
- Jaeger, K.-E.; Eggert, T. (2002): Lipases for biotechnology. In: *Curr. Opin. Biotechnol.* 13 (4), S. 390–397.
- Jenney, F. E.; Adams, M. W. W. (2008): The impact of extremophiles on structural genomics (and vice versa). In: *Extremophiles* 12 (1), S. 39–50.
- Jensen, S.; Samuelson, O. B.; Andersen, K.; Torkildsen, L.; Lambert, C.; Choquet, G. *et al.* (2003): Characterization of strains of *Vibrio splendidus* and *V. tapetis* isolated from corkwing wrasse *Symphodus melops* suffering vibriosis. In: *Dis. Aquat. Org.* 53 (1), S. 25–31.
- Jeon, J. H.; Kim, S.-J.; Lee, H. S.; Cha, S.-S.; Lee, J. H.; Yoon, S.-H. *et al.* (2011): Novel metagenome-derived carboxylesterase that hydrolyzes  $\beta$ -lactam antibiotics. In: *Appl Environ Microbiol.* 77 (21), S. 7830–6.
- Jiang, Z.; Wang, H.; Ma, Y.; Wei, D. (2006): Characterization of two novel lipase genes isolated directly from environmental sample. In: *Appl. Microbiol. Biotechnol.* 70 (3), S. 327–332.
- Kakirde, K. S.; Parsley, L. C.; Liles, M. R. (2010): Size Does Matter: Application-driven Approaches for Soil Metagenomics. In: *Soil Biol. Biochem.* 42 (11), S. 1911–1923.
- Kakirde, K. S.; Wild, J.; Godiska, R.; Mead, D. A.; Wiggins, A. G.; Goodman, R. M. *et al.* (2011): Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. In: *Gene* 475 (2), S. 57–62.
- Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. (2014): Data, information, knowledge and principle: back to metabolism in KEGG. In: *Nucleic Acids Res.* 42 (1), S. D199–205.
- Keeling, P. J.; Doolittle, W. F. (1995): Archaea: narrowing the gap between prokaryotes and eukaryotes. In: *Proc. Natl. Acad. Sci. U.S.A.* 92 (13), S. 5761–5764.
- Kellner, H.; Luis, P.; Portetelle, D.; Vandenbol, M. (2011): Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants. In: *Microbiol. Res.* 166 (5), S. 360–368.
- Kim, D.; Park, B. H.; Jung, B.-W.; Kim, M.-K.; Hong, S.-I.; Lee, D.-S. (2006): Identification and molecular modeling of a family 5 endocellulase from *Thermus caldophilus* GK24, a cellulolytic strain of *Thermus thermophilus*. In: *IJMS* 7 (12), S. 571–589.
- Kim, H.-W.; Kataoka, M.; Ishikawa, K. (2012): Atomic resolution of the crystal structure of the hyperthermophilic family 12 endocellulase and stabilizing role of the DxDxDG calcium-binding motif in *Pyrococcus furiosus*. In: *FEBS Lett.* 586 (7), S. 1009–1013.
- Kim, S.-B.; Lee, W.; Ryu, Y.-W. (2008): Cloning and characterization of thermostable esterase from *Archaeoglobus fulgidus*. In: *J. Microbiol.* 46 (1), S. 100–107.

- Kouker, G.; Jaeger, K.-E. (1987): Specific and sensitive plate assay for bacterial lipases. In: *Appl. Environ. Microbiol.* 53 (1), S. 211–213.
- Koyama, Y.; Hoshino, T.; Tomizuka, N.; Furukawa, K. (1986): Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp. In: *J. Bacteriol.* 166 (1), S. 338–340.
- Krogh, A.; Larsson, B.; Heijne, G. von; Sonnhammer, E. L. (2001): Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. In: *J. Mol. Biol.* 305 (3), S. 567–580.
- Kudla, G.; Murray, A. W.; Tollervey, D.; Plotkin, J. B. (2009): Coding-sequence determinants of gene expression in *Escherichia coli*. In: *Science* 324 (5924), S. 255–258.
- Lakhdari, O.; Cultrone, A.; Tap, J.; Gloux, K.; Bernard, F.; Ehrlich, S. D. *et al.* (2010): Functional metagenomics: a high throughput screening method to decipher microbiota-driven NF- $\kappa$ B modulation in the human gut. In: *PLoS ONE* 5 (9).
- Lane, D.J (1991): 16S/23S rRNA sequencing. In: E. Stackebrandt und M. Goodfellow (Hg.): *Nucleic acid techniques in bacterial systematics*. Chichester, New York: Wiley (Modern microbiological methods), S. 115–175.
- Le Breton, Y.; Mohapatra, N. P.; Haldenwang, W. G. (2006): In vivo random mutagenesis of *Bacillus subtilis* by use of TnYLB-1, a mariner-based transposon. In: *Appl. Environ. Microbiol.* 72 (1), S. 327–333.
- Lee, S.-T.; Lee, J. J. (1997): Insoluble dye substrate for screening and assay of xylan-degrading enzymes. In: *Journal of Microbiological Methods* 29 (1), S. 1–5.
- Leis, B.; Angelov, A.; Liebl, W. (2013): Screening and expression of genes from metagenomes. In: *Adv. Appl. Microbiol.* 83, S. 1–68.
- Leis, B.; Angelov, A.; Mientus, M.; Li, H.; Pham, V. T. T.; Lauinger, B.; Bongen, P.; Pietruszka, J.; Gonçalves, L. G.; Santos, H.; Liebl, W. (2015a): Identification of novel esterase-active enzymes from hot environments by use of the host bacterium *Thermus thermophilus*. *Front. Microbiol.* 6, 95.
- Leis, B.; Heinze, S.; Angelov, A.; Pham, V. T. T.; Thürmer, A.; Jebbar, M.; Golyshin, P. N.; Streit, W. R.; Daniel, R.; Liebl, W. (2015b): Functional Screening of Hydrolytic Activities Reveals an Extremely Thermostable Cellulase from a Deep-Sea Archaeon. *Front. Bioeng. Biotechnol.* 3, 275.
- Leis, B.; Angelov, A.; Li, H.; Liebl, W. (2014): Genetic analysis of lipolytic activities in *Thermus thermophilus* HB27. *J. Biotechnol.* 191, 150–157.
- Levisson, M.; van der Oost, J.; Kengen, S. W. M. (2009): Carboxylic ester hydrolases from hyperthermophiles. In: *Extremophiles* 13 (4), S. 567–581.
- Lichev, A. (2011): Genetisches Screening für die Identifikation lichtempfindlicher Promotoren in *Thermus thermophilus* HB27. Bachelorthesis. Technische Universität München, Lehrstuhl für Mikrobiologie.
- Liebl, W.; Götz, F. (1986): Studies on lipase directed export of *Escherichia coli* beta-lactamase in *Staphylococcus carnosus*. In: *Mol. Gen. Genet.* 204 (1), S. 166–173.
- Liebl, W.; Ruile, P.; Bronnenmeier, K.; Riedel, K.; Lottspeich, F.; Greif, I. (1996): Analysis of a *Thermotoga maritima* DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes. In: *Microbiology (Reading, Engl.)* 142 (Pt 9), S. 2533–2542.
- Liebl, W. (2004): Genomics taken to the extreme. In: *Nat. Biotechnol.* 22 (5), S. 524–525.
- Liebl, W. (2011): Metagenomics. In: Joachim Reitner und Volker Thiel (Hg.): *Encyclopedia of Geobiology: Springer Netherlands (Encyclopedia of Earth Sciences Series)*, S. 553–558. Online verfügbar unter [http://dx.doi.org/10.1007/978-1-4020-9212-1\\_133](http://dx.doi.org/10.1007/978-1-4020-9212-1_133).
- Limauro, D.; Cannio, R.; Fiorentino, G.; Rossi, M.; Bartolucci, S. (2001): Identification and molecular characterization of an endoglucanase gene, celS, from the extremely thermophilic archaeon *Sulfolobus solfataricus*. In: *Extremophiles* 5 (4), S. 213–219.
- Lin, Z.; Zheng, J. (2010): Occurrence, characteristics, and applications of fructosyl amine oxidases (amadoriases). In: *Appl. Microbiol. Biotechnol.* 86 (6), S. 1613–1619.
- Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. (2013): The carbohydrate-active enzymes database (CAZy) in 2013. In: *Nucleic Acids Res.*
- Lorenz, P.; Eck, J. (2005): Metagenomics and industrial applications. In: *Nat. Rev. Microbiol.* 3 (6), S. 510–516.

- Lussier, F.-X.; Chambenoit, O.; Côté, A.; Hupé, J.-F.; Denis, F.; Juteau, P. *et al.* (2011): Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector. In: *J. Ind. Microbiol. Biotechnol.* 38 (9), S. 1321–1328.
- MacElroy, R. D. (1974): Some comments on the evolution of extremophiles. In: *Biosystems* 6 (1), S. 74–75. Online verfügbar unter <http://www.sciencedirect.com/science/article/pii/0303264774900264>.
- Madigan, M. T.; Martinko, J. M.; Parker, J.; Brock, T. D.; Goebel, W. (2001): *Mikrobiologie*. Berlin [u.a.]: Spektrum Akad. Verl. (Spektrum-Lehrbuch).
- Martinez, A.; Kolvek, S. J.; Yip, C. L. T.; Hopke, J.; Brown, K. A.; MacNeil, I. A.; Osburne, M. S. (2004): Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. In: *Appl. Environ. Microbiol.* 70 (4), S. 2452–2463.
- Martínez-Martínez, M.; Alcaide, M.; Tchigvintsev, A.; Reva, O.; Polaina, J.; Bargiela, R. *et al.* (2013): Biochemical diversity of carboxyl esterases and lipases from Lake Arreo (Spain): a metagenomic approach. In: *Appl. Environ. Microbiol.* 79 (12), S. 3553–3562.
- Maseda, H.; Hoshino, T. (1995): Screening and analysis of DNA fragments that show promoter activities in *Thermus thermophilus*. In: *FEMS Microbiol. Lett.* 128 (2), S. 127–134.
- McMahon, M. D.; Guan, C.; Handelsman, J.; Thomas, M. G. (2012): Metagenomic analysis of *Streptomyces lividans* reveals host-dependent functional expression. In: *Appl. Environ. Microbiol.* 78 (10), S. 3622–3629.
- Mergulhão, F. J. M.; Summers, D. K.; Monteiro, G. A. (2005): Recombinant protein secretion in *Escherichia coli*. In: *Biotechnol. Adv.* 23 (3), S. 177–202.
- Meyer, Q. C.; Burton, S. G.; Cowan, D. A. (2007): Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. In: *Biotechnol J* 2 (1), S. 36–40.
- Michaelis, C. A. (2011): Functional Analysis of the Light-Dependent Transcriptional Regulator LitR of *Thermus thermophilus* HB27. Bachelorthesis. Technische Universität München, Lehrstuhl für Mikrobiologie.
- Mientus, M.; Brady, S.; Angelov, A.; Zimmermann, P.; Wemheuer, B.; Schuldes, J. *et al.* (2013): Thermostable Xylanase and  $\beta$ -Glucanase Derived from the Metagenome of the Avachinsky Crater in Kamchatka (Russia). In: *CBIOT* 2 (4), S. 284–293.
- Miller, G. L. (1959): Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. In: *Anal. Chem.* 31 (3), S. 426–428.
- Mokoena, N.; Mathiba, K.; Tsekoa, T.; Steenkamp, P.; Rashamuse, K. (2013): Functional characterisation of a metagenome derived family VIII esterase with a deacetylation activity on  $\beta$ -lactam antibiotics. In: *Biochem Biophys Res Commun.* 437 (3), S. 342–8.
- Monnier, V. M. (2005): Bacterial enzymes that can deglycate glucose- and fructose-modified lysine. In: *Biochem. J.* 392 (2), S. e1–3.
- Moraïs, S.; Barak, Y.; Lamed, R.; Wilson, D. B.; Xu, Q.; Himmel, M. E.; Bayer, E. A. (2012): Paradigmatic status of an endo- and exoglucanase and its effect on crystalline cellulose degradation. In: *Biotechnol Biofuels* 5 (1), S. 78.
- Moreno, R.; Haro, A.; Castellanos, A.; Berenguer, J. (2005): High-level overproduction of His-tagged Tth DNA polymerase in *Thermus thermophilus*. In: *Appl. Environ. Microbiol.* 71 (1), S. 591–593.
- Nakamura, T.; Mine, S.; Hagihara, Y.; Ishikawa, K.; Ikegami, T.; Uegaki, K. (2008): Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from *Pyrococcus furiosus*. In: *J. Mol. Biol.* 381 (3), S. 670–680.
- Nimchua, T.; Thongaram, T.; Uengwetwanit, T.; Pongpattanakitsote, S.; Eurwilaichitr, L. (2012): Metagenomic analysis of novel lignocellulose-degrading enzymes from higher termite guts inhabiting microbes. In: *J. Microbiol. Biotechnol.* 22 (4), S. 462–469.
- Ogram, A.; Sayler, G. S.; Barkay, T. (1987): The extraction and purification of microbial DNA from sediments. In: *Journal of Microbiological Methods* 7 (2–3), S. 57–66. Online verfügbar unter <http://www.sciencedirect.com/science/article/pii/016770128790025X>.
- Ohta, T.; Tokishita, S.-I.; Imazuka, R.; Mori, I.; Okamura, J.; Yamagata, H. (2006): beta-Glucosidase as a reporter for the gene expression studies in *Thermus thermophilus* and constitutive expression of DNA repair genes. In: *Mutagenesis* 21 (4), S. 255–260.

- Ohtani, N.; Tomita, M.; Itaya, M. (2010): An extreme thermophile, *Thermus thermophilus*, is a polyploid bacterium. In: *J. Bacteriol.* 192 (20), S. 5499–5505.
- Olsen, G. J.; Lane, D. J.; Giovannoni, S. J.; Pace, N. R.; Stahl, D. A. (1986): Microbial ecology and evolution: a ribosomal RNA approach. In: *Annu. Rev. Microbiol.* 40, S. 337–365.
- Ortiz-Guerrero, J. M.; Polanco, M. C.; Murillo, F. J.; Padmanabhan, S.; Elías-Arnanz, M. (2011): Light-dependent gene regulation by a coenzyme B12-based photoreceptor. In: *Proc. Natl. Acad. Sci. U.S.A.* 108 (18), S. 7565–7570.
- Oshima, T.; Imahori, K. (1974): Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a Nonsporulating Thermophilic Bacterium from a Japanese Thermal Spa. In: *International Journal of Systematic Bacteriology* 24 (1), S. 102–112.
- Overbeek, R.; Begley, T.; Butler, R. M.; Choudhuri, J. V.; Chuang, H.-Y.; Cohoon, M. *et al.* (2005): The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. In: *Nucleic Acids Res.* 33 (17), S. 5691–5702.
- Parachin, N. S.; Gorwa-Grauslund, M. F. (2011): Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library. In: *Biotechnol Biofuels* 4, S. 9.
- Park, H.-S.; Kayser, K. J.; Kwak, J.-H.; Kilbane, J. J. (2004): Heterologous gene expression in *Thermus thermophilus*: beta-galactosidase, dibenzothiophene monooxygenase, PNB carboxy esterase, 2-aminobiphenyl-2,3-diol dioxygenase, and chloramphenicol acetyl transferase. In: *J. Ind. Microbiol. Biotechnol.* 31 (4), S. 189–197.
- Park, S.-J.; Kang, C.-H.; Chae, J.-C.; Rhee, S.-K. (2008): Metagenome microarray for screening of fosmid clones containing specific genes. In: *FEMS Microbiol. Lett.* 284 (1), S. 28–34.
- Pathak, G. P.; Gärtner, W. (2010): Detection and isolation of selected genes of interest from metagenomic libraries by a DNA microarray approach. In: *Methods Mol. Biol.* 668, S. 299–312.
- Peng, Q.; Zhang, X.; Shang, M.; Wang, X.; Wang, G.; Li, B. *et al.* (2011): A novel esterase gene cloned from a metagenomic library from neritic sediments of the South China Sea. In: *Microb Cell Fact* 10 (1), S. 95.
- Petersen, T. N.; Brunak, S.; Heijne, G. von; Nielsen, H. (2011): SignalP 4.0: discriminating signal peptides from transmembrane regions. In: *Nat. Methods* 8 (10), S. 785–786.
- Punta, Marco; Coghill, Penny C.; Eberhardt, Ruth Y.; Mistry, Jaina; Tate, John; Bournnell, Chris *et al.* (2012): The Pfam protein families database. In: *Nucleic Acids Res.* 40 (Database issue), S. D290-301.
- Ramírez-Arcos, S.; Fernández-Herrero, L. A.; Berenguer, J. (1998a): A thermophilic nitrate reductase is responsible for the strain specific anaerobic growth of *Thermus thermophilus* HB8. In: *Biochim. Biophys. Acta* 1396 (2), S. 215–227.
- Ramírez-Arcos, S.; Fernández-Herrero, L. A.; Marín, I.; Berenguer, J. (1998b): Anaerobic growth, a property horizontally transferred by an Hfr-like mechanism among extreme thermophiles. In: *J. Bacteriol.* 180 (12), S. 3137–3143.
- Reigstad, L. J.; Bartossek, R.; Schleper, C. (2011): Preparation of high-molecular weight DNA and metagenomic libraries from soils and hot springs. In: *Meth. Enzymol.* 496, S. 319–344.
- Rhee, J.-K.; Ahn, D.-G.; Kim, Y.-G.; Oh, J.-W. (2005): New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library. In: *Appl. Environ. Microbiol.* 71 (2), S. 817–825.
- Robertson, H. M.; Lampe, D. J. (1995): Distribution of transposable elements in arthropods. In: *Annu. Rev. Entomol.* 40, S. 333–357.
- Rutherford, K.; Parkhill, J.; Crook, J.; Horsnell, T.; Rice, P.; Rajandream, M. A.; Barrell, B. (2000): Artemis: sequence visualization and annotation. In: *Bioinformatics* 16 (10), S. 944–945.
- Sakamoto, K.; Agari, Y.; Agari, K.; Kuramitsu, S.; Shinkai, A. (2010): Structural and functional characterization of the transcriptional repressor CsoR from *Thermus thermophilus* HB8. In: *Microbiology (Reading, Engl.)* 156 (Pt 7), S. 1993–2005.
- Sambrook, J.; Fritsch, E. F.; Maniatis, T.; Sambrook-Fritsch-Maniatis (1989): Molecular cloning. A laboratory manual; [prepared for use in the CSH course on the molecular cloning of eukaryotic genes]. 2. Aufl. Plainview, NY: Cold Spring Harbor Laboratory Press.

- Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, C. A. *et al.* (1977): Nucleotide sequence of bacteriophage phi X174 DNA. In: *Nature* 265 (5596), S. 687–695.
- Santos, C. R.; Paiva, J. H.; Sforça, M. L.; Neves, J. L.; Navarro, R. Z.; Cota, J. *et al.* (2012): Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. In: *Biochem. J.* 441 (1), S. 95–104.
- Satokari, R. M.; Vaughan, E. E.; Akkermans, A. D.; Saarela, M.; Vos, W. M. de (2001): Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. In: *Appl. Environ. Microbiol.* 67 (2), S. 504–513.
- Saxena, S.; Bahadur, J.; Varma, A. (1992): Effect of cobalt and nickel on growth and carboxymethyl cellulase activity of *Cellulomonas* spp. In: *Biometals* 5 (4), S. 209–212. Online verfügbar unter <http://dx.doi.org/10.1007/BF01061219>.
- Schallmeyer, M.; Ly, A.; Wang, C.; Meglei, G.; Voget, S.; Streit, W. R. *et al.* (2011): Harvesting of novel polyhydroxyalkanoate (PHA) synthase encoding genes from a soil metagenome library using phenotypic screening. In: *FEMS Microbiol. Lett.* 321 (2), S. 150–156.
- Schmidt, T. M.; DeLong, E. F.; Pace, N. R. (1991): Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. In: *J. Bacteriol.* 173 (14), S. 4371–4378.
- Schwarzenlander, C.; Averhoff, B. (2006): Characterization of DNA transport in the thermophilic bacterium *Thermus thermophilus* HB27. In: *FEBS J.* 273 (18), S. 4210–4218.
- Seegerer, A. H.; Trincone, A.; Gahrtz, M.; Stetter, K. O. (1991): *Stygiolobus azoricus* gen. nov., sp. nov. Represents a Novel Genus of Anaerobic, Extremely Thermoacidophilic Archaeobacteria of the Order Sulfolobales. In: *International Journal of Systematic Bacteriology* 41 (4), S. 495–501.
- Selvarani, P.; Shanthi, V.; Rajesh, C. K.; Saravanan, S.; Sekar, K. (2004): BSDD: Biomolecules Segment Display Device--a web-based interactive display tool. In: *Nucleic Acids Res.* 32 (Web Server issue), S. W645–8.
- Sevostyanova, A.; Djordjevic, M.; Kuznedelov, K.; Naryshkina, T.; Gelfand, M. S.; Severinov, K.; Minakhin, L. (2007): Temporal regulation of viral transcription during development of *Thermus thermophilus* bacteriophage phiYS40. In: *J. Mol. Biol.* 366 (2), S. 420–435.
- She, Q.; Singh, R. K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M. J. *et al.* (2001): The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. In: *Proc. Natl. Acad. Sci. U.S.A.* 98 (14), S. 7835–7840.
- Shokralla, S.; Spall, J. L.; Gibson, J. F.; Hajibabaei, M. (2012): Next-generation sequencing technologies for environmental DNA research. In: *Mol. Ecol.* 21 (8), S. 1794–1805.
- Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W. *et al.* (2011): Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. In: *Mol. Syst. Biol.* 7, S. 539.
- Sigrist, C. J. A.; Castro, E. de; Cerutti, L.; Cuche, B. A.; Hulo, N.; Bridge, A. *et al.* (2013): New and continuing developments at PROSITE. In: *Nucleic Acids Res.* 41 (Database issue), S. D344–7.
- Sigurgísladóttir, S.; Konráðsdóttir M.; Jónsson Á.; Kristjánsson J.K; and Matthiasson E. (1993): Lipase activity of thermophilic bacteria from icelandic hot springs. In: *Biotechnol Lett* 15 (4), S. 361–366.
- Simon, C.; Daniel, R. (2009): Achievements and new knowledge unraveled by metagenomic approaches. In: *Appl. Microbiol. Biotechnol.* 85 (2), S. 265–276.
- Simon, C.; Daniel, R. (2011): Metagenomic analyses: past and future trends. In: *Appl. Environ. Microbiol.* 77 (4), S. 1153–1161.
- Simon, C.; Herath, J.; Rockstroh, S.; Daniel, R. (2009): Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. In: *Appl. Environ. Microbiol.* 75 (9), S. 2964–2968.
- Simpson, P. J.; Xie, H.; Bolam, D. N.; Gilbert, H. J.; Williamson, M. P. (2000): The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. In: *J. Biol. Chem.* 275 (52), S. 41137–41142.
- Snyder, L.; Champness, W. (2007): *Molecular genetics of bacteria*. 3. Aufl. Washington, D.C: ASM Press.
- Sommer, M. O. A.; Church, G. M.; Dantas, G. (2010): A functional metagenomic approach for expanding the synthetic biology toolbox for biomass conversion. In: *Mol. Syst. Biol.* 6, S. 360.
- Sorek, R.; Zhu, Y.; Creevey, C. J.; Francino, M. P.; Bork, P.; Rubin, E. M. (2007): Genome-wide experimental determination of barriers to horizontal gene transfer. In: *Science* 318 (5855), S. 1449–1452.

- Sørensen, H. P.; Mortensen, K. K. (2005a): Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. In: *J. Biotechnol.* 115 (2), S. 113–128.
- Sørensen, H. P.; Mortensen, K. K. (2005b): Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. In: *Microb. Cell Fact.* 4 (1), S. 1.
- Southern, E.M (1975): Detection of specific sequences among DNA fragments separated by gel electrophoresis. In: *Journal of Molecular Biology* 98 (3), S. 503–517.
- Staden, R.; Beal, K. F.; Bonfield, J. K. (2000): The Staden package, 1998. In: *Methods Mol. Biol.* 132, S. 115–130.
- Staley, J. T.; Konopka, A. (1985): Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. In: *Annu. Rev. Microbiol.* 39, S. 321–346.
- Steele, H. L.; Jaeger, K.-E.; Daniel, R.; Streit, W. R. (2009): Advances in recovery of novel biocatalysts from metagenomes. In: *J. Mol. Microbiol. Biotechnol.* 16 (1-2), S. 25–37.
- Sternier, R.; Liebl, W. (2001): Thermophilic adaptation of proteins. In: *Crit. Rev. Biochem. Mol. Biol.* 36 (1), S. 39–106.
- Takano, H.; Kondo, M.; Usui, N.; Usui, T.; Ohzeki, H.; Yamazaki, R. *et al.* (2011): Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in *Thermus thermophilus*. In: *J. Bacteriol.* 193 (10), S. 2451–2459.
- Tamakoshi, M.; Yaoi, T.; Oshima, T.; Yamagishi, A. (1999): An efficient gene replacement and deletion system for an extreme thermophile, *Thermus thermophilus*. In: *FEMS Microbiology Letters* 173 (2), S. 431–437.
- 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. In: *Mol. Biol. Evol.* 28 (10), S. 2731–2739.
- Tanaka, T.; Fujiwara, S.; Nishikori, S.; Fukui, T.; Takagi, M.; Imanaka, T. (1999): A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. In: *Appl. Environ. Microbiol.* 65 (12), S. 5338–5344.
- Taupp, M.; Mewis, K.; Hallam, S. J. (2011): The art and design of functional metagenomic screens. In: *Curr. Opin. Biotechnol.* 22 (3), S. 465–472.
- Tekedar, H. C.; Sanlı-Mohamed, G. (2011): Molecular cloning, over expression and characterization of thermoalkalophilic esterases isolated from *Geobacillus* sp. In: *Extremophiles* 15 (2), S. 203–211.
- Ten, L. N.; Im, W-T; Kim, M-K; Lee, S-T (2005): A plate assay for simultaneous screening of polysaccharide- and protein-degrading micro-organisms. In: *Lett. Appl. Microbiol.* 40 (2), S. 92–98.
- Terpe, K. (2006): Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. In: *Appl. Microbiol. Biotechnol.* 72 (2), S. 211–222.
- Terzi, F. (2012): Licht-induzierbare Expression heterologer Gene in *Thermus thermophilus* HB27. Masterthesis. Technische Universität München, Lehrstuhl für Mikrobiologie.
- Thauer, R. K.; Jungermann, K.; Decker, K. (1977): Energy conservation in chemotrophic anaerobic bacteria. In: *Bacteriol Rev* 41 (1), S. 100–180.
- Thomas, T.; Gilbert, J.; Meyer, F. (2012): Metagenomics - a guide from sampling to data analysis. In: *Microb Inform Exp* 2 (1), S. 3.
- Tirawongsaroj, P.; Sriprang, R.; Harnpicharnchai, P.; Thongaram, T.; Champreda, V.; Tanapongpipat, S. *et al.* (2008): Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. In: *J. Biotechnol.* 133 (1), S. 42–49.
- Tolstrup, N.; Sensen, C. W.; Garrett, R. A.; Clausen, I. G. (2000): Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. In: *Extremophiles* 4 (3), S. 175–179.
- Torsvik, V.; Øvreås, L.; Thingstad, T. F. (2002): Prokaryotic Diversity--Magnitude, Dynamics, and Controlling Factors. In: *Science* 296 (5570), S. 1064–1066.
- Torsvik, V.; Goksøyr, J.; Daae, F. L. (1990): High diversity in DNA of soil bacteria. In: *Appl. Environ. Microbiol.* 56 (3), S. 782–787.
- Torsvik, V. L.; Goksøyr, J. (1978): Determination of bacterial DNA in soil. In: *Soil Biology and Biochemistry* 10 (1), S. 7–12.

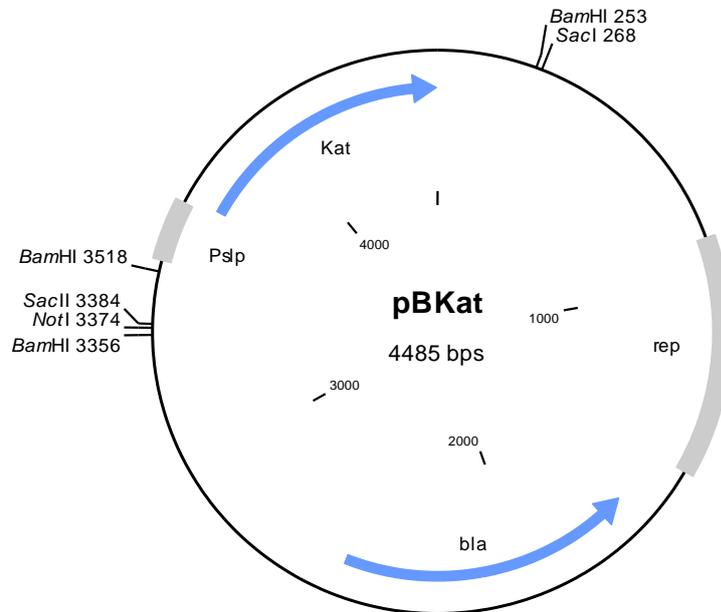
- Tyo, K. E. J.; Ajikumar, P. K.; Stephanopoulos, G. (2009): Stabilized gene duplication enables long-term selection-free heterologous pathway expression. In: *Nat. Biotechnol.* 27 (8), S. 760–765.
- Uchiyama, T.; Abe, T.; Ikemura, T.; Watanabe, K. (2005): Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. In: *Nat. Biotechnol.* 23 (1), S. 88–93.
- Uchiyama, T.; Miyazaki, K. (2010): Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. In: *Appl. Environ. Microbiol.* 76 (21), S. 7029–7035.
- Uyaguari, M. I.; Fichot, E. B.; Scott, G. I.; Norman, R. S. (2011): Characterization and quantitation of a novel  $\beta$ -lactamase gene found in a wastewater treatment facility and the surrounding coastal ecosystem. In: *Appl. Environ. Microbiol.* 77 (23), S. 8226–8233.
- van Aalten, D. M.; Komander, D.; Synstad, B.; Gåseidnes, S.; Peter, M. G.; Eijssink, V. G. (2001): Structural insights into the catalytic mechanism of a family 18 exo-chitinase. In: *Proc. Natl. Acad. Sci. U.S.A.* 98 (16), S. 8979–8984.
- Verger, R. (1997): Interfacial activation of lipases: facts and artifacts. In: *Trends Biotechnol.* 15 (1), S. 32–38. Online verfügbar unter <http://linkinghub.elsevier.com/retrieve/pii/S016779996100640>.
- Verma, D.; Kawarabayasi, Y.; Miyazaki, K.; Satyanarayana, T. (2013): Cloning, expression and characteristics of a novel alkalistable and thermostable xylanase encoding gene (Mxyl) retrieved from compost-soil metagenome. In: *PLoS ONE* 8 (1), S. e52459.
- Villegas, A.; Kropinski, A. M. (2008): An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation. In: *Microbiology (Reading, Engl.)* 154 (Pt 9), S. 2559–2661.
- Wang, B.; Lu, D.; Gao, R.; Yang, Z.; Cao, S.; Feng, Y. (2004): A novel phospholipase A2/esterase from hyperthermophilic archaeon *Aeropyrum pernix* K1. In: *Protein Expr. Purif.* 35 (2), S. 199–205.
- Wang, G.; Meng, K.; Luo, H.; Wang, Y.; Huang, H.; Shi, P. *et al.* (2012): Phylogenetic Diversity and Environment-Specific Distributions of Glycosyl Hydrolase Family 10 Xylanases in Geographically Distant Soils. In: *PLoS ONE* 7 (8), S. e43480.
- Wang, Q.; Yang, G.; Liu, Y.; Feng, Y. (2006): Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic *Aeropyrum pernix* K1 by a single mutation. In: *J. Biol. Chem.* 281 (27), S. 18618–18625.
- Warren, R. L.; Freeman, J. D.; Levesque, R. C.; Smailus, D. E.; Flibotte, S.; Holt, R. A. (2008): Transcription of foreign DNA in *Escherichia coli*. In: *Genome Res.* 18 (11), S. 1798–1805.
- Whitman, W. B.; Coleman, D. C.; Wiebe, W. J. (1998): Prokaryotes: the unseen majority. In: *Proc. Natl. Acad. Sci. U.S.A.* 95 (12), S. 6578–6583.
- Wiezer, A. (2001): Untersuchung möglicher Kompetenzgene in *Thermus thermophilus*. Diplomarbeit. Georg-August-Universität zu Göttingen, Institut für Mikrobiologie und Genetik.
- Williamson, L. L.; Borlee, B. R.; Schloss, P. D.; Guan, C.; Allen, H. K.; Handelsman, J. (2005): Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. In: *Appl. Environ. Microbiol.* 71 (10), S. 6335–6344.
- Wooley, J. C.; Godzik, A.; Friedberg, I. (2010): A primer on metagenomics. In: *PLoS Comput. Biol.* 6 (2), S. e1000667.
- Yokoyama, S.; Hirota, H.; Kigawa, T.; Yabuki, T.; Shirouzu, M.; Terada, T. *et al.* (2000): Structural genomics projects in Japan. In: *Nat. Struct. Biol.* 7, S. 943–945.
- Zengler, K.; Toledo, G.; Rappe, M.; Elkins, J.; Mathur, E. J.; Short, J. M.; Keller, M. (2002): Cultivating the uncultured. In: *Proc. Natl. Acad. Sci. U.S.A.* 99 (24), S. 15681–15686.
- Zhou, Y.; Asahara, H.; Gaucher, E. A.; Chong, S. (2012): Reconstitution of translation from *Thermus thermophilus* reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. In: *Nucleic Acids Res.* 40 (16), S. 7932–7945.
- Zhu, Y.; Li, J.; Cai, H.; Ni, H.; Xiao, A.; Hou, L. (2013): Characterization of a new and thermostable esterase from a metagenomic library. In: *Microbiol. Res.* 168 (9), S. 589–597.

Zumft, W. G. (1997): Cell biology and molecular basis of denitrification. In: *Microbiol. Mol. Biol. Rev.* 61 (4), S. 533–616.

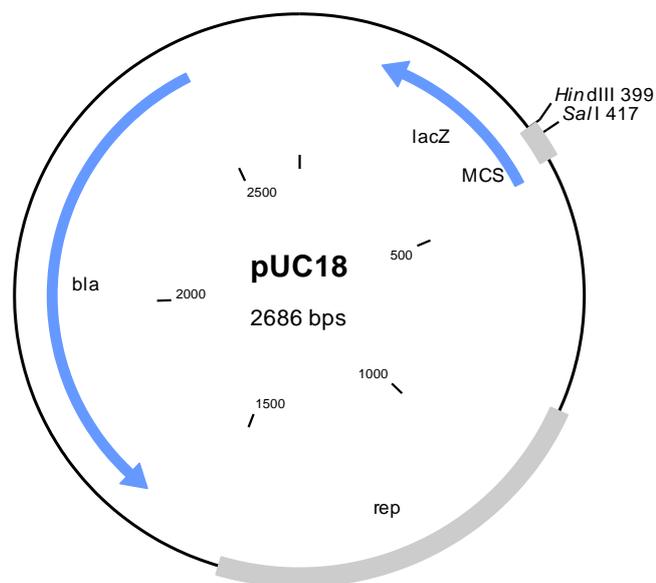
## IX. Anhang

### Plasmidkarten:

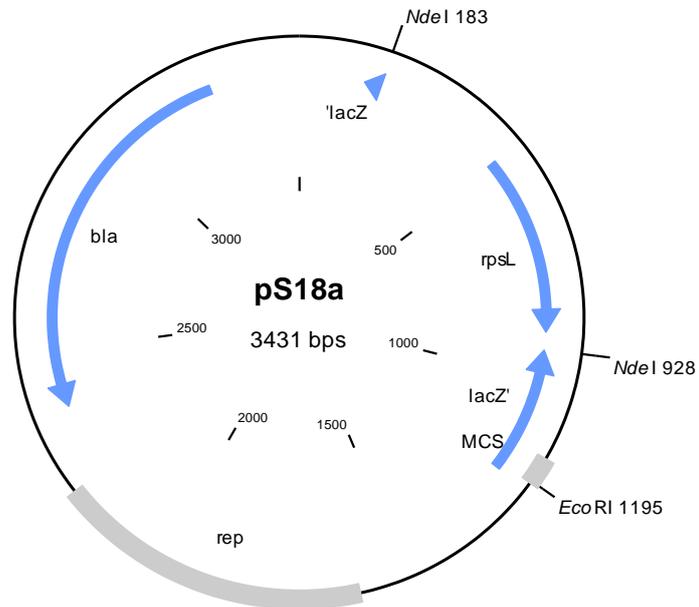
pBKat (A. Wiezer, Diplomarbeit, 2001):



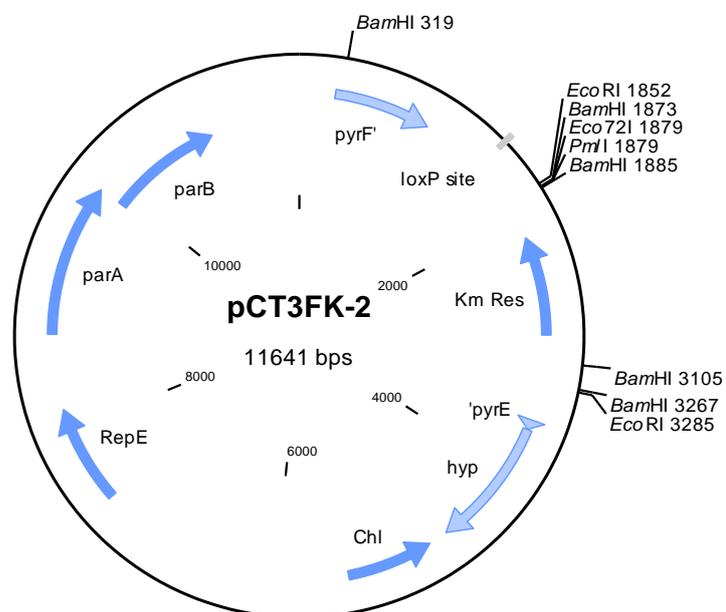
pUC18 (Fermentas, Vilnius, Litauen):



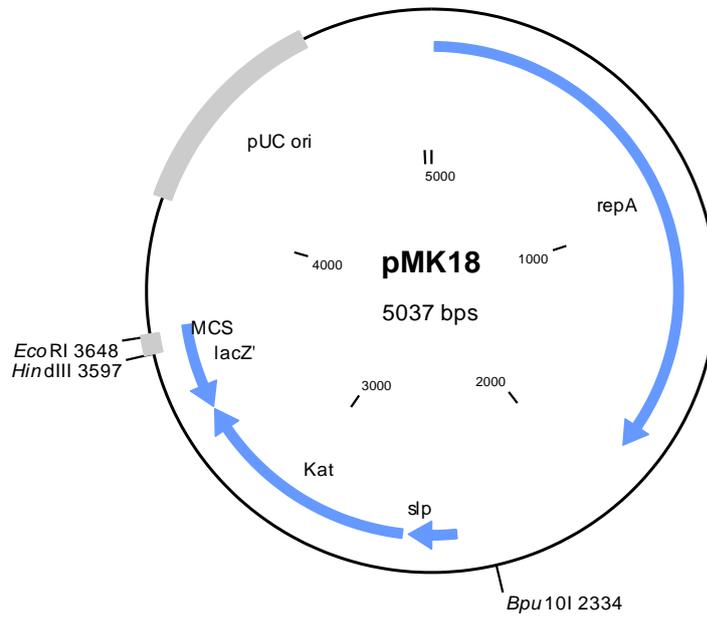
pS18a (Blas-Galindo *et al.*, 2007):



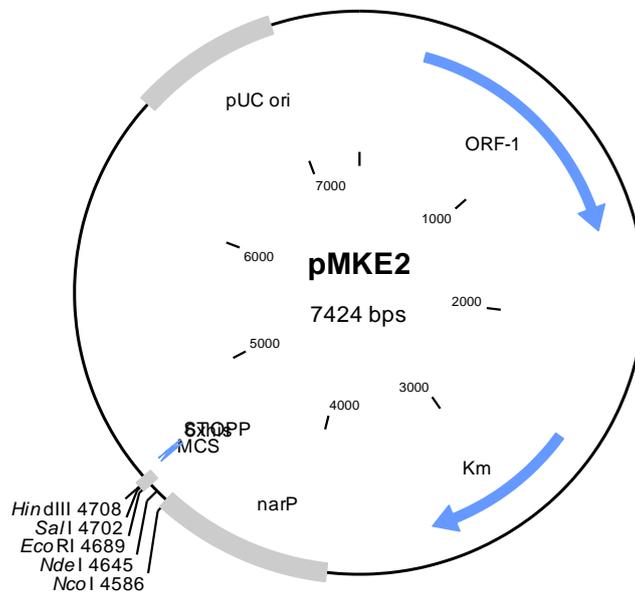
pCT3FK-2 (invertierte Sequenz nach Angelov *et al.*, 2009; modifiziert nach pCC1FOS, Epicentre):



pMK18 (de Grado *et al.*, 1999):



pMKE2 (Moreno *et al.*, 2005):



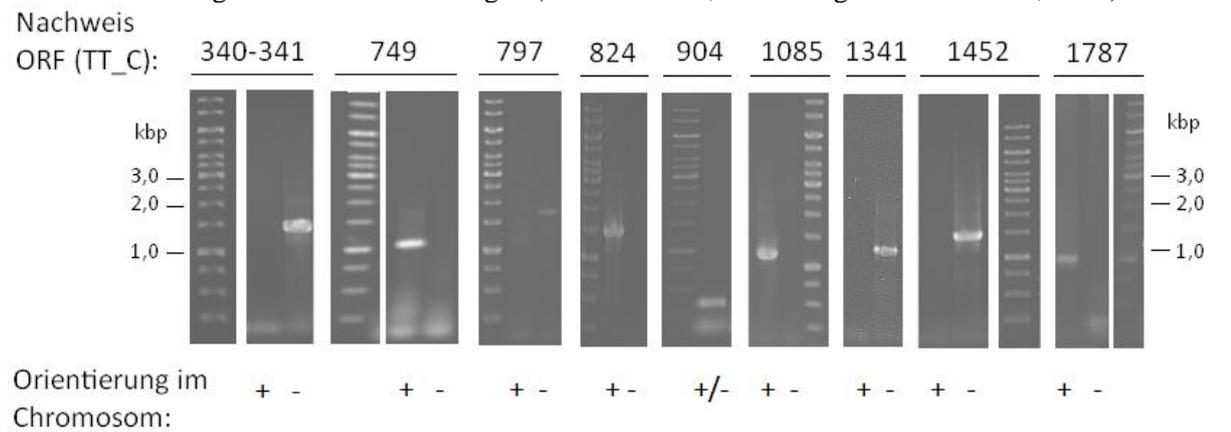
**Tabelle 39: Klonierung der *Suicide*-Plasmide zur Inaktivierung der lipolytischen Aktivität in *T. thermophilus* HB27. Tabelle mit Daten von D. Lehmann und C. Zörgiebel (Praktikumsreport, 2010).**

Zielsequenz (TT_C)	PCR Produktgröße, Annealing	Klonierungsstrategie	For-Primer	Rev-Primer
340-1	3,4 kb, 61 °C 6,92 kb, 60 °C	in pCR-TOPO, <i>site-directed mutagenesis</i>	Δ340_for Δ340_mut_for	Δ341_rev Δ341_mut_for
749	0,72 kb, 61 °C	in pBKat über <i>SacI</i>	Δ749_for	Δ749_rev
797	0,51 kb, 66 °C	in pBKat über <i>SacII</i>	Δ797_for	Δ797_rev
824	0,56 kb, 64 °C	in pBKat über <i>NotI</i>	Δ824_for	Δ824_rev
904	0,52 kb, 60 °C	in pBKat über <i>SacI</i>	Δ904_for	Δ904_rev
1085	0,54 kb, 64 °C	in pBKat über <i>NotI</i>	Δ1085_for	Δ1085_rev
1341	0,55 kb; 61 °C	in pBKat über <i>SacII</i>	Δ1341_for	Δ1341_rev
1452	0,54 kb, 61 °C	in pBKat über <i>NotI</i>	Δ1452_for	Δ1452_rev
1787	0,72 kb, 60 °C	in pBKat über <i>SacI</i>	Δ1787_for	Δ1787_rev

**Tabelle 40: PCR-Nachweise der erfolgten Genunterbrechungen im *T. thermophilus* Chromosom. Tabelle mit Daten von D. Lehmann und C. Zörgiebel (Praktikumsreport, 2010).**

Insertionsort (TT_C)	Lage des Vektors/Kat-Kassette relativ zur Zielsequenz im Genom (+ / -)	Produktgröße (kb) (+ / -)	For-Primer	Rev-Primer
340-1	Allelaustausch mit Kat-Kassette, antiparallel, -	1,47 (+) 1,43 (-)	Δ340_for Δ341_rev	pMK18_NdeI_rev
749	parallel, +	1,01 (+) 0,87 (-)	FP-FOS	tth749_rev tth749_for
797	antiparallel, -	2,00 (+) 1,76 (-)	tth797_for tth797_rev	pMK18_NdeI_rev
824	parallel, +	1,38 (+) 2,06 (-)	tth824_for tth824_rev	pMK18_NdeI_rev
904	parallel oder antiparallel, + oder -	0,35 (+/-)	pMK18_NdeI_f or	pMK18_NdeI_rev
1085	parallel, +	1,19 (+) 2,16 (-)	tth1085_for tth1085_rev	pMK18_NdeI_rev
1341	antiparallel, -	1,46 (+) 1,34 (-)	tth1341_for tth1341_rev	pMK18_NdeI_rev
1452	antiparallel, -	2,47 (+) 1,48 (-)	tth1452_for tth1452_rev	pMK18_NdeI_rev
1787	parallel, +	0,96 (+) 1,57 (-)	FP-FOS	rtta1781R PCR Δ1787_for

Zusammenfassung der Genunterbrechungen (Gelkontrollen, Abbildung nach Leis *et al.*, 2014):



**Tabelle 41: Liste der klassifizierten Esterase/Lipase-Familien.**

Klassifizierungen gemäß Arpigny und Jaeger (1999); Hausmann und Jaeger (2010); Liste nach Chow <i>et al.</i> (2012).				
Familie	Sequenzidentifikationsnummer	Annotation	Herkunft	
I-1	gi 1752730 dbj BAA09135.1	lipase	<i>Pseudomonas</i>	<i>aeruginosa</i>
I-1	gi 3098280 gb AAC15585.1	lipase	<i>Pseudomonas</i>	<i>fluorescens</i>
I-1	gi 45513 emb CAA32193.1	unnamed protein product	<i>Pseudomonas</i>	<i>fragi</i>
I-1	gi 870829 emb CAA56780.1	lipase	<i>Acinetobacter</i>	<i>calcoaceticus</i>
I-2	gi 557867 gb AAA50466.1	lipase	<i>Burkholderia</i>	<i>cepacia</i>
I-2	gi 49206 emb CAA49812.1	lipase	<i>Burkholderia</i>	<i>glumae</i>
I-2	gi 75538862 sp O68551	triacylglycerol lipase	<i>Pseudomonas</i>	<i>luteola</i>
I-3	gi 216915 dbj BAA02012.1	triacylglycerol lipase	<i>Pseudomonas</i>	<i>fluorescens</i>
I-3	gi 216950 dbj BAA02519.1	lipase	<i>Serratia</i>	<i>marcescens</i>
I-4	gi 75449375 sp Q8RJP5	extracellular esterase precursor	<i>Bacillus</i>	<i>megaterium</i>
I-4	gi 1568338 emb CAA02196.1	lipase	<i>Bacillus</i>	<i>pumilus</i>
I-4	gi 143154 gb AAA22574.1	lipase	<i>Bacillus</i>	<i>subtilis</i>
I-5	gi 3015530 gb AAC12257.1	lipase	<i>Geobacillus</i>	<i>stearothermophilus</i>
I-5	gi 1321706 emb CAA64621.1	triacylglycerol lipase	<i>Geobacillus</i>	<i>thermocatenulatus</i>
I-5	gi 122688384 sp Q842J9	thermostable lipase	<i>Geobacillus</i>	<i>zalihae</i>
I-6	gi 81325158 sp Q9RGZ6	lipase	<i>Staphylococcus</i>	<i>haemolyticus</i>
I-6	gi 126334 sp P04635.1	triacylglycerol lipase	<i>Staphylococcus</i>	<i>hyicus</i>
I-7	gi 81761821 sp Q8NU60	predicted acetyltransferases and hydrolases with the alpha/beta fold	<i>Corynebacterium</i>	<i>glutamicum</i>
I-7	gi 1430921 emb CAA67627.1	triacylglycerol lipase	<i>Propionibacterium</i>	<i>acnes</i>
I-7	gi 2435400 gb AAB71210.1	lipase LipA	<i>Streptomyces</i>	<i>cinnamoneus</i>
I-8	gi 123634260 sp Q48AN1	hypothetical protein	<i>Colwellia</i>	<i>psychrerythraea</i>
I-8	gi 123532849 sp Q2SGZ8	putative esterase/lipase	<i>Hahella</i>	<i>chejuensis</i>
I-8	gi 123587578 sp Q3IF07	hypothetical protein	<i>Pseudoalteromonas</i>	<i>haloplanktis</i>
II	gi 118572649 sp P10480.3	phosphatidylcholine-sterol precursor	<i>Aeromonas</i>	<i>hydrophila</i>
II	gi 75388798 sp Q7X4K7	esterase EstE	<i>Xanthomonas</i>	<i>vesicatoria</i>
II	gi 48550 emb CAA47020.1	triacylglycerol lipase	<i>Photorhabdus</i>	<i>luminescens</i>
II	gi 2218156 gb AAB61674.1	lipase/esterase	<i>Pseudomonas</i>	<i>aeruginosa</i> PAO1
II	gi 2896133 gb AAC38796.1	outer membrane esterase	<i>Salmonella enterica</i>	<i>serovar Typhimurium</i>
III	gi 44525 emb CAA37220.1	unnamed protein product	<i>Moraxella</i>	<i>sp.</i>
III	gi 153517 gb AAB51445.1	triacylglycerol acylhydrolase	<i>Streptomyces</i>	<i>sp.</i>
IV	gi 81320830 sp Q7SIG1	hydrolase	<i>Alicyclobacillus</i>	<i>acidocaldarius</i>
IV	gi 695278 gb AAC41424.1	lipase-like enzyme	<i>Cupriavidus</i>	<i>necator</i>
IV	gi 296842 emb CAA37862.1	triacylglycerol lipase	<i>Moraxella</i>	<i>sp.</i>
IV	gi 2853612 gb AAC38151.1	lipase	<i>Pseudomonas</i>	<i>sp. B11-1</i>
V	gi 151443 gb AAA25933.1	PHA-depolymerase	<i>Pseudomonas</i>	<i>oleovorans</i>
V	gi 747876 emb CAA47949.1	triacylglycerol lipase	<i>Psychrobacter</i>	<i>immobilis</i>
V	gi 3342450 gb AAC67392.1	lipolytic enzyme	<i>Sulfolobus</i>	<i>acidocaldarius</i>
VI	gi 546789 gb AAB30793.1	serine esterase	<i>Arthrospira</i>	<i>platensis</i>
VI	gi 244501 gb AAC60403.1	esterase II	<i>Pseudomonas</i>	<i>fluorescens</i>
VI	gi 2073490 emb CAA72452.1	esterase	<i>Rickettsia</i>	<i>prowazekii str. Madrid E</i>
VI	gi 123585942 sp Q3BXV6	carboxylesterase	<i>Xanthomonas</i>	<i>campestris pv. vesicatoria</i>
VII	gi 400739 sp Q01470.1	phenylcarbamate hydrolase	<i>Arthrobacter</i>	<i>oxydans</i>
VII	gi 68845777 sp P37967.2	para-nitrobenzyl esterase	<i>Bacillus</i>	<i>subtilis</i>
VII	gi 4160323 emb CAA22794.1	putative carboxylesterase	<i>Streptomyces</i>	<i>coelicolor A3(2)</i>
VII	gi 348008 gb AAA99492.1	carboxylic ester hydrolase	<i>Arthrobacter</i>	<i>globiformis</i>
VIII	gi 14019249 gb AAC60471.2	esterase III	<i>Pseudomonas</i>	<i>fluorescens</i>

VIII	gi 3649751 emb CAA78842.1	esterase A	<i>Streptomyces</i>	<i>anulatus</i>
LipS	gi 315603833 gb ADU46368.1	acylglycerol lipase	<i>Rhodopseudomonas</i>	<i>palustris</i> DX-1
LipS	gi 51856872 dbj BAD41030.1	esterase	<i>Symbiobacterium</i>	<i>thermophilum</i> IAM 14863
LipS	gi 317394016 gb EFV74766.1	esterase	<i>Bacillus</i>	<i>sp.</i> 2 A
LipS	gi 89088253 gb EAR67363.1	esterase	<i>Bacillus</i>	<i>sp.</i> NRRL B-14911
LipS	gi 345447308 gb AEN92268.1	EstGtA2	<i>Geobacillus</i>	<i>thermodenitrificans</i>
LipS	gi 39651563 emb CAE30086.1	putative carboxylesterase	<i>Rhodopseudomonas</i>	<i>palustris</i> CGA009
LipT	gi 218243229 gb EED09760.1	putative esterase	<i>Thermus</i>	<i>aquaticus</i> Y51MC23
LipT	gi 320150044 gb ADW21422.1	putative esterase	<i>Thermus</i>	<i>scotoductus</i> SA-01
LipT	gi 359290424 gb AEV15941.1	hypothetical protein	<i>Thermus</i>	<i>sp.</i> CCB US3
LipT	gi 46196833 gb AAS81248.1	putative esterase	<i>Thermus</i>	<i>thermophilus</i> HB27
UF	gi 9454060 gb AAF87663.1	esterase1	uncultured	bacterium
UF	gi 166163631 gb ABY83635.1	esterase	uncultured	bacterium
UF	gi 292486093 gb ADE28719.1	EstGK1	uncultured	bacterium
UF	gi 292486095 gb ADE28720.1	EstZ3	uncultured	bacterium
UF	gi 307073935 gb ADN26553.1	EstD2	uncultured	bacterium
UF	gi 194739309 gb ACF94292.1	esterase	uncultured	bacterium
UF	gi 223558101 gb ACM91105.1	lipase	uncultured	bacterium
UF	gi 9454064 gb AAF87665.1	esterase	uncultured	bacterium
UF	gi 343482756 gb AEM45123.1	hypothetical protein	uncultured	organism
UF	gi 9454058 gb AAF87662.1	lipase	uncultured	bacterium
UF	gi 91984610 gb ABE69172.1	probable lipase	uncultured	bacterium

Tabelle 42: Verwendete Oligonukleotide und PCR-Primer.

Primernamen 5' → 3' Sequenz <sup>1</sup>	Verwendungszweck
Δ340_for: GAGGATGGCCACCTTCTTCG Δ341_rev: AACGCGGTCCAGACCCGATTCTC	Flankiert bicistronisches Operon TT_C0340-1, Klonierung in pCR-TOPO, Detektion der Gendeletion, Erstellung der Southern Blot Sonde
Δ340_mut_for: TCCTCAAGGAGGTAGATCTATGAGGTTCCGCGT Δ341_mut_for: CCCTCGAGGAGATCTTAAGCCTGGCCA	Site-directed Mutagenesis mit <i>Bgl</i> III, Erstellung der Southern Blot Sonde
Δ749_for: ATCTGAGCTCTAGGACGAAGGAGCGCCACCT Δ749_rev: TGATGAGCTCCGGTAGGCGTCTTTCCCCCGG	Genunterbrechung von TT_C0749, kloniert via <i>Sac</i> I in pBKat
tth749_for: ACGATGCCGACCTTCATGGG tth749_rev: TTCCAGGTGACCCAGTTGTG	Detektion der Genunterbrechung von TT_C0749
Δ797_for: ACTACCGCGGTAACGCGGCGTGGCCTTGGCCTT Δ797_rev: ACTACCGCGGTTACAGGAGGCCGCCCATGA	Genunterbrechung von TT_C0797, kloniert via <i>Sac</i> II in pBKat
tth797_for: CCAGGGACCTTCTGGCCTAC tth797_rev: TGAGGGCTATGGCCTCTTC	Detektion der Genunterbrechung von TT_C0797
Δ824_for: ACTAGCGGCCGCTAGAGGAAGGCGATCGCCGGGCTT Δ824_rev: ACTAGCGGCCGCTAGCCAGGTCATGTTGGTGCTCCCC	Genunterbrechung von TT_C0824, kloniert via <i>Not</i> I in pBKat
tth824_for: GCGGACTTCTCAAGACCTC tth824_rev: CTCCTCCACCACCTCCTTTG	Detektion der Genunterbrechung von TT_C0824
Δ904_for: ACTAGAGCTCCAGCGGCTTATCGCGTGGTC Δ904_rev: TGATGAGCTCCCAAAGCACCGCCCCGTAGGC	Genunterbrechung von TT_C0904, kloniert via <i>Sac</i> I in pBKat
Δ904_Sall_for: ATCAGTCGACCACGGAGAACGACCGCCGTTAC	Flankiert ORF TT_C0904 mit <i>Sall</i> Schnittstelle, Klonierung in pUC18, Detektion der Gendeletion
Δ904_HindIII_rev: ACTAAAGCTTGATCCTGGAGGAGTCCACGGACGG	Flankiert TT_C0904 mit <i>Hind</i> III Schnittstelle, Klonierung in pUC18, Detektion der Gendeletion
rtta900F: CCCAGGGCATTTACAACC orf900F: ACTAGAATTCGAGGACACCTTCCGCCTCAAGCGCATCAAG	Detektion der Genunterbrechung von TT_C0904
Δ904_SOE_for: GGACGTGGAGGGTACAATGATCAGGCCGAAGCCCCGGAACCACTC Δ904_SOE_rev: GGTCCGGGGCTTCGGCCTGATCATTGTACCCTCCACGTCCAAAC	Sequence Overlap Extension PCR mit <i>Bcl</i> II Schnittstelle
Δ1085_for: ACTAGCGGCCGCTAGGCCCTTCCGGGGCAAGCGGAA Δ1085_rev: ACTAGCGGCCGCTAGCCACCCAGGCCGCTTGGC	Genunterbrechung von TT_C1085, kloniert via <i>Not</i> I in pBKat
tth1085_for: GACAGGCTTCTGGTGGAGAC tth1085_rev: CCGCGTAGAAGACCTGGTAG	Detektion der Genunterbrechung von TT_C1085
Δ1341_for: ACTACCGCGGTAGCTCCTCCACGGCTTCACTTCC Δ1341_rev: ACTACCGCGGCTACTTACCACCCGGTCTTTCC	Genunterbrechung von TT_C1341, kloniert via <i>Sac</i> II in pBKat
tth1341_for: TCGGGCTCAAAGACGAAGGG tth1341_rev: GATGGCAGGCCTTCCATAC	Detektion der Genunterbrechung von TT_C1341
Δ1452_for: ACTAGCGGCCGCTAGAAATCCC GCCCCGCAACCAG Δ1452_rev: ACTAGCGGCCGCTAGCGGATGGGCACCTTGAACCTC	Genunterbrechung von TT_C1452, kloniert via <i>Not</i> I in pBKat
tth1452_for: CTACCGCTGGCCTACTTTG tth1452_rev: ACGTGAAGCTCCTCCAGAAG	Detektion der Genunterbrechung von TT_C1452
Δ1787_for: ACTAGAGCTCATCTCCGCTTCGTGAGGGACTAC Δ1787_rev: TGATGAGCTCAGGTGAAGAGGTAGGCGTAGG	Genunterbrechung von TT_C1787, kloniert via <i>Sac</i> I in pBKat
Δ1787_Sall_for: ATCAGTCGACGTGGAGGCCCGGGGCGGGATG	Flankiert ORF TT_C1787 mit <i>Sall</i> Schnittstelle, Klonierung in pUC18
Δ1787_HindIII_rev: ACTAAAGCTTGGGAAAGGCCCGCTCGTGGAGAGC	Flankiert ORF TT_C1787 mit <i>Hind</i> III Schnittstelle, Klonierung in pUC18, Detektion der Gendeletion und Herstellung der Southern Blot Sonde
Δ1787_SOE_for: GGCCATCTTATACTTAGGGCTGATCAGGGCATGAGGGCCCGCG Δ1787_SOE_rev: CGGGCCCTCATGCCCTGATCAGCCCTAAGTATAAGATGGCCCCAT	Sequence Overlap Extension PCR mit <i>Bcl</i> II Restriktionsstelle, Erstellung der Southern Blot Sonde
tth1787_for: ATACTTAGGGCTGATCAGGGCA tth1787_rev: TTCCAGGAGACCCGCACGAAC	Detektion der markerlosen Δ1787 Mutanten aus einem Gemisch von Wildtyp-Zellen
PCR_Δ1787_for: GGGCCACCTCCTTGACCTTC	ΔTT_C1787 <i>Knockout</i> -Nachweis in <i>Thermus thermophilus</i>
rtta1781R: TGCCTCCTTTGGGCTTCTTGG	Detektion der Genunterbrechung in TT_C1787
pMK18-Kat-f: CCCGACTGGAAAGCGGGCAGTGAGCGCAA pMK18-Kat-r: GGGGCTCCTGGCCGTCCACCGC	Klonierung des pBKat <i>Suicide</i> -Plasmids (A. Wiezer, Diplomarbeit)
d42_seq_for: TACCTTCCCGAGGACATCAC d42_seq_rev: GCCGTCTACATCTTCTCAC	Nachweis der Gendeletion von TT_C0042
HACMC-F: TTTAAGAAGGAGATATAACAATGCAGGAGACAACAGTGCTGGA	Klonierung von Cel12E ohne Signalpeptid
SP-HACMC-F: TTTAAGAAGGAGATATAACAATGAAAAGCATTGCACTTG	Klonierung von Cel12E mit Signalpeptid



**Tabelle 43 : Konstruktion der Marker-freien Deletionsvektoren und Mutantendetektion.**

Erstellung der Marker-freien Multideletionsmutante *T. thermophilus* BL03 sowie die Nachweise aus Detektions-PCR und Southern Blot-Analyse (Fragmentlängen als kalkulierte Werte angegeben). Die erste Mutante  $\Delta bgl$  wurde vollständig von H. Li (unveröffentlichte Ergebnisse) erstellt.

Konstruktion des Deletionsvektors				PCR-basierte Detektion der markerlosen Gendeletionen				Southern Blot Nachweis der Gendeletionen					
ORF (Stamm)	Vektor	PCR Primer	Methode zur Klonierung	Resultier. Vektor	PCR Primer	Annealing-temperatur °C	WT Produkt Größe (kb)	BL03 Produkt Größe	Primer-Name	Länge der Sonde (kb)	Verdau der gDNA	SB WT (kb)	SB BL03 (kb)
TT_P0042 ( $\Delta bgl$ )	pCR-TOPO	42.F 42.R	site directed mutagenesis	pCR- $\Delta 42$ (6,87 kb)	d42_seq_for d42_seq_rev	60	2,11	0,83	42SB.F 42SB.R	0,66	<i>BamHI</i>	1,86	1,21
TT_C1787 (BL01)	pUC18	$\Delta 1787\_Sall\_for$ $\Delta 1787\_HindIII\_rev$	splicing by overlap extension PCR	pUC- $\Delta 1787$ (4,68 kb)	PCR $\Delta 1787\_for$ $\Delta 1787\_HindIII\_rev$	65	2,79	1,24	$\Delta 1787\_SOE\_for$ $\Delta 1787\_HindIII\_rev$	1,04	<i>SacII</i>	2,55	1,81
TT_C0340-341 (BL02)	pCR-TOPO	$\Delta 340\_for$ $\Delta 341\_rev$	site directed mutagenesis	pCR- $\Delta 340-1$ (5,44 kb)	$\Delta 340\_for$ $\Delta 341\_rev$	70	3,4	1,92	$\Delta 340\_for$ $\Delta 340\_mut\_for$	0,96	<i>SacII</i>	4,80	3,33
TT_C0904 (BL03)	pUC18	$\Delta 904\_Sall\_for$ $\Delta 904\_HindIII\_rev$	splicing by overlap extension PCR	pUC- $\Delta 904$ (4,68 kb)	$\Delta 904\_Sall\_for$ $\Delta 904\_HindIII\_rev$	70	3,01	2,03	Fragment von pUC18- $\Delta 904$ mittels Verdau mit <i>XhoI</i> und <i>HindIII</i> geschnitten und aufgereinigt	0,83	<i>HindIII</i>	5,47	4,49

Tabelle 44: Assemblierungsstatistik der 454-Sequenzierung *EST6*.

Fosmid	AZ3-14-D11	AZ3-14-D2	M12-4-G10	AZ3-33-C12	AZ3-32-E5
Zahl überlappende <i>reads</i> (%)	12006 (95.55%)	12912 (96.97%)	12611 (97.26%)	17308 (95.83%)	31312 (98.07%)
Zahl überlappende Basen (%)	5735887 (96.80%)	6255430 (98.43%)	6033363 (99.09%)	8225536 (96.98%)	15347368 (99.50%)
<i>read</i> Error (%)	27697 (0.48%)	30918 (0.49%)	26038 (0.43%)	43525 (0.53%)	73898 (0.48%)
Assemblierte <i>reads</i>	11000	11909	12518	16284	30804
Partielle <i>reads</i>	1006	1003	93	1024	508
singletons	317	122	94	428	76
repeats	0	0	0	0	0
outlier	11	20	2	26	40
Zu kurze <i>reads</i>	231	262	259	300	500
Anzahl <i>contigs</i>	2	4	1	1	3
Anzahl Basen	32684	34714	50871	39462	28967
Durchschnittliche <i>contig</i> -Größe (bp)	16342	8678	50871	39462	9655
Größtes <i>contig</i> (bp)	21047	12694	50871	39462	16376

(lt. Newbler-Auswertung, LGC-Genomics, Berlin)

Tabelle 45: Sequenzanalyse des Fosmids HA-cmc-1 (contig 4).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01	Hypothetisches Protein	<i>Thermococcus sp.</i> CL1	89 % (0,0)	Radical SAM
02c	Hypothetisches Protein	<i>Thermococcus sp.</i> AM4	86 % (4 × e-41)	Plasmidstabilisierungsprotein
03c	Hypothetisches Protein	<i>Thermococcus sp.</i> AM4	89 % (2 × e-13)	-
04	Hypothetisches Protein	-	-	-
05c	Multiple Substrat Aminotransferase	<i>Thermococcus sp.</i> 4557	87 % (0,0)	Aminotransferase Klasse I und II
06	Ribose-1,5-bisphosphat-isomerase	<i>Thermococcus sp.</i> AM4	92 % (0,0)	Initiationsfaktor II Untereinheit
07	C/D box Methylierungs guide Ribonukleoproteinkomplex, aNOP56 Untereinheit	<i>Thermococcus sp.</i> 4557	86 % (0,0)	NOSIC (NUC001) Domäne, putative snoRNA Bindedomäne
08	Fibrillarin	<i>Thermococcus kodakarensis</i> KOD1	94 % (3 × e-154)	Fibrillarin
09	Hydroxymethylglutaryl-CoA synthase	<i>Thermococcus sp.</i> AM4	94 % (0,0)	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III, C-Terminus
10	Acetyl-CoA Acetyltransferase	<i>Thermococcus sp.</i> 4557	93 % (0,0)	Thiolase, N- und C-termini
11	AcaC	<i>Thermococcus zilligii</i>	90 % (2 × e-62)	Unbekannte Funktion, DUF35 OB
12	Hypothetisches Protein	<i>Thermococcus gammatolerans</i> EJ3	59 % (1 × e-11)	-
13	Nukleotidbindeprotein	<i>Thermococcus gammatolerans</i> EJ3	58 % (2 × e-32)	PIN Domäne
14	MT Familie Permease	<i>Pyrococcus sp.</i> ST04	55 % (2 × e-92)	EamA-ähnlicher Transporter
15c	Hypothetisches Protein	<i>Thermococcus sp.</i> AM4	79 % (4 × e-36)	-
16c	N-acetyltransferase	<i>Thermococcus sp.</i> 4557	79 % (1 × e-87)	Acetyltransferase (GNAT)
17	Hypothetisches Protein	-	-	-
18	Putativer Transkriptionsfaktor	<i>Thermococcus sp.</i> CL1	89 % (5 × e-105)	Helix-turn-helix (HTH)
19c	Hypothetisches Protein	-	-	-
20	Hypothetisches Protein	<i>Thermococcus sp.</i> AM4	95 % (0,0)	Unbekannte Funktion, DUF505
21	Transkriptionsfaktor	<i>Thermococcus sp.</i> 4557	98 % (8 × e-130)	Transcription factor TFIID, TATA-Bindeprotein
22	Histon-Deacetylase	<i>Thermococcus sp.</i> AM4	78 % (2 × e-127)	Histon-Deacetylase Domäne
23	Endo-1,4-beta-glucanase b	<i>Thermococcus sp.</i> AM4	45 % (2 × e-82)	Glykosidhydrolasefamilie (GHF) 12, Cellulose Bindedomäne (CBM) 2
24	Na <sup>+</sup> /H <sup>+</sup> Antiporter	<i>Thermococcus sp.</i> CL1	84 % (0,0)	Natrium/Protonen Austauscher
25c	Indolpyruvat Oxidoreduktase Untereinheit IorB	<i>Thermococcus sp.</i> AM4	95 % (2 × e-138)	Pyruvat Ferredoxin/Flavodoxin Oxidoreduktase (POR)
26c	Indolpyruvat Oxidoreduktase Untereinheit IorA	<i>Thermococcus sp.</i> AM4	92 % (0,0)	POR, Thiamin diPh-Bindung Thiaminpyrophosphat (TPP) Bindung, Eisen-Schwefel (4Fe-4S) Cluster
27c	Acetyl-CoA Synthetase II Alpha-Untereinheit	<i>Thermococcus onnurineus</i> NA1	90 % (0,0)	CoA Bindedomäne, Succinyl-CoA Ligase, Flavodoxin-ähnlich
28	Metall-abhängige Hydrolase	<i>Thermococcus sp.</i> 4557	89 % (6 × e-143)	Beta-Lactamase Superfamilie B3
29	Putativer Transkriptionsregulator	<i>Thermococcus sp.</i> 4557	68 % (2 × e-165)	Putativer Zellwandbinde-Repeat 2
30c	Nukleinsäure-Bindeprotein	<i>Thermococcus sp.</i> 4557	87 % (6 × e-22)	Zink-Finger (C2H2-Typ)
31c	Hypothetisches Protein	<i>Thermococcus sp.</i> 4557	65 % (2 × e-105)	-
32c	Membranprotein	<i>Thermococcus onnurineus</i> NA1	51 % (4 × e-57)	Unbekannte Funktion, DUF1129
33c	Hypothetisches Protein	<i>Thermococcus kodakarensis</i> KOD1	47 % (5 × e-74)	Uncharakterisierte Proteinfamilie, UPF0104
34c	Spermidinsynthase	<i>Thermococcus onnurineus</i> NA1	86 % (3 × e-179)	Spermin/Spermidin Synthase
35c	Pyruvoyl-abhängige Arginin-Decarboxylase	<i>Thermococcus sp.</i> 4557	94 % (8 × e-104)	Pyruvoyl-abhängige Arginindecaboxylase (PvlArgDC)
36c	Natrium/Hydrogen Antiporter, NapA-Typ	<i>Thermococcus onnurineus</i> NA1	87 % (0,0)	Natrium/Hydrogen Austauscher
37	Hypothetisches Protein	<i>Thermococcus onnurineus</i> NA1	85 % (6 × e-63)	ACT Domäne
38c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit D	<i>Thermococcus onnurineus</i> NA1	69 % (0,0)	NADH-Ubiquinon/Plastoquinon (Komplex I), verschiedene Ketten
39c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit C	<i>Thermococcus onnurineus</i> NA1	92 % (1 × e-45)	NADH-Ubiquinon/Plastoquinon Oxidoreduktase, Kette 4L

40c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit B	<i>Thermococcus onnurineus</i> NA1	92 % (9 × e-73)	Zu MnhB Untereinheit des Na <sup>+</sup> /H <sup>+</sup> Antiporters zugehörige Domäne
41c	Multiuntereinheit Natrium/Hydrogen Antiporter Untereinheit	<i>Thermococcus onnurineus</i> NA1	81 % (2 × e-32)	Zu MnhB Untereinheit des Na <sup>+</sup> /H <sup>+</sup> Antiporters zugehörige Domäne
42c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit G	<i>Thermococcus onnurineus</i> NA1	92 % (3 × e-62)	Na <sup>+</sup> /H <sup>+</sup> Antiporter Untereinheit
43c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit F	<i>Thermococcus onnurineus</i> NA1	90 % (3 × e-44)	Multiple Resistenz- und pH Regulationsprotein F (MrpF / PhaF)
44c	Putativer monovalenter Kationen/H <sup>+</sup> Antiporter, Untereinheit E	<i>Thermococcus onnurineus</i> NA1	84 % (2 × e-122)	Na <sup>+</sup> /H <sup>+</sup> Ionen Antiporter Untereinheit
45	Glycin-Spaltungssystem Protein H	<i>Thermococcus kodakarensis</i> KOD1	93 % (1 × e-82)	Glycin-Spaltungssystem Protein H
46	Hypothetisches Protein	<i>Thermococcus zilligii</i>	91 % (5 × e-63)	-
47	Putative ATPase (AAA+ Superfamilie) 3	<i>Thermococcus sp.</i> CL1	86 % (0,0)	Archaeelle ATPase
48c	Putative Oxidoreduktase	<i>Thermococcus sp.</i> CL1	97 % (1 × e-74)	Pyridin Nukleotid-Disulphid-Oxidoreduktase

**Tabelle 46: Sequenzanalyse des Fosmids ES-6-B5 (contig 8).**

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01	DJ-1/PfpI Proteinfamilie	marines gamma Proteobakterium HTCC2148	61 % (4 × e-142)	Putative Amidotransferase (DUF4066) Helix-turn-helix Domäne
02	Acetyl-CoA Acetyltransferase	marines gamma Proteobakterium HTCC2148	76 % (0,0)	Thiolase, N- und C-terminus
03	3,4-Dihydroxy-2-butanon 4-phosphat Synthase	marines gamma Proteobakterium HTCC2148	84 % (0,0)	3,4-Dihydroxy-2-butanon 4-phosphat Synthase GTP Cyclohydrolase II
04	RND Familie Efflux Transporter, MFP Untereinheit	<i>Caulobacter vibrioides</i>	46 % (5 × e-85)	HlyD Sekretionsprotein
05	Multidrug Transporter AcrB	<i>Comamonas sp.</i> B-9	59 % (0,0)	AcrB/AcrD/AcrF Familie Drug Transporter
06c	Hypothetisches Protein	marines gamma Proteobakterium HTCC2148	85 % (0,0)	Luciferase-ähnliche Monooxygenase
07	Molybdopterin Dinukleotid-Bindeprotein	<i>Haliangium ochraceum</i> DSM 14365	45 % (0,0)	Molybdopterin Oxidoreduktase, Dinukleotid-Bindedomäne, Fe <sub>4</sub> S <sub>4</sub> -Domäne
08c	Enoyl-CoA Hydratase	marines gamma Proteobakterium HTCC2148	72 % (1 × e-129)	Enoyl-CoA Hydratase/Isomerase
09c	Nitroreduktase	marines gamma Proteobakterium HTCC2148	76 % (2 × e-108)	Nitroreduktase
10c	Enoyl-CoA Hydratase	<i>Spongiibacter tropicus</i>	81 % (2 × e-161)	Enoyl-CoA Hydratase/Isomerase
11c	AMP-Bindeprotein	marines gamma Proteobakterium HTCC2148	81 % (0,0)	AMP-bindendes Enzym
12c	Acetyl-CoA Acetyltransferase	marines gamma Proteobakterium HTCC2148	92 % (0,0)	-
13c	Hypothetisches Protein	marines gamma Proteobakterium HTCC2148	85 % (0,0)	-
14c	Putatives Protein mit konservierter Domäne	marines gamma Proteobakterium HTCC2148	81 % (4 × e-90)	Rubredoxin-ähnliche zinc ribbon Domäne (DUF35)
15c	Putatives Protein mit konservierter Domäne	marines gamma Proteobakterium HTCC2148	82 % (8 × e-82)	DUF35 OB-Faltungsdomäne
16c	Hypothetisches Protein	-	-	-
17	Glyoxalase Familie Protein	marines gamma Proteobakterium HTCC2148	73 % (8 × e-157)	Glyoxalase/Bleomycin Resistenzprotein/Dioxygenase
18	Hypothetisches Protein	<i>Catelliglobosipora koreensis</i>	50 % (1 × e-106)	Glycosylhydrolase Familie 18 (GHF 18), insignifikante CBM_5_12
19c	Putatives Protein der Phosphotransferase Enzymfamilie	marines gamma Proteobakterium HTCC2148	62 % (5 × e-158)	Phosphotransferase
20c	Hypothetisches Protein	-	-	-
21c	NAD Kinase	<i>Thiocapsa marina</i>	46 % (4 × e-161)	Na <sup>+</sup> /P <sub>i</sub> -Cotransporter
22c	Hypothetisches Protein	<i>Mariniradius saccharolyticus</i>	58 % (5 × e-56)	Ionenkanal
23c	Putative Glykosidase	<i>Rhodococcus erythropolis</i> PR4	38 % (2 × e-76)	Cellulase, Glycosylhydrolase Familie 5 (GHF 5)

Tabelle 47: Sequenzanalyse des Fosmids TB-est-7 (contig 9).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01	Methyl-akzeptierendes Chemotaxis Sensor-Protein	<i>Rhodothermus marinus</i> DSM 4252	41 % ( $4 \times e^{-25}$ )	Methyl-akzeptierendes Chemotaxis Protein (MCP)
02c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	89 % ( $2 \times e^{-105}$ )	Domäne unbekannter Funktion (DUF 99)
03c	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	94 % ( $3 \times e^{-130}$ )	-
04	Aspartatkinase	<i>Rhodothermus marinus</i> DSM 4252	87 % (0,0)	Aminosäurekinasefamilie, ACT Domäne, Pyridoxal-abhängige Decarboxylase
05	Dihydrodipicolinat-Reductase	<i>Rhodothermus marinus</i> DSM 4252	89 % ( $2 \times e^{-164}$ )	Dihydrodipicolinat-Reductase, C- und N-termini
06	Dihydrodipicolinat-Synthase	<i>Rhodothermus marinus</i> DSM 4252	95 % (0,0)	Dihydrodipicolinat-Synthetase Familie
07	2,3,4,5-Tetrahydropyridin-2,6-Dicarboxylat N-Succinyltransferase	<i>R. marinus</i> SG0.5JP17-172	92 % (0,0)	Tetrahydrodipicolinat N-Succinyltransferase, N-terminal Bakterielle Transferase Hexapeptid
08	Phosphoenolpyruvat - Carboxylase	<i>R. marinus</i> SG0.5JP17-172	89 % (0,0)	Phosphoenolpyruvat-Carboxylase
09-11	Hypothetisches Protein	-	-	-
12c	Propionyl-CoA Carboxylase	<i>Rhodothermus marinus</i> DSM 4252	96 % (0,0)	Carboxyltransferase Domäne
13c	CsbD-ähnliches Protein	<i>Rhodothermus marinus</i> DSM 4252	88 % ( $1 \times e^{-28}$ )	-
14c	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	84 % ( $7 \times e^{-76}$ )	-
15	Protoporphyrinogen-Oxidase	<i>Rhodothermus marinus</i> DSM 4252	88 % (0,0)	Flavin-haltige Amin-Oxidoreduktase
16	Cystathionine Gamma-synthase	<i>Rhodothermus marinus</i> DSM 4252	92 % (0,0)	Cys/Met Metabolismus PLP-abhängiges Enzym
17	Uroporphyrin-III C-Methyltransferase	<i>Rhodothermus marinus</i> DSM 4252	86 % ( $4 \times e^{-162}$ )	Tetrapyrrol (Corrin/Prophyrin) Methylase
18	Sirohäm-Synthase	<i>Rhodothermus marinus</i> DSM 4252	86 % ( $1 \times e^{-142}$ )	Putative NAD(P)-Bindung
19	Rhodanase-Domänenhaltiges Protein	<i>Rhodothermus marinus</i> DSM 4252	95 % (0,0)	Rhodanase-ähnliche Domäne
20c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	94 % ( $3 \times e^{-172}$ )	-
21c	Transferase Hexapeptid Repeat-haltiges Protein	<i>Rhodothermus marinus</i> DSM 4252	94 % ( $2 \times e^{-126}$ )	Bakterielle Transferase, Hexapeptid (sechs repeats)
22c	Oxidoreduktase Domänenhaltiges Protein	<i>Rhodothermus marinus</i> DSM 4252	87 % (0,0)	Oxidoreduktase Familie, NAD-Bindung, Rossmann Faltung, Oxidoreduktase Familie alpha/beta Domäne
23c	Nukleotid Zucker Dehydrogenase	<i>Rhodothermus marinus</i> DSM 4252	94 % (0,0)	DP-Glucose/GDP-Mannose Dehydrogenase, Bindedomäne, UDP-Bindedomäne und zentrale Domäne
24	Hypothetisches Protein	-	-	-
25c	Hypothetisches Protein	-	-	-
26c	Hypothetisches Protein	-	-	-
27	Cystein-Synthase	<i>Rhodothermus marinus</i> DSM 4252	89 % (0,0)	Pyridoxal-Phosphat-abhängiges Enzym
28	Mov34/MPN/PAD-1 Familien-Protein	<i>R. marinus</i> SG0.5JP17-172	93 % ( $9 \times e^{-91}$ )	Prokaryotisches Homolog der JAB Domäne
29	Schwefel-Transferprotein ThiS	<i>Rhodothermus marinus</i> DSM 4252	93 % ( $5 \times e^{-59}$ )	ThiS Familie
30	UBA/THIF-Typ NAD/FAD Bindeprotein	<i>Rhodothermus marinus</i> DSM 4252	94 % (0,0)	ThiF Familie, MoeZ/MoeB Domäne, Rhodanase-ähnliche Domäne
31c	Thiol-Disulfid Oxidoreductase DCC	<i>R. marinus</i> SG0.5JP17-172	86 % ( $2 \times e^{-84}$ )	Domäne unbekannter Funktion (DUF 393)
32c	Patatin	<i>Rhodothermus marinus</i> DSM 4252	90 % (0,0)	Patatin-ähnliche Phospholipase
33	Alpha-Glucosidase	<i>Rhodothermus marinus</i> DSM 4252	79 % ( $2 \times e^{-79}$ )	Galactose Mutarotase-ähnlich

Tabelle 48: Sequenzanalyse des Fosmids TB-xyl-3 (contig 5).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01c	Schwefeltransferprotein ThiS	<i>Rhodothermus marinus</i> DSM 4252	74 % (2 × e-36)	ThiS Familie
02c	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	86 % (0,0)	MacB-ähnliches periplasmatische Core-Domäne, FtsX-ähnliche Permease
03	CMP/dCMP Deaminase Zink-bindendes Protein	<i>R. marinus</i> SG0.5JP17-172	70 % (2 × e-122)	Cytidin und Deoxycytidylat-Deaminase Zinkbindung
04	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	72 % (0,0)	-
05	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	74 % (0,0)	Lamin endständige Domäne, FlgD <sub>ig</sub> , FlgD Ig-ähnliche Domäne
06	Eisen-Schwefel-Cluster Assemblierungsapparat	<i>R. marinus</i> SG0.5JP17-172	99 % (6 × e-61)	Eisen-Schwefel-Cluster Biosynthese
07	ATPase AAA-2 Domänen- haltiges Protein	<i>R. marinus</i> SG0.5JP17-172	99 % (0,0)	Clp N- und C-terminale Domäne, ATPase Familie, UvrB/uvrC-Motiv
08c	N(4)-(beta-N- Acetylglucosaminy)-L- Asparaginase	<i>Rhodothermus marinus</i> DSM 4252	83 % (5 × e-177)	Asparaginase
09c	Glycosidhydrolase	<i>Rhodothermus marinus</i> DSM 4252	76 % (0,0)	Glycosylhydrolase Familie 18
10	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	86 % (1 × e-111)	AP2 Domäne
11c	Putatives uncharakterisiertes Protein	<i>Parabacteroides</i> sp. CAG:409	24 % (3 × e-07)	-
12c	Peptidase M16 Domänen- enthaltendes Protein	<i>Rhodothermus marinus</i> DSM 4252	92 % (0,0)	Insulinase (Peptidase Familie M16)
13c	Peptidase M16 Domänen- enthaltendes Protein	<i>Rhodothermus marinus</i> DSM 4252	91 % (0,0)	Insulinase (Peptidase Familie M16)
14c	DNA-Polymerase I	<i>Rhodothermus marinus</i> DSM 4252	92 % (0,0)	DNA-Polymerase Familie A, 5'-3' Exonuklease, 3'-5'-Exonuklease
15	dTDP-glucose 4,6- Dehydratase	<i>R. marinus</i> SG0.5JP17-172	95 % (0,0)	NAD-abhängige Epimerase/Dehydratase Familie
16	Nukleotidyltransferase	<i>Rhodothermus marinus</i> DSM 4252	98 % (5 × e-155)	Nukleotidyltransferase
17	dTDP-4-Dehydrothamnose 3 5-Epimerase-ähnliches Protein	<i>Rhodothermus marinus</i> DSM 4252	92 % (4 × e-102)	dTDP-4-Dehydrothamnose 3,5-Epimerase
18	Haloacid-Dehalogenase Domänen-haltiges Hydrolaseprotein	<i>R. marinus</i> SG0.5JP17-172	86 % (1 × e-143)	HAD-Hyrolase-ähnlich
19c	Fructosaminkinase	<i>Rhodothermus marinus</i> DSM 4252	90 % (2 × e-145)	Fructosaminkinase
20c	Protein Tyrosinphosphatase	<i>Rhodothermus marinus</i> DSM 4252	86 % (7 × e-103)	Phosphotyrosin Proteinphosphatase
21	Exodeoxyribonuklease 7 große Untereinheit	<i>R. marinus</i> SG0.5JP17-172	80 % (0,0)	Exonuklease VII, große Untereinheit OB-fold Nukleinsäure-Bindedomäne
22c	Beta-Lactamase	<i>Rhodothermus marinus</i> DSM 4252	90 % (0,0)	Beta-Lactamase
23	N-Acyl-D-Aminosäure Deacylase	<i>Rhodothermus marinus</i> DSM 4252	90 % (0,0)	Amidohydrolase, D-Aminoacylase
24c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	71 % (3 × e-12)	-
25	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	82 % (1 × e-137)	ATP-Binderegion
26	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	83 % (2 × e-76)	-
27c	Sauerstoff-unabhängige Coproporphyrinogen III Oxidase	<i>Rhodothermus marinus</i> DSM 4252	86 % (0,0)	Radical SAM Superfamilie HemN C-terminale Domäne
28	Hypothetisches Protein	-	-	-
29c	Gruppe 1 Glycosyl- transferase	<i>Rhodothermus marinus</i> DSM 4252	91 % (0,0)	Glycosyltransferase Familie 4 Glycosyltransferase Gruppe 1
30c	OsmC Proteinfamilie	<i>Rhodothermus marinus</i> DSM 4252	88 % (2 × e-85)	OsmC-ähnliches Protein
31	Peptidase C26	<i>Rhodothermus marinus</i> DSM 4252	76 % (1 × e-117)	Peptidase C26

Tabelle 49: Sequenzanalyse des Fosmids TB01-A10 (contig 1).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	70 % (3 × e-102)	-
02c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	68% (6 × e-83)	-
03c	Typ II Sekretionspfad, ATPase PulE/Tfp Pilus Assemblierungspfad, ATPase PilB	<i>Rhodothermus marinus</i> DSM 4252	84 % (0,0)	Typ II/IV Sekretionssystem Protein
04c	Typ II Sekretionspfad, Komponente PulF / Typ IV fimbrisches Assemblierungsprotein PilC	<i>Rhodothermus marinus</i> DSM 4252	93 % (0,0)	Typ II Sekretionssystem (T2SS), Protein F
05c	Hypothetisches Protein	<i>Balneola vulgaris</i>	34 % (2 × e-13)	-
06c	TPR Wiederholung	<i>R. marinus</i> SG0.5JP17-172	86 % (0,0)	TPR_1, Tetratricopeptid
07	Klasse I Hitzeschockprotein	<i>Rhodothermus marinus</i> DSM 4252	95 % (7 × e-98)	Hsp20/Alpha Kristalline Familie
08	Hypothetische NUDIX Hydrolase YeaB	<i>Rhodothermus marinus</i> DSM 4252	83% (2 × e-96)	NUDIX Domäne
09c	Thiosulfat-Sulfurtransferase, Rhodanase (EC 2.8.1.1)	<i>R. marinus</i> SG0.5JP17-172	87 % (0,0)	Rhodanase
10	Hypothetisches Protein	-	-	-
11c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	93 % (1 × e-173)	Domäne unbekannter Funktion (DUF4249)
12c	TonB-abhängiger Rezeptor	<i>R. marinus</i> SG0.5JP17-172	90 % (0,0)	Carboxypeptidase Regulator-ähnliche Domäne; TonB-abhängige Rezeptor Anschluss-Domäne
13c	Glycerophosphoryl-Diester Phosphodiesterase (EC3.1.4.46)	<i>R. marinus</i> SG0.5JP17-172	87% (3 × e-167)	Glycerophosphoryldiester Phosphodiesterase Familie
14	Hypothetisches Protein	-	-	-
15	DNA-Gyrase Untereinheit A (EC 5.99.1.3)	<i>R. marinus</i> SG0.5JP17-172	95 % (0,0)	DNA-Gyrase/Topoisomerase IV, Untereinheit A; DNA-Gyrase C-terminale Domäne, beta-Propeller
16c	L-Lysine 6-Transaminase	<i>R. marinus</i> SG0.5JP17-172	94 % (0,0)	Aminotransferase Klasse III
17	Lipoatesynthase	<i>R. marinus</i> SG0.5JP17-172	97 % (0,0)	Radical SAM Superfamilie
18c	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	73 % (4 × e-27)	Antitoxin Phd_YefM, Typ II Toxin-Antitoxin System
19	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	84 % (5 × e-31)	Protein unbekannter Funktion (DUF2892)
20c	Hypothetisches Protein	<i>Rhodothermus marinus</i> DSM 4252	96 % (2 × e-63)	SCP-2 Steroltransfer-Familie
21c	L-Carnitindehydratase/Gallensäure induzierbares Protein F	<i>Rhodothermus marinus</i> DSM 4252	88 % (0,0)	CoA-Transferase Familie III
22c	putativer transcriptioneller Regulator	<i>Rhodothermus marinus</i> DSM 4252	60 % (3 × e-79)	Bakterielles Regulatorprotein, tetR Familie; YsiA-ähnliches Protein, C-terminale Region
23c	TonB-Abhängiger Rezeptor	<i>Rhodothermus marinus</i> DSM 4252	89 % (0,0)	Carboxypeptidase Regulations-ähnliche Domäne; TonB-abhängiger Rezeptor Anschlussdomäne; TonB-abhängiger Rezeptor
24c	Lang-kettige Fettsäure--CoA Ligase (EC 6.2.1.3)	<i>R. marinus</i> SG0.5JP17-172	93 % (0,0)	AMP-bindendes Enzym
25c	Kurz-kettige Dehydrogenase/Reductase SDR	<i>Rhodothermus marinus</i> DSM 4252	95 % (0,0)	Kurz-kettige Dehydrogenase
26	Enoyl-CoA Hydratase (EC 4.2.1.17)	<i>Rhodothermus marinus</i> DSM 4252	92 % (8 × e-172)	Enoyl-CoA Hydratase/Isomerase Familie
27	Aspartyl-tRNA(Asn) Amidotransferase Untereinheit A (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) Amidotransferase Untereinheit A (EC 6.3.5.7)	<i>R. marinus</i> SG0.5JP17-172	88 % (0,0)	Amidase
28	Carbonyldehydrogenase (NADPH)	<i>R. marinus</i> SG0.5JP17-172	88 % (7 × e-163)	Kurzkettige Dehydrogenase
29	Acyl-CoA Dehydrogenase-enthaltendes Protein	<i>Rhodothermus marinus</i> DSM 4252	92 % (0,0)	Acyl-CoA Dehydrogenase, N-terminale, C-terminale, Mitteldomäne
30	Acyl-CoA Dehydrogenase-enthaltendes Protein	<i>Rhodothermus marinus</i> DSM 4252	83 % (0,0)	Acyl-CoA Dehydrogenase, C-terminale und Mitteldomäne
31	Aminoglycosid Phosphotransferase	<i>Rhodothermus marinus</i> DSM 4252	89 % (0,0)	Phosphotransferase Enzymfamilie
32c	Prolyl-Endopeptidase (EC 3.4.21.26)	<i>R. marinus</i> SG0.5JP17-172	90 % (0,0)	Prolyl-Oligopeptidase Familie
33c	Hypothetisches Protein	-	-	-

34c	Hypothetisches Protein	<i>R. marinus</i> SG0.5JP17-172	77 % ( $2 \times e^{-54}$ )	-
35c	Molybdän Kofaktor Biosyntheseprotein MoaA	<i>Rhodothermus</i> <i>marinus</i> DSM 4252	92 % (0,0)	Radical SAM Superfamilie; Fe <sub>4</sub> -S <sub>4</sub> Einzelcluster-Domän; Molybdän Cofaktor-Synthese C
36c	GntR Familie transkriptioneller Regulator	<i>R. marinus</i> SG0.5JP17-172	83 % ( $3 \times e^{-156}$ )	Bakterielles Regulatorprotein, gntR Familie; UTRA Domäne
37	Nukleosiddiphosphatkinase (EC 2.7.4.6)	<i>Rhodothermus</i> <i>marinus</i> DSM 4252	99 % ( $4 \times e^{-94}$ )	Nukleosiddiphosphatkinase

Tabelle 50: Sequenzanalyse des Fosmids TN-9-B5 (contig 6).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01c	30S ribosomales Protein S16	<i>Opitutaceae</i> Bakterium TAV1	80 % ( $2 \times e^{-41}$ )	Ribosomales Protein S16
02c	tRNA (Guanine-N1)- Methyltransferase	Unklassifizierte <i>Opitutaceae</i>	82 % ( $3 \times e^{-133}$ )	tRNA (Guanine-1)-Methyltransferase
03c	50S ribosomales Protein L19	<i>Opitutus terrae</i> PB90-1	78 % ( $3 \times e^{-57}$ )	Ribosomales Protein L19
04c	Hypothetisches Protein	-	-	-
05c	Transketolase	<i>Opitutaceae</i> Bakterium TAV1	78 % (0,0)	Transketolase, Thiamindiphosphat und Pyrimidin-bindende Domäne, C-terminus
06c	-	-	-	-
07	MarR Familie transkriptioneller Regulator	<i>Stigmatella aurantiaca</i> DW4/3-1	51 % ( $9 \times e^{-18}$ )	MarR Familie
08	RND Familie Efflux- Transporter MFP Untereinheit	<i>Opitutus terrae</i> PB90-1	48 % ( $4 \times e^{-105}$ )	HlyD Familie Sekretionsprotein
09c	Acriflavin Resistenzprotein B	<i>Chthoniobacter flavus</i>	60 % (0,0)	AcrB/AcrD/AcrF Familie Transporter
10	Transporter	<i>Opitutaceae</i> Bakterium TAV1	57 % ( $9 \times e^{-148}$ )	Efflux-Protein der äußeren Membran
11c	Hypothetisches Protein	-	-	-
12c	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	29 % ( $2 \times e^{-75}$ )	-
13c	Hypothetisches Protein	-	-	-
14	ATPase-ähnliches Protein	<i>Opitutus terrae</i> PB90-1	53 % ( $2 \times e^{-111}$ )	AAA Domäne
15	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	65 % ( $2 \times e^{-75}$ )	-
16c	AMP-abhängige Synthetase	<i>Opitutaceae</i> Bakterium TAV1	66 % (0,0)	Acyltransferase, AMP-Bindung
17	FAD-gekoppelte Oxidase	<i>Opitutaceae</i> Bakterium TAV1	83 % (0,0)	FAD-gekoppelte Oxidase, FAD Bindedomäne
18c	Chromosomales Segregationsprotein SMC	<i>Opitutus terrae</i> PB90-1	70 % (0,0)	RecF/RecN/SMC N-terminale Domäne, SMC Proteine, Flexible Achsendomäne
19	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	31 % ( $9 \times e^{-65}$ )	-
20c	Phytochrom Sensorprotein	<i>Opitutus terrae</i> PB90-1	79 % (0,0)	Typ II/IV Sekretionssystem-Protein, GAF Domäne
21c	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	68 % ( $2 \times e^{-36}$ )	UNcharakterisierter Protein Familie UPF0102
22c	Glucose-6-Phosphat-1- Dehydrogenase	Nicht klassifizierte <i>Opitutaceae</i>	84 % (0,0)	Glucose-6-phosphate Dehydrogenase, NAD-bindende Domäne, C-terminus
23c	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	56 % ( $1 \times e^{-126}$ )	Glucose-6-phosphat-Dehydrogenase Untereinheit
24c	Hypothetisches Protein	<i>Verrucomicrobiae</i> Bakterium DG1235	51 % ( $2 \times e^{-75}$ )	Glucosamine-6-Phosphatisomerasen/6- Phosphogluconolactonase
25c	L-Fucoseisomerase-ähnliches Protein	<i>Thermoanaerobacterium</i> <i>saccharolyticum</i> JW/SL- YS485	41 % ( $8 \times e^{-106}$ )	L-Fucoseisomerase, C-terminale Domäne
26	Hypothetisches Protein	-	-	-
27	SAM-abhängige Methyltransferase	<i>Methylacidiphilum</i> <i>inferorum</i> V4	31 % ( $3 \times e^{-39}$ )	Methyltransferase-Domäne
28	Hypothetisches Protein	-	-	-
29	Hypothetisches Protein	<i>Geobacter</i> sp. M21	38 % ( $3 \times e^{-28}$ )	Protein unbekannter Funktion (DUF4254)
30	Flp Pilus Assemblierungsprotein Tadd	<i>Singulisphaera acidiphila</i> DSM 18658	48 % ( $2 \times e^{-111}$ )	TPR Wiederholung

Tabelle 51: Sequenzanalyse des Fosmids ES-7-E10 (contig 7).

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
01c	Cytochrom c Proteinfamilie	<i>Leptothrix cholodnii</i> SP-6	45 % (1 × e-60)	-
02c	Crp/Fnr Familie transkriptioneller Regulator	<i>Thermoanaerobacter</i>	34 % (5 × e-33)	Zyklische Nukleotid-Bindedomäne, Crp-ähnliche Helix-Turn-Helix
03c	Hypothetisches Protein	<i>Azoarcus sp. KH32C</i>	57 % (2 × e-126)	NHL Wiederholung
04	Cytochrom c Klasse I	<i>Leptothrix cholodnii</i> SP-6	43 % (9 × e-50)	Cytochrom C Oxidase, cbb3-Typ, Untereinheit III
05c	Zweikomponenten-Regulator, Propeller-Domänen enthaltendes Protein	<i>Leptothrix cholodnii</i> SP-6	66 % (1 × e-164)	Zweikomponenten-Regulator Propeller
06c	Zweikomponenten-Regulator, Propeller-Domänen enthaltendes Protein	<i>Leptothrix cholodnii</i> SP-6	64 % (1 × e-149)	Zweikomponenten-Regulator Propeller
07c	Putatives Cytochrom c55x	<i>Candidatus Nitrospira defluvii</i>	43 % (2 × e-04)	-
08c	Protein mit Sell-repeat	<i>Burkholderiales bacterium JOSHI_001</i>	42 % (2 × e-19)	Sell Wiederholung
09	Cytochrom c, Klasse I	<i>Azoarcus sp. KH32C</i>	49 % (1 × e-95)	Cytochrom C Oxidase, cbb3-Typ, Untereinheit III
10c	Hypothetisches Protein	-	-	-
11	Sensor Histidinkinase Response Regulator, HAMP und Hpt Domänen-haltig	<i>Geobacter sulfurreducens</i> PCA	39 % (7 × e-117)	Histidinkinase A (Phospho-Akzeptor), ATPase, Response Regulator Empfänger
12	LuxR Familie transkriptioneller Regulator	<i>Ktedonobacter racemifer</i>	44 % (4 × e-57)	Response Regulator Empfänger, Bakterieller Regulatorprotein der luxR Familie
13c	Zwei-Komponenten Hybridsensor und Regulator, partiell	<i>Cyanotheca sp. CCY0110</i>	34 % (5 × e-100)	PAS fold, Histidinkinase A (Phosphor-Akzeptor), ATPase, Response Regulator Empfänger, Hpt Domäne
14c	Protein D des generellen Sekretionspfads	Endosymbiont von <i>Riftia pachyptila</i>	49 % (0,0)	Bakterielles Typ II/III Sekretionssystem, kurze Domäne
15	Protein E des generellen Sekretionspfads	<i>Nitrosococcus halophilus</i> Nc 4	59 % (0,0)	Typ II/IV Sekretionssystem Protein
16	Protein F des generellen Sekretionspfads	<i>Nitrosococcus oceani</i> ATCC 19707	49 % (5 × e-130)	Typ II Sekretionssystem (T2SS), Protein F
17	Hypothetisches Protein	<i>Pseudomonas sp. EGD-AK9</i>	61 % (1 × e-55)	Typ II Sekretionssystem (T2SS), Protein G
18	Protein H des generellen Sekretionspfads	<i>Thiorhodococcus drewsii</i>	37 % (6 × e-15)	-
19	Hypothetisches Protein	<i>Methylobacter marinus</i>	43 % (1 × e-23)	-
20	Hypothetisches Protein	<i>Thiothrix flexilis</i>	31 % (4 × e-22)	Typ IV Pilin N-terminale Methylierungsstelle GFxxxE
21	Protein K des generellen Sekretionspfads	<i>Nitrosococcus halophilus</i> Nc 4	36 % (5 × e-47)	Typ II Sekretionssystem (T2SS), Protein K
22	fimbrische Assemblierungsprotein-Familie	<i>Nitrosococcus halophilus</i> Nc 4	33 % (6 × e-56)	fimbrisches Assemblierungsprotein (PilN)
23	Hypothetisches Protein	<i>Thioalkalivibrio sp. ALE11</i>	29 % (4 × e-16)	Typ II Sekretionssystem (T2SS), Protein M Subtyp b
24	Hypothetisches Protein	Endosymbiont von <i>Riftia pachyptila</i>	34 % (2 × e-11)	Typ IV Pilus-Biogenese
25	Hypothetisches Protein	<i>Azoarcus toluclasticus</i>	49 % (8 × e-51)	SCO1/SenC
26	Putatives Cytochrom c	<i>Pseudomonas pseudoalcaligenes</i>	39 % (5 × e-82)	Periplasmatisches Bindeprotein
27c	Putatives Cytochrom c	Planctomycet KSU-1	37 % (3 × e-106)	Periplasmatisches Bindeprotein, Cytochrom C Domäne
28c	Elektronentransportprotein SCO1/SenC	<i>Leptothrix cholodnii</i> SP-6	49 % (7 × e-58)	SCO1/SenC
29c	PPIC-Typ PPIASE Domänen-haltiges Protein	<i>Spirochaeta thermophila</i> DSM 6192	33 % (3 × e-43)	Rotamase PPIC-Typ PPIASE Domäne SurA N-terminale Domäne
30c	Hypothetisches Protein	<i>Cystobacter fuscus</i>	49 % (0,0)	Cytochrom D1 Hämdomäne

**Tabelle 52: Sequenzanalyse *contig 10* (MB-est-2/3/5 und 6)**

ORF-Position	Annotation	Höchste Übereinstimmung ( <i>blastp</i> )	Identität ( <i>e-value</i> )	Proteinklassifikation (Pfam), signifikante Domänen und Familien
1464...19	Exonuklease I	<i>Pseudoalteromonas haloplanktis</i>	89 % (0,0)	Exonuklease
2021...2572	Hypothetisches Protein	<i>Pseudoalteromonas haloplanktis</i>	44 % (2 × e-49)	-
2565...2945	Hypothetisches Protein	<i>Pseudoalteromonas haloplanktis</i>	36 % (8 × e-17)	-
3543...2959	Hypothetisches Protein	<i>Pseudoalteromonas haloplanktis</i>	36 % (8 × e-17)	-
4385...3462	Permease	<i>Pseudoalteromonas haloplanktis</i>	85 % (5 × e-166)	Domäne unbekannter Funktion (DUF897)
4513...5463	Transkriptioneller Regulator	<i>Pseudoalteromonas sp. SM9913</i>	80 % (0,0)	Bakterieller Regulator Helix-Turn-Helix Protein, LysR Substrat Bindung
6623...5481	Diguanylatcyclase	<i>Pseudoalteromonas haloplanktis</i>	64 % (1 × e-168)	GGDEF Domäne
7598...6705	Epimerase	<i>Pseudoalteromonas haloplanktis</i>	69 % (7 × e-150)	NAD-abhängige Epimerase/Dehydratasefamilie, Domäne unbekannter Funktion (DUF1731)
8342...9862	Sensorische Signalübertragungs-Histidinkinase	<i>Pseudoalteromonas agarivorans</i>	63 % (0,0)	PAS fold Domain, Histidinkinase A (Phospho-Akzeptor) Domäne, ATPase, Response Regulator Empfänger
9983...11251	Signalpeptid vorhergesagt	<i>Pseudoalteromonas undina</i>	77 % (0,0)	Bakterielles extrazelluläres Solut-Bindeprotein
11330...13549	Histidinkinase	<i>Pseudoalteromonas sp. BSi20439</i>		Histidinkinase A (Phospho-Akzeptor) Domäne, ATPase, Response Regulator Empfänger
15879...13651	Isocitratdehydrogenase	<i>Pseudoalteromonas agarivorans</i>	89 % (0,0)	Monomerische Isocitrat- Dehydrogenase
16722...17039	ATP-abhängiges Clp Protease ClpS	<i>Pseudoalteromonas</i>	62 % (7 × e-69)	ATP-abhängige Clp Protease Adaptorprotein ClpS
17111...19375	ATP-abhängiges Clp Protease ATP-bindendes Protein	<i>Pseudoalteromonas haloplanktis</i>	93 % (0,0)	Clp amino terminal domain, ClpB protein, ATPase
20562...19858	Arginyl-tRNA-Proteintransferase	<i>Pseudoalteromonas haloplanktis</i>	78 % (3 × e-176)	Arginine-tRNA-Proteintransferase, N und C terminus
21271...20555	Leucyl/Phenylalanyl-tRNA-Proteintransferase	<i>Pseudoalteromonas sp. PAMC 22718</i>	72 % (3 × e-128)	Leucyl/Phenylalanyl-tRNA Proteintransferase
22327...21377	Thioredoxinreduktase	<i>Pseudoalteromonas haloplanktis</i>	96 % (0,0)	Pyridinnukleotid-Disulphid Oxidoreduktase
22615...23940	Phosphatidylserinsynthase	<i>Pseudoalteromonas haloplanktis</i>	84 % (0,0)	Phospholipase D (PLD)-ähnliche Domäne
24773...24000	Oxoacyl-ACP-Reductase	<i>Pseudoalteromonas</i>	91 % (3 × e-169)	kurzkettige Dehydrogenase
24972...27404	Hypothetisches Protein	<i>Pseudoalteromonas sp. BSi20439</i>	71 % (0,0)	Bakterielle Virulenzfaktor Lipase N-terminal, Alpha/Beta Hydrolase 5
27604...28494	Peptidase	<i>Pseudoalteromonas haloplanktis</i>	82 % (8 × e-167)	Peptidase Familie S49

**Tabelle 53: Sequenzanalyse *contig 25* (MB-est- 7 und 8).**

ORF-Position	Annotation	Höchste Übereinstimmung ( <i>blastp</i> )	Identität ( <i>e-value</i> )	Proteinklassifikation (Pfam), signifikante Domänen und Familien
453...1	Thermostable Carboxylesterase Est50	<i>Geobacillus thermoleovorans</i> CCB_US3_UF5	97 % (3 × e-97)	Carboxylesterase Familie
2442...1111	Phosphopyruvat Hydratase	<i>Geobacillus kaustophilus</i>	99 % (0,0)	Enolase, N- und C-terminale Domänen
3980...2445	2,3-Bisphosphoglycerat-unabhängige Phosphoglyceratmutase	<i>Geobacillus thermoleovorans</i> CCB_US3_UF5	100 % (0,0)	Metalloenzyme Superfamilie, BPG-unabhängige PGAM, N-terminus
4734...3973	Triosephosphat-Isomerase	<i>Geobacillus kaustophilus</i> HTA426	100 % (0,0)	Triosephosphat-Isomerase
5980...4796	Phosphoglyceratkinase	<i>G. kaustophilus</i> HTA426	99 % (0,0)	Phosphoglyceratkinase

Tabelle 54: Sequenzanalyse der nicht zuordenbaren *contigs*.

ORF-Position	Annotation	Höchste Übereinstimmung ( <i>blastp</i> )	Identität ( <i>e-value</i> )	Proteinklassifikation (Pfam), signifikante Domänen und Familien
<i>contig 27 (6,12 kbp Größe)</i>				
100...1194	Hypothetisches Protein	<i>Alistipes indistinctus</i>	31 % (9 × e-34)	Protein unbekannter Funktion (DUF3089)
1250...3424	Hypothetisches Protein	<i>Dehalobacter sp. UNSWDHB</i>	34 % (6 × e-63)	Amidohydrolase
3654...4019	Hypothetisches Protein	-	-	-
4112...5110	Hypothetisches Protein	<i>Eubacterium sp. 14-2</i>	41 % (6 × e-66)	Protein unbekannter Funktion (DUF3089)
6081...5695	Hypothetisches Protein	-	-	-
<i>contig 33 (4,47 kbp Größe)</i>				
6...1376	Glycosidhydrolase Familie 5	<i>Kribbella flavida</i> DSM 17836	38 % (7 × e-96)	Glycosidhydrolase Familie 5
1409...2677	Hypothetisches Protein	<i>Alicyclobacillus acidoterrestris</i>	43 % (5 × e-121)	Bakterielles extrazelluläres Solut-bindendes Protein
2739...3590	Hypothetisches Protein	<i>Sulfobacillus acidophilus</i> TPY	43 % (2 × e-70)	Bindeprotein-abhängige Komponente des inneren Membrantransportsystems
3587...4414	ABC Transporter Permease	<i>Alicyclobacillus acidoterrestris</i>	50 % (1 × e-80)	Bindeprotein-abhängige Komponente des inneren Membrantransportsystems
<i>contig 41 (3,80 kbp Größe)</i>				
775...2	Hypothetisches Protein	<i>Opitutus terrae</i> PB90-1	59 % (6 × e-112)	Sulfatase-modifizierender Faktor Enzym 1
1616...924	Xylanase	Unkultiviertes Bakterium	76 % (3 × e-102)	Glycosidhydrolase Familie 11
3467...2334	Protein mit konservierter Domäne	<i>Verrucomicrobiae</i> Bakterium DG1235	54 % (2 × e-123)	Domäne unbekannter Funktion (DUF1080)
<i>contig 49 (2,84 kbp Größe)</i>				
1474...2	Putative Endo-1,4-beta-Xylanase	<i>Fibrobacter succinogenes subsp. succinogenes</i> S85	63 % (0,0)	Glycosidhydrolase Familie 43 (GHF 43), 2x CBM 6
1814...1344	Putative Endo-1,4-beta-Xylanase	<i>Fibrobacter succinogenes subsp. succinogenes</i> S85	67 % (5 × e-46)	Glycosidhydrolase Familie 43
2475...1957	Hypothetisches Protein	<i>Fibrobacter succinogenes subsp. succinogenes</i> S85	37 % (3 × e-24)	-
<i>contig 61 (2,2 kbp Größe)</i>				
1037...63	Xylosidase/Arabinosidase	<i>Bacteroides sp. CAG:754</i>	71 % (1 × e-166)	Glycosidhydrolase Familie 43
2139...1066	Endo-1,4-beta-Xylanase	<i>Bacteroides</i>	49 % (3 × e-109)	Glycosidhydrolase Familie 10
<i>contig 88 (1,36 kbp Größe)</i>				
1037...3	Glycosidhydrolase Familie 10	Unkultivierter Organismus	55 % (1 × e-119)	Glycosidhydrolase Familie 10
<i>contig 100 (1,15 kbp Größe)</i>				
630...55	Glycosidhydrolase	<i>Rhodothermus marinus</i> DSM 4252	99 % (3 × e-120)	Glycosidhydrolase Familie 10

**Tabelle 55: Sequenzanalyse des contigs12.**

ORF-Position	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
261...2921	DNA-Topoisomerase I Untereinheit Omega	<i>Pseudoalteromonas agarivorans</i>	92 % (0,0)	DNA Topoisomerase, C4 Zinkfinger, Zink-ribbon-ähnlich
3067...3522	Carboxymuconolacton-Decarboxylase	<i>Halomonas</i>	53 % ( $2 \times e-48$ )	Carboxymuconolacton-Decarboxylase Familie
3532...3717	Hypothetisches Protein	<i>Alteromonas macleodii</i> str. 'Black Sea 11'	57 % ( $2 \times e-08$ )	YrhK-ähnliches Protein
3982...3785	Hypothetisches Protein	<i>Pseudoalteromonas haloplanktis</i>	78 % ( $7 \times e-29$ )	-
5269...4817	Hypothetisches Protein	<i>Pseudoalteromonas</i> sp. PAMC 22718	82 % ( $6 \times e-90$ )	Uncharakterisierte Proteinfamilie, UPF0114
6514...4877	Alpha-Glucosidase	<i>Pseudoalteromonas haloplanktis</i>	86 % (0,0)	Alpha Amylase, katalytische Domäne
7185...9974	Ligand-abhängiges Kanalprotein	<i>Pseudoalteromonas haloplanktis</i>	90 % (0,0)	TonB-abhängiger Rezeptor Anschlussdomäne
10106...11575	Tryptophanhalogenase	<i>Pseudoalteromonas haloplanktis</i>	78 % (0,0)	Tryptophanhalogenase
11578...13428	Alpha-Amylase	<i>Pseudoalteromonas arctica</i>	71 % (0,0)	Cyclomaltodextrinase, Alpha-Amylase, katalytische Domäne
13523...15397	Glycosidase	<i>Pseudoalteromonas haloplanktis</i> TAC125	71 % (0,0)	Alpha Amylase, katalytische Domäne
15390...16874	Zuckertransporter	<i>Pseudoalteromonas</i> sp. SM9913	88 % (0,0)	Major Facilitator Superfamilie, Transporter
16916...19294	Glucan 1,4-alpha-Glucosidase	<i>Pseudoalteromonas</i>	84 % (0,0)	Glucodextranase, Glycosidhydrolase Familie 15 (GHF 15)

**Tabelle 56: Sequenzanalyse des contigs 17 .**

ORF-Position	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
946...2169	Hypothetisches Protein	<i>Fibrella aestuarina</i> BUZ 2	33 % ( $5 \times e-04$ )	-
2166...3629	NHL-repeat-enthaltendes Protein	<i>Opitutus terrae</i> PB90-1	39 % ( $1 \times e-29$ )	-
4833...3703	Lipase GDSL Familie	<i>Opitutus terrae</i> PB90-1	61 % ( $3 \times e-132$ )	GDSL-ähnliche Lipase/Acylhydrolasefamilie
6567...5023	Glycosidhydrolase	<i>Opitutus terrae</i> PB90-1	59 % (0,0)	Pectatelyase Protein Superfamilie
7247...6696	Ketopantoatreduktase	<i>Caulobacter vibrioides</i>	52 % ( $1 \times e-28$ )	Ketopantoatreduktase PanE/ApbA
7401...7781	Transkriptioneller Regulator, PadR-Familie	<i>Terriglobus roseus</i> DSM18391	60 % ( $6 \times e-39$ )	Transkriptioneller Regulator PadR-ähnlicheFamilie
8907...7807	Hypothetisches Protein	-	-	-

Tabelle 57: Sequenzanalyse des Fosmids AZ2-4-B6.

ORF-Position	Annotation	Höchste Übereinstimmung ( <i>blastp</i> )	Identität ( <i>e-value</i> )	Proteinklassifikation (Pfam), signifikante Domänen und Familien
<i>contig 14</i>				
6 ... 461	Hypothetisches Protein	<i>Caldimonas manganoxidans</i>	73 % (7 × e-71)	CRISPR-assoziiertes Protein Cse2
352 ... 1599	Hypothetisches Protein	<i>Caldimonas manganoxidans</i>	75 % (0,0)	CT1975-ähnliches Protein
1603... 2316	Hypothetisches Protein	<i>Caldimonas manganoxidans</i>	79 % (3 × e-115)	CRISPR-assoziiertes Protein (Cas_Cas5)
2313...3011	Hypothetisches Protein	<i>Caldimonas manganoxidans</i>	75 % (4 × e-119)	CRISPR-assoziiertes Protein
3008...3907	Hypothetisches Protein	<i>Caldimonas manganoxidans</i>	89 % (1 × e-172)	CRISPR-assoziiertes Protein Cas1
4431...5246	Transkriptioneller Regulator	<i>Azoarcus sp.</i> BH72	55 % (6 × e-75)	Bakterielle Regulatorproteine gntR Familie, FCD Domäne
6177...5386	4-oxalocrotonat- Decarboxylase	<i>Burkholderia multivorans</i>	69 % (2 × e-125)	Fumarylacetoacetate (FAA) Hydrolasefamilie
7056...6268	2-hydroxypent-2,4-dienoat- Hydratase	<i>Comamonas testosteroni</i>	69 % (1 × e-113)	Fumarylacetoacetate (FAA) Hydrolasefamilie
7928...7053	Meta-fission Produkt Hydrolase	<i>Dyella ginsengisoli LA-4</i>	71 % (1 × e-149)	Alpha/beta Hydrolase Familie 6
9592...8051	Putative 5-carboxymethyl-2- Hydroxymuconat Semialdehydedehydrogenase Oxidoreduktase Protein	<i>Burkholderia vietnamiensis</i>	79 % (0,0)	Aldehyddehydrogenase Familie
10400...9696	Tryptophansynthase Alpha- Untereinheit	<i>Acidobacterium capsulatum</i> ATCC 51196	96 % (4 × e-118)	Tryptophansynthase Alpha-Kette

Tabelle 58: Sequenzanalyse des *shotgun*-Klons aus Fosmid M12-4-D9.

ORF-Position	Annotation	Höchste Übereinstimmung ( <i>blastp</i> )	Identität ( <i>e-value</i> )	Proteinklassifikation (Pfam), signifikante Domänen und Familien
1 ... 86	Sequenz des pCR-TOPO Vektorrückgrates			
392...1006	Hypothetisches Protein Carboxylesterase Est30	<i>Caldibacillus debilis</i> <i>Geobacillusstearothermophilus</i>	99 % (1 × e-141) 71 % (8 × e-102)	Alpha/beta Hydrolase Familie 6

Tabelle 59: Sequenzanalyse des Fosmids AZ3-14-D2.

ORF-Position	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
<i>contig 1</i>				
1 ... 11571	entspricht der Sequenz des pCT3FK Fosmidvektor-Rückgrats			
12240...12689	Pyrimidindimer DNA-Glycosylase	<i>Acidianus hospitalis</i> W1	72 % (4 × e-68)	-
<i>contig 2</i>				
203 ... 1102	AMP-abhängige Synthetase und Ligase	<i>Metallosphaera sedula</i> DSM 5348	72% (3 × e-153)	AMP bindendes Enzym, C-terminale Domäne
1169 ... 2359	Acetyl-CoA Acetyltransferase	<i>Metallosphaera sedula</i> DSM 5348	84 % (0,0)	Thiolase, N- und C-terminale Domänen
2356 ... 2886	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	68 % (1 × e-82)	Domänen unbekannter Funktion (DUF35)
3370 ... 3011	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	44 % (4 × e-15)	-
4014 ... 4160	Hypothetisches Protein	-	-	-
5166 ... 4972	Hypothetisches Protein	-	-	-
5973 ... 5488	N-Acetyltransferase GCN5	<i>Acidianus hospitalis</i> W1	63 % (2 × e-56)	Acetyltransferase (GNAT) Familie
6361 ... 6675	Fe <sub>2</sub> -S <sub>4</sub> Ferredoxin	<i>Acidianus hospitalis</i> W1	94 % (8 × e-67)	Fe <sub>2</sub> -S <sub>4</sub> Dicluster Domäne
7366 ... 8091	Elektronentransfer- Flavoprotein Beta-Untereinheit EtfB	<i>Sulfolobus solfataricus</i> P2	71 % (6 × e-120)	Elektronentransfer Flavoprotein Domäne
8088 ... 8945	Elektronentransfer- Flavoprotein Alpha-Untereinheit	<i>Sulfolobus islandicus</i> Y.G.57.14	68 % (2 × e-136)	Elektronentransfer Flavoprotein Domäne, FAD-Bindedomäne
10870 ... 9341	AMP-abhängige Synthetase und Ligase	<i>Sulfolobus islandicus</i> HVE10/4	96 % (0,0)	AMP bindendes Enzym, C-terminale Domäne
11895...11050	Hypothetisches Protein	-	-	-
<i>contig 3</i>				
507 ... 199	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	59 % (4 × e-33)	-
734 ... 1498	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	66 % (2 × e-122)	Dioxygenase
2874 ... 1693	Gluconatdehydrogenase	<i>Acidianus hospitalis</i> W1	94 % (0,0)	Mandelatracemase, Muconat-lactonisierendes Protein, Enolase
2931 ... 3419	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	62 % (7 × e-67)	NTPase Familie
4446 ... 3499	Alkoholdehydrogenase	<i>Metallosphaera sedula</i> DSM 5348	68 % (7 × e-152)	Alkoholdehydrogenase Domäne Zink-bindende Dehydrogenase
6411 ... 4447	3-Hydroxyacyl-CoA Dehydrogenase	<i>Metallosphaera sedula</i> DSM 5348	58 % (0,0)	3-Hydroxyacyl-CoA Dehydrogenase, NAD-Bindung und C-terminale Domäne, Enoyl-CoA Hydratase/Isomerase Familie
7621 ... 6413	Acyl-CoA Dehydrogenase Domänen-haltiges Protein	<i>Metallosphaera sedula</i> DSM 5348	75 % (0,0)	Acyl-CoA Dehydrogenase, N-terminale, Mittel- und C-terminale Domäne
<i>contig 4</i> mit 1,46 kbp Länge:				
3 ... 743	Transposase	<i>Escherichia coli</i>	99 % (0,0)	Transposase DDE Domäne
1140 ... 724	Transposase	<i>Escherichia coli</i>	86 % (1 × e-76)	-

Tabelle 60: Sequenzanalyse des Fosmids AZ3-14-D11.

ORF-Position	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
<i>contig 1</i>				
1607 ... 966	30S ribosomales Protein S6	<i>Sulfolobus acidocaldarius</i> DSM 639	71 % (4 × e-111)	Ribosomales Protein S6e
1671 ... 2024	Sirohydrochlorin Cobaltochelatase	<i>Sulfolobus tokodaii</i> str. 7	74 % (5 × e-57)	Cobalaminbiosynthese CbiX
2468 ... 2010	30S ribosomales Protein S15P	<i>Sulfolobus tokodaii</i> str. 7	79 % (5 × e-76)	Ribosomale S13/S15 N-terminale Domäne, Ribosomales Protein S15
3490 ... 2474	Methioninsynthase	<i>Sulfolobus tokodaii</i> str. 7	80 % (0,0)	Cobalamin-unabhängige Synthase, Katalytische Domäne
4457 ... 3483	5-Methyltetrahydropteroyltriglutamat-homocystein Methyltransferase	<i>Sulfolobus tokodaii</i> str. 7	58 % (2 × e-126)	Cobalamin-unabhängige Synthase, N-terminale Domäne
5197 ... 4457	Proliferierendes nukleäres Zellantigen	<i>Sulfolobus tokodaii</i> str. 7	68 % (5 × e-122)	Proliferierendes nukleäres Zellantigen, N und C-terminale Domäne
5500 ... 6543	Cobyrinsäure A,C-Diamidsynthase	<i>Sulfolobus acidocaldarius</i> DSM 639	61 % (3 × e-151)	CobB/CobQ-ähnliche Glutamin Amidotransferase Domäne
7281 ... 6514	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	76 % (7 × e-137)	Vorhergesagtes Membranprotein (DUF2208)
7656 ... 7925	Hypothetisches Protein	-	-	-
7943 ... 8602	3,4-Dihydroxy-2-Butanon 4-Phosphatsynthase	<i>Sulfolobus tokodaii</i> str. 7	72 % (2 × e-110)	3,4-Dihydroxy-2-Butanon 4-Phosphatsynthase
8599 ... 9072	Riboflavinsynthase	<i>Sulfolobus islandicus</i> M.14.25	82 % (4 × e-86)	6,7-Dimethyl-8-Ribityllumazinsynthase
9050 ... 9520	6,7-Dimethyl-8-Ribityllumazinsynthase	<i>Sulfolobales</i> archaeon Acd1	83 % (1 × e-89)	6,7-Dimethyl-8-Ribityllumazinsynthase
9538 ... 10176	Hypothetisches Protein	<i>Sulfolobales</i> archaeon Acd1	45 % (3 × e-69)	GGDN Familie
10158 ... 11267	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	48 % (8 × e-120)	Amidohydrolase family
11986 ... 11234	DNA-Polymerase DNA Kammer (sliding clamp) B	<i>Sulfurisphaera ohwakuensis</i>	61 % (4 × e-103)	Proliferierendes nukleäres Antigen, C-terminale Domäne
12324 ... 11983	Hypothetisches Protein	<i>Sulfolobus acidocaldarius</i> DSM 639	58 % (5 × e-17)	RNA-Polymerase Rpo13 Untereinheit HTH Domäne
13549 ... 12317	Hypothetisches Protein	<i>Sulfolobus acidocaldarius</i> DSM 639	58 % (8 × e-170)	ATPases assoziierte mit diversen zellulären Aktivitäten
14027 ... 13551	Hypothetisches Protein	<i>Sulfolobus acidocaldarius</i> N8	34 % (1 × e-07)	-
14068 ... 16803	DEAD/DEAH Box-Helicase	<i>Sulfolobus acidocaldarius</i> DSM 639	67 % (0,0)	DEAD/DEAH Box-Helicase und zugehörige Domänen
16845 ... 17135	50S ribosomales Protein L14e	<i>Sulfolobus acidocaldarius</i> DSM 639	77 % (2 × e-48)	Kyprides/Ouzounis/Woese (KOW)-Motif
17137 ... 17388	H/ACA RNA-Proteinkomplex-Komponente, Cbf5p N-terminales Exon	<i>Acidianus hospitalis</i> W1	74 % (1 × e-38)	DKCLD (NUC011) Domäne
17495 ... 18175	Cbf5	<i>Stygiolobus azoricus</i>	100% (5 × e-126)	Pseudouridinsynthase und Archaeosintransglycosylase (PUA)
18162 ... 18734	Hypothetische Methyltransferase	<i>Sulfolobus acidocaldarius</i> DSM 639	65 % (7 × e-82)	Methyltransferase kleine Domäne
19774 ... 20469	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	71 % (2 × e-111)	-
<i>contig 2</i> mit 11,64 kbp Länge entspricht vollständig der pCT3FK Fosmidvektor Rückgrat-Sequenz				

**Tabelle 61: Sequenzanalyse des Fosmids AZ3-32-E5.**

ORF-Position	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
<i>contig 1</i>				
166 ... 1870	Cytochrom c Oxidase Polypeptid I	<i>Sulfolobus tokodaii</i> str. 7	74 % (0,0)	Cytochrom C und Quinol-Oxidase Polypeptid I (COX)
2289 ... 2654	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	85 % (6 × e-66)	Periplasmatisches Bindeprotein
3225 ... 3091	Hypothetisches Protein	<i>Sulfolobus islandicus</i> REY15A	76 % (5 × e-07)	-
4880 ... 3234	ABC-Transporter Permease	<i>Metallosphaera yellowstonensis</i>	61 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran
5800 ... 4877	ABC-Transporter in Verbindung mit bakteriellem G-3-P Import	<i>Metallosphaera cuprina</i> Ar-4	66 % (6 × e-144)	ABC Transporter und Transport-assoziierte OB
7036 ... 5834	Familie 1 Extrazelluläres Solut-Bindeprotein	<i>Acidianus hospitalis</i> W1	69 % (0,0)	Bakterielles extrazelluläres Solut-Bindeprotein
7077 ... 7313	Hypothetisches Protein	-	-	-
8010 ... 7813	Hypothetisches Protein	-	-	-
8086 ... 7813	Hypothetisches Protein	-	-	-
8692 ... 8901	Hypothetisches Protein	<i>Sulfolobus acidocaldarius</i> DSM 639	68 % (4 × e-18)	Protein unbekannter Funktion (DUF3311)
8914 ... 10536	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	71 % (0,0)	Protein unbekannter Funktion (DUF2434)
11186 ... 10869	Beta-Lactamase Domänen-enthaltendes Protein	<i>Acidianus hospitalis</i> W1	62 % (2 × e-33)	Metallo-beta-Lactamase Superfamilie
11660 ... 13303	Dipeptidylaminopeptidase/ Acylaminoacyl-Peptidase	<i>Sulfolobus islandicus</i> LAL14/1	60 % (0,0)	Prolyl-Oligopeptidase Familie S9
13412 ... 15016	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	79 % (0,0)	Pyrrolo-quinolin Quinon (PQQ)-ähnliche Domäne
15692 ... 15549	Hypothetisches Protein	-	-	-
15945 ... 15796	Hypothetisches Protein	-	-	-
<i>contig 2</i> mit 11,64 kbp Länge entspricht vollständig der pCT3FK Fosmidvektor Rückgrat-Sequenz				
<i>contig 3</i> umfasst 952 bp Sequenzinformation				
523... 2	Putative Transposase, partiell	<i>Escherichia coli</i> O26:H11 str. CVM9942	100 % (3 × e-124)	Transposase DDE Domäne
843... 415	Transposase	<i>Escherichia coli</i>	87 % (3 × e-81)	-

Tabelle 62: Sequenzanalyse des Fosmids AZ3-33-C12.

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
1	Protein der PP-loop-Familie	<i>Acidianus hospitalis</i> W1	75 % ( $4 \times e^{-55}$ )	-
2	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	66 % ( $1 \times e^{-59}$ )	-
3	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	60 % ( $6 \times e^{-23}$ )	-
4	Hypothetisches Protein	-	-	-
5c	Hypothetisches Protein	<i>Thermoplasmatales archaeon</i> Gpl	65 % ( $4 \times e^{-7}$ )	-
6c	Hypothetisches Protein	<i>Metallosphaera sedula</i> DSM 5348	60 % ( $7 \times e^{-43}$ )	-
7	Hypothetisches Protein	<i>Sulfolobus tokodaii</i> str. 7	73 % (0,0)	Dimerisierungsdomäne, Peptidase M20/M25/M40 Familie
8c	Threoninsynthase	<i>Sulfolobus tokodaii</i> str. 7	76 % (0,0)	Pyridoxalphosphat-abhängiges Enzym (PALP)
9	Acetyl-Coenzyme A Synthetase Alpha-Domäne-enhaltendes Protein	<i>Acidianus hospitalis</i> W1	85 % (0,0)	CoA Bindedomäne Succinyl-CoA Ligase-ähnliche Flavodoxin Domäne ATP-grasp Domäne
10	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	80 % ( $2 \times e^{-151}$ )	-
11	Sulfid:Quinon Oxidoreduktase	<i>Acidianus hospitalis</i> W1	84 % (0,0)	Pyridinnukleotiddisulphid-Oxidoreduktase
12c	Chloridkanal Protein	<i>Acidianus hospitalis</i> W1	69 % (0,0)	Spannungsgesteuerter Chloridkanal CBS Domänen
13	Aminosäuretransporter	<i>Sulfolobus tokodaii</i> str. 7	83 % ( $9 \times e^{-159}$ )	Aminosäure-Permease
14 - 27 (DNA Positionen 13765 bis 23524: pCT3FK Fosmidvektor Rückgrat)				
28	Tricorn Peptidase	<i>Pyrobaculum aerophilum</i> str. IM2	47 % (0,0)	Tricorn Protease C1 und PDZ Domäne, Peptidase Familie S41
29	Aminosäure Permease-assoziiertes Protein	<i>Acidianus hospitalis</i> W1	73 % (0,0)	Aminosäure-Permease
30	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	63 % ( $2 \times e^{-81}$ )	Protein unbekannter Funktion (DUF1404)
31c	Natrium:Protonen-Antiporter	<i>Metallosphaera yellowstonensis</i>	60 % ( $2 \times e^{-143}$ )	Natrium/Protonen Austauscher
32	Hypothetisches Protein	<i>Metallosphaera sedula</i> DSM 5348	49 % ( $1 \times e^{-109}$ )	-
33c	Hypothetisches Protein	<i>Metallosphaera sedula</i> DSM 5348	51 % ( $5 \times e^{-82}$ )	Domäne unbekannter Funktion (DUF929)
34c	Hypothetisches Protein	<i>Sulfolobus solfataricus</i> P2	47 % ( $5 \times e^{-57}$ )	-
35	Hypothetisches Protein	<i>Metallosphaera sedula</i> DSM 5348	47 % ( $7 \times e^{-14}$ )	Protein unbekannter Funktion (DUF3311)
36	Na <sup>+</sup> /solute-Symporter	<i>Acidianus hospitalis</i> W1	72 % (0,0)	Natrium:Solut Symporter Familie
37c	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	62 % ( $8 \times e^{-34}$ )	-
38c	DEAD/DEAH Box-Helicase	<i>Metallosphaera sedula</i> DSM 5348	67 % ( $2 \times e^{-170}$ )	DEAD/DEAH Box-Helicase konservierte C-terminale Domäne
39c	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	43 % ( $3 \times e^{-64}$ )	-
40c	Hypothetisches Protein	<i>Sulfolobus islandicus</i> Y.G.57.14	64 % ( $1 \times e^{-06}$ )	-
41	Hypothetisches Protein	<i>Metallosphaera sedula</i> DSM 5348	58 % ( $5 \times e^{-70}$ )	Zink-ribbon Domäne
42c	Hypothetisches Protein	<i>Acidianus hospitalis</i> W1	68 % ( $3 \times e^{-89}$ )	-

**Tabelle 63: Sequenzanalyse des Fosmids M12-4-G10.**

ORF	Annotation	Höchste Übereinstimmung (blastp)	Identität (e-value)	Proteinklassifikation (Pfam), signifikante Domänen und Familien
1c	HemK Familie Modifikation Methylase	<i>Sphaerobacter thermophilus</i> DSM 20745	94 % (8 × e-160)	Methyltransferasedomäne
2	Hypothetisches Protein	-	-	-
3c	Peptidkette Release Factor 1	<i>Sphaerobacter thermophilus</i> DSM 20745	98 % (0,0)	Peptidketten Release Factor Domäne (PCRF und RF-1)
4c	Rossmann-Faltung Nukleotidbindendes Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	98 % (7 × e-104)	Mögliche Lysindecaboxylase
5c	2'-5' RNA-Ligase	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (4 × e-87)	2'-5' RNA-Ligase Superfamilie
6	Vorhergesagter Co/Zn/Cd Kationtransporter	Unkultiviertes Bakterium	40 % (4,4)	-
7c	RDD Domänen-haltiges Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	88 % (5 × e-171)	-
8c	Familie 5 extrazelluläres Solutbindendes Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	95 % (0,0)	Familie 5 bakterielles, extrazelluläres Solutbindendes Protein, mittlere Domäne
9c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran N-terminale Transmembrandomäne einer Oligopeptidtransport-Permease C
10c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	97 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran
11c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran N-terminale Transmembrandomäne einer Oligopeptidtransport-Permease C
12c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran
13	Hypothetisches Protein	-	-	-
14c	4-Phytase	<i>Sphaerobacter thermophilus</i> DSM 20745	94 % (0,0)	Familie 5 bakterielles, extrazelluläres Solutbindendes Protein, mittlere Domäne
15	Hypothetisches Protein	-	-	-
16	Prolindehydrogenase	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Prolindehydrogenase
17c	DegT/DnrJ/EryC1/StrS Aminotransferase	<i>Sphaerobacter thermophilus</i> DSM 20745	94 % (0,0)	DegT/DnrJ/EryC1/StrS Aminotransferase-Familie
18c	Hypothetisches Protein	<i>Sphingobium quisquiliarum</i> P25	40 % (2 × e-141)	Metallo-beta-Lactamase Superfamilie Klasse B Alkyl sulfatase C-terminale und Dimerisationsdomänen
19	Prephenatdehydratase	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Prephenatdehydratase, ACT Domäne
20	Methyltransferase	<i>Sphaerobacter thermophilus</i> DSM 20745	97 % (9 × e-135)	Konserviertes hypothetisches Protein 95
21	Pantetheinphosphat Adenyltransferase	<i>Sphaerobacter thermophilus</i> DSM 20745	98 % (1 × e-116)	Cytidylyltransferase
22	Vaskuläre-Typ H <sup>+</sup> -ATPase Untereinheit H	<i>Sphaerobacter thermophilus</i> DSM 20745	99 % (2 × e-90)	-
23	Hypothetisches Protein	-	-	-
24	Hypothetisches Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	98 % (1 × e-26)	-
25	Hypothetisches Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	91 % (5 × e-177)	-
26c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	97 % (1 × e-177)	N-terminale Transmembrandomäne einer Oligopeptidtransport-Permease C Bindeprotein-abhängige Transportsystemkomponente der inneren Membran
27c	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran	<i>Sphaerobacter thermophilus</i> DSM 20745	98 % (0,0)	Bindeprotein-abhängige Transportsystemkomponente der inneren Membran

28c	Familie 5 extrazelluläres Solut-bindendes Protein, mittlere Domäne	<i>Sphaerobacter thermophilus</i> DSM 20745	95 % (0,0)	Familie 5 bakterielles, extrazelluläres Solut-bindendes Protein, mittlere Domäne
29	Hypothetisches Protein	-	-	-
30	Anthranilatsynthase Komponente I	<i>Sphaerobacter thermophilus</i> DSM 20745	97 % (0,0)	Anthranilatsynthase Komponente I, N-terminale Region, Chorismat Bindeprotein
31	Anthranilatphosphoribosyltransferase	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Glycosyltransferase Familie, Helikales Bündel und a/b Domäne
32 bis 44 (entsprechende DNA-Koordinaten 29,10 bis 40,7 kbp) entsprechen dem pCT3FK Fosmidvektor Rückgrat				
45	Hypothetisches Protein	-	-	-
46c	Hypothetisches Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	94 % (5 × e-132)	Menaquinone biosynthesis
47	Hypothetisches Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	84 % (0,0)	-
48c	Hypothetisches Protein	-	-	-
49c	ATP-abhängige DNA-Helikase RecG	<i>Sphaerobacter thermophilus</i> DSM 20745	96 % (0,0)	Helikase konservierte C-terminale Domäne, DEAD/DEAH Box-Helikase, OB-fold Nukleinsäure-bindende Domäne
50c	degV Proteinfamilie	<i>Sphaerobacter thermophilus</i> DSM 20745	92 % (4 × e-173)	Uncharakterisiertes Protein, DegV Familie COG1307
51c	degV Proteinfamilie, Fettsäure-bindendes Protein	<i>Sphaerobacter thermophilus</i> DSM 20745	94 % (0,0)	Uncharakterisiertes Protein, DegV Familie COG1307
52c	Dak Phosphatase	<i>Sphaerobacter thermophilus</i> DSM 20745	95 % (0,0)	Dihydroxyacetonkinase Familie, DAK2 Domäne
53	50S ribosomales Protein L28	<i>Nitrolancetus hollandicus</i>	76 % (1 × e-09)	Ribosomale L28 Familie

### Nukleotid- und Aminosäuresequenz der Endoglucanase Cel12E.

Grau hinterlegte Sequenz entspricht dem vermuteten Signalpeptid, die laut HMM-Modell vorhergesagten Proteinmodule sind eine Glycosidhydrolase 12 (GHF12, rot unterlegt) sowie zwei Cellulose Bindedomänen der Familie 2 (CBM2, blau unterlegt).

```

1  atgaaaagca ttgcaactgg tgtgagtatc ttaatgatac tcgcagtagt gtccactttt
   m k s i a l g v s i l m i l a v v s t f
61  tcggtttcag cccaggagac aacagtgtcg gagttcccgg gaaccagggg atccacagca
   s v s a q e t t v l e f p g t q g s t a
121 ttaatcgaca tgaatggaga cggcatttct gatttcatta tggaataaaa tccgtggaac
   l i d m n g d g i s d f i m e i n p w n
181 atccaagatg ctcagggaaa gcaaataatg gactacgacc cgaacaacaa cgaaattaga
   i q d a q g k q i m e y d p n n n e i r
241 ttctccagta atctcacgga catcgttccc aaaaattcag acaggtggat ttacgggtat
   f s s n l t d i v p k n s d r w i y g y
301 cccggtgat acttcgggag caagccttgg aacagcaaag tagcaggggg acttctgaaa
   p e v y f g s k p w n s k v a g g l l k
361 cttccgaaaa aagtctctga cctctcaggg ttcacgttga agttcgaata tagcctcgaa
   l p k k v s d l s g f t l k f e y s l e
421 cagcatccga atcttccaat aaactctgca atggaaacct ggtaacgac cgaccagctg
   h d p n l p i n l a m e t w l t t d q l
481 gcaacaagcg ggttaaagc aggggatc gagataatgg tgtggctcta cagtagcaag
   a t s g v k a g d i e i m v w l y s s k
541 ctaaaccccg cggcagaaaa aatcgatacc gtcacggttc caatgatagt caacggcaga
   l n p a g r k i d t v t v p m i v n g r
601 ttagtcaacg agagttttga agtctggaag atggagggta tgggaagcgg ttggacgtac
   l v n e s f e v w k m e g m g s g w t y
661 ttcgcgttca ggtaacaac acccctgaag agtgcagaaa taggaattga cccacgttg
   f a f r l t t p l k s a e i g i d p t l
721 ttcatcaaga aagtctgaag gtacattcag acgaaactgg agaactctca catgcaggac
   f i k k v e g y i q t n v e n l y m q d
781 tgggaaatcg gcaactgagt cggcaatcca actaccacat cggcgtttt taactggacc
   w e i g t e f g n p t t t s a l f n w t
841 ataaggaacc tagaagtcaa cgaagaggca ttgttgccat caagcacctc ctcaccggag
   i r n l e v n e e a l l p s s t s s p e
901 cagaacaaa ctcaccgctc caactacata aactgataa aaccggggag tttggcgtt
   q n q t s p s n t t n l i k p g s l d v
961 aaagttaaca gttggggaag tgccaccag tactcgtgca cgttatacct cggggcccag
   k v h s w g s a t q y s c t l y l g a q
1021 tacgactgga gcgtgaaagt agagctgaaa gacggttctg agatcacaag ctactggagt
   y d w s v k v e l k d g s e i t s y w s
1081 gccagacca ctgaagaaaa tggaccgctc gtgtttacgc cgaagagctg gaacaggggg
   a d a t e e n g t v v f t p k s w n r g
1141 cccacagcaa gcttcggatt tatagcgtcc ggggatgtgc cagtggaatc aataacccta
   p t a s f g f i a s g d v p v e s i t l
1201 atcgttgacg gaaagtttg ggacgtttgg ccaaatgaag ctcaagcgaa cacaacatca
   i v d g k v w d v w p n e a q a n t t s
1261 tcaaatcaaa cttcaacca aatgccaca gccacgaacc aaaccaccaa taccacatca
   s n q t s t q n a t a t n q t t n t t s
1321 acaaacacca cctcaaactc gaatcaaacc tcgaccgatc agaccaaacac aacaaataca
   t n t t s n s n q t s t d q t n t t n t
1381 acatctgaca cctcaaacc cggttctctg agcgtgagca tcaacgactg gggcaccgga
   t s d t l k p g s l s v s i n d w g t g
1441 ggacagtcca acatcaactt ggacctcggg gggcagtatg agtgggtcgt caaagtcaaa
   g q f n i n l d l g g q y e w v v k v k
1501 gttgaatccc ccaccacaat tggcaactac tggagcggcc agaagagcga ggagaacggt
   v e s p t t i g n y w s a q k s e e n g
1561 tggatcatat tcacgccaac gaactggaac aaaggccaa ccgcaagctt tggattcatt
   w i i f t p t n w n k g p t a s f g f i
1621 gtaaatggcc ccgtaaccgg agtcaaacag ataatactgg aagttaacgg ggaattgtgg
   v n g p v t g v k q i i l e v n g e l w
1681 gatattctgga cgccacagtg a
   d i w t p q -

```

**Nukleotid- und Aminosäuresequenz der archaeellen Esterase EstA2.**

Grau hinterlegte Sequenz entspricht dem konservierten Pentapeptid (GXSGX).

```

1  atggttcaga gtaaaaagat aaagatttta cttaattatc ctgatgagac tcctgcagga
   m v q s k k i k i l l n y p d e t p a g
61  tactctatat atgatggaat tttttcaaaa gtatacgatg aaaagggaga attactcttt
   y s i y d g i f s k v y d e k g e l l f
121 gaggttaatg gattatttcc tccaagaata actacaagaa attattcctg gatagagaaa
   e v n g l f p p r i t t r n y s w i e k
181 atattgaatt ctggattatc tgatggaaga aaaagattca tattatatgt agcttcaagg
   i l n s g l s d g r k r f i l y v a s r
241 tacctagtta atgtgaagaa agttgatgaa gaagaagcat taaaagatct gagagatttt
   y l v n v k k v d e e e a l k d l r d f
301 tactataaga acggttcagg aagaatata gatgcatggt taagatctgt aataagggga
   y y k n g s g r i y d a w l r s v i r g
361 gttcaagaga aaaaacttct tcccccttct cttaaaaaca tccaagatag ggataaggaa
   v q e k k l l p p s l k n i q d r d k e
421 ctatatgaag agattacaaa gatttttagaa aagagataa
   l y e e i t k i l e k r -

```

## Nukleotid- und Aminosäuresequenz der bakteriellen Esterase EstB2.

Die vorhergesagten Proteinmodule sind eine Metallo- $\beta$ -Lactamase B, (rot unterlegt) sowie eine Alkyl-Sulfatase Dimerisierungsdomäne (blau) und die C-terminale Domäne der Alkyl-Sulfatase (grün). Die Grau hinterlegte Sequenz entspricht dem konservierten Pentapeptid (GXSGX).

```

1  gtgccaatcc cgaaccact cttctgacg ggaagaacc tgccttcac cgaggaggcc
   v p i p e p l f l t g k n l a f t e e a
61  gcgaggtcc cgccgcgact gaccggccac gcgcgacga tggaccaggg cgtctaccag
   a t v p p r l t g h a r t m d q g v y q
121  gtggccgaga acctctacgt cgccgtcggc tatgggcagg ccaacatgac gatggtggtc
   v a e n f y v a v g y g q a n m t m v v
181  gggaccgacg gctgctgat catcgactgc ctcgagaacg aggagaacgc gcaccaggcg
   g t d g v l i i d c l e n e e n a h q a
241  ctgagcgacc tgcggaaggt cagcgacaag ccgatccagg cgttgatcta cagccacagc
   l s d l r k v s d k p i q a l i y s h s
301  cacccgatc acatctcgg ggtcgcgcg gttctcgatt cggagcaggt cgaaaacggg
   h p d h i s g v r a v l d s e q v e n g
361  caggtggcga tctacgcca cgagcggctg ctcgccgggg tgcggggcaa tcccagtctc
   q v a i y a h e r l l a g v r g n p s l
421  gggatcgtgc cggcgtccg actggcgtac acgttcggct tcgacctgga gcgcgggccc
   g i v p a l r l a y t f g f d l e r g p
481  gaaggtggg tggagaccg cctcgggacg cagtccaca ccggcacgat cggcttcac
   e g w v e t g l g t q f h t g t i g f i
541  ccgccgaaa ccgtgtcca gggcacgctc gacattcagg tggccgggat caccgtgcag
   p p t t v f q g t l d i q v a g i t v q
601  ctccgcgagg cgccgagcga gtcggacgat gagatcgta tgtggttccc ggtcacggc
   l r e a p s e s d d e i v m w f p d h g
661  gtctgcatg tcgcgagct gatccagggc gagacgctac cgaacgtgta ccttttcga
   v l h v a d v i q g e t l p n v y p l r
721  ggccggtcc gtgatccgct gcagtgggtt cggcgcgatc acctgctccg cgagtacgac
   g a v r d p l q w v r a i d l l r e y d
781  gcacaggcac tcattctcgg gcatggctgc ccgctcgtgg gcaaggatga cgcccgggca
   a q a l i f g h g r p l v g k d d a r a
841  ttgctcatgg cctaccggga tgcggtgcag tacgtccatg accaggccgt ccggtgatg
   l l m a y r d a v q y l h d q a v r l m
901  gcccgggct acacccgga cgagctcgca gaaacctca ccgagctacc gtctgcctt
   a r g y t r d e l a e t l t e l p s r l
961  cgtcatcacc cctggctcgg ggagtctac ggcaccgtga agcagacggt gcggcagatc
   r h h p w l g e f y g t v k q t v r q
1021  taccacaact actacgggtg gttccagggg gaccccgctc cgatcgaccc actgccccgc
   y h n y y g w f q g d p a s i d p l p r
1081  cggaacgcy cggccgcta cgtcgcgcy atggggagac gtagcgcggt aattgccaac
   r e r a a r y v d a m g g r d a v i a n
1141  gccgaggcag cgtacaacgc gggcgatttc cgtggggtcg cggagctcct gacgcacctg
   a e a a y n a g d f r w v a e l l t h l
1201  ctccggcttg acccggagga tcaggacgca cgccggctca agcgggatgc cctgcgacag
   l r l d p e d q d a r r l k a d a l r q
1261  ctgggtatc ggaccagcaa cccatctgg cgacaact acctgatggg cgccaaggag
   l g y r t s n p i w r n n y l m g a k e
1321  atcgacggga cgctggatcg tgcccgcctc ctcgccacgc tccgcgcgct ggggaaccgg
   i d g t l d r a r l l a t l r a l g n p
1381  gacgtcgccc ccaccatgcc gatcccgctg ctccttcgcy cgctcgccac ccgctgaac
   d v a a t m p i p l l l r a l a t r l n
1441  ccggcacaga gcgaggggaa gcacctcag gtaggcttcc actgcaccga tacggggcgc
   p a q s e g t h l t v g f h c t d t g a
1501  agctacggcc tcgcatccg ctcgcaogtg gcgaggttcc tcgcaacggc tctgcccagc
   s y g l a i r s d v a e v l a t a p a d
1561  gcggcactgg agatccagac caccgagccg acgctgcgcy gcctcctcac cggccggcta
   a a l e i q t t e p t l r g l l t g r l
1621  ccgtggggcg aggcgctcgc ggacggcgcy gcgacggtca cccagggcac gcccgagggc
   p w a q a v a d g a a t v t q g t p e g
1681  gcaacgcata tctggagcct gttcgatccg cccatggvgy agctcccggc actcgcactg
   a t h l w s l f d p p m g e l p a l a l
1741  cggtag
   r -

```

**Makro-Quellcode (ColourSelectedCells) zur Visualisierung der Daten als *Heatmap*-Repräsentation (für OpenOffice Calc) (aufgefunden auf der Internetseite vom 6. Januar 2014: <https://yacf.co.uk/forum/index.php?topic=18781.0>; A. Angelov, persönliche Mitteilung).**

```
Sub ColourSelectedCells
  oSelect=ThisComponent.CurrentSelection
  oColumn=oselect.Columns
  oRow=oSelect.Rows
  Dim valMax
  Dim valMin
  Dim valTmp

  valMax = oselect.getCellByPosition(0,0).Value
  valMin = oselect.getCellByPosition(0,0).Value

  For nc = 0 To oColumn.getCount-1
    For nr = 0 To oRow.getCount-1
      oCell = oselect.getCellByPosition(nc,nr)
      If oCell.Value > valMax Then
        valMax = oCell.Value
      End If
      If oCell.Value < valMin Then
        valMin = oCell.Value
      End If
    Next nr
  Next nc

  If (valMax-valMin) <> 0 Then
    For nc = 0 To oColumn.getCount-1
      For nr = 0 To oRow.getCount-1
        oCell = oselect.getCellByPosition(nc,nr)
        valTmp = 255*(valMax-oCell.Value)/(valMax-valMin)
        oCell.CellBackColor = RGB (255, valTmp, valTmp)
      Next nr
    Next nc
  End If

End Sub
```

## X. Danksagungen

Mein Dank gilt Prof. Dr. Wolfgang Liebl, der mir die Möglichkeit gegeben hat, an diesem interessanten Forschungsthema teilzuhaben. Als mein Doktorvater und Betreuer bin ich für sein Interesse am Thema, sein offenes Ohr in fachlichen Fragen und für Lösungsvorschläge sehr dankbar.

Dem Zweitprüfer Prof. Dr. Volker Sieber sowie dem Prüfungsvorsitzenden Prof. Dr. Erwin Grill möchte ich herzlich für ihre Bereitschaft an der Teilnahme zur Doktorprüfung danken.

Prof. em. Dr. Karl-Heinz Schleifer sei seiner Bereitschaft als mein Mentor für die TUM-GS gedankt.

Den Kooperationspartnern des ExpresSys-Projektes meinen besten Dank für ausgetauschte Materialien und Know-how: Prof. Dr. Garabed Antranikian, Dr. Kerstin Sahm, Dr. Saskia Blank, Prof. Dr. Jörg Pietruszka, Benjamin Lauinger, Prof. Dr. Wolfgang Streit, Dr. Jennifer Chow, Dr. Thomas Drepper, Prof. Dr. Helena Santos, Luís Gonçalves, Prof. Dr. Peter Golyshin und Dr. Andrea Thürmer.

Meinem Arbeitsgruppenleiter und direkten fachlichen Betreuer Dr. Angel Angelov bin ich für seinen scheinbar unendlichen Ideenreichtum, die motivierende Arbeitsatmosphäre und dem freundschaftlichen Miteinander in allen Fragen rund ums und außerhalb des Labors sehr dankbar. Einen speziellen Dank noch für die tollen „Trips“ an die Felsen und in die Berge, die „Päuschen“ an der frischen Luft und die geistreichen „Gespräche“ mit Filterkaffee. Das war ne echt coole Zeit!

Den 002ern danke ich für die Hilfen um alle kleinen Fragen des Laboralltags und die tolle Zeit im Labor: Beate, Li, Trang, Markus, Milad, Chandrajit, Paul und Philipp. Meinen Studenten gebührt ebenfalls Dank: Denise, Corinna, Matthias, Marie-Therese, Firat, Regine, Wolfgang, Katy, Moe und Flo. Euch viel Erfolg für die Zukunft! Den Nachbarlabors und Kollegen danke ich herzlichst für die kollegiale Zusammenarbeit und den einen oder anderen Plausch im Gang, im Sozialraum, in der Mensa, im Foyer, im Innenhof beim Grillen oder einfach mal ganz privat: Insbesondere meinen Kollegen Claudia, Vladimir, Daniela, Alex, Katharina sowie vor allem Chrissi, Nina, Lena, Wolfgang, Al, Robert, Justyna, Sepp, Liu, Xu, Johanna, Hedwig, Tanja, Jan, Ching-Ning, Marco und die „oberen“ Etagen Schorsch und Timo. David, die Zeit im und außerhalb des Labors bleibt unvergessen!

Spezieller Dank an Angel, Claudia und Trang für das Durchlesen und die Weiterführung der Arbeit.

Meiner Familie, die so einiges in den letzten Jahren miterleben musste, danke ich für die Geduld, die Liebe und den Glauben an mich und meine Arbeit: Insbesondere meinen Eltern Anna und Josef und meiner Oma Angela. Ohne Eure tatkräftige Unterstützung wäre das alles nicht möglich gewesen!

A mi esposa valorosa Alma Delia y a nuestros retoños agradezco su paciencia, su apoyo y su fuente de amor y energía, que son infinitos. Este trabajo también está dedicado a toda la familia en México.

## XI. Lebenslauf

14.06.1983	Geburt in Starnberg
9/1990 – 7/1994	Besuch der Grundschule in Münsing
9/1994 – 6/2003	Besuch des Gymnasiums in Geretsried
10/2003 – 9/2009	Studium der Biologie (Bachelor und Master of Science) an der Technischen Universität München (TUM), Fakultät Wissenschaftszentrum Weihenstephan
3/2009 – 9/2009	Masterthesis unter der Anleitung von Prof. Dr. Martin Loessner: „Design, Expression und biochemische Charakterisierung von rekombinanten bakteriolytischen Enzymen ( <i>enzybotics</i> )“, durchgeführt an der Eidgenössischen Technischen Hochschule (ETH) in Zürich, Schweiz
1/2010	Beginn mit der experimentellen Arbeit zur vorliegenden Dissertation
seit 2/2013	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Mikrobiologie an der Technischen Universität München, Freising  Thema: Entwicklung und Herstellung eines synthetischen Multi-Enzymkomplexes zum effizienten Abbau von cellulosehaltiger Biomasse