TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Mikrobiologie

Isolierung und Charakterisierung von cellulolytischen Bakterien aus der mikrobiellen Lebensgemeinschaft in Biogasanlagen

Daniela Ernestine Köck

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.

Vorsitzende(r): Univ.-Prof. Dr. R. Hückelhoven

Prüfer der Dissertation:

1. Univ.-Prof. Dr. W. Liebl
2. Univ.-Prof. Dr. R. F. Vogel
3. Assist. Prof. Dr. J. Ph. Benz

Die Dissertation wurde am 09.09.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 16.12.2015 angenommen.

I. INHALTSVERZEICHNIS

I. Inhaltsverzeichnis III
II. Abbildungsverzeichnis VII
III. Tabellenverzeichnis XI
IV. Abbkürzungsverzichnis XIV
V. Zusammenfassung XV
VI. Abstract XVI
VII. Publikationen XVII
1 Einleitung 1
1.1 Biogas als regenerative Energiequelle 1
1.2 Lignocellulose als Substrat in Biogasanlagen 4
1.3 Cellulolytische Bakterien 6
1.4 Mechanismen und Enzyme zum Abbau von Cellulose 6
1.5 Die Enzymsysteme cellulolytischer anaerober Bakterien 9
1.5.1 Das Cellulosom, ein effizienter Cellulase-Komplex 9
1.5.2 Nicht-cellulosomale freie Cellulasen 11
1.5.3 Multifunktionale Enzyme 11
1.5.4 Vergleich der Cellulose-Abbaueffizienz 12
1.6 Cellulolytische Mikroorganismen in Biogasanlagen 12
1.7 Analyse der mikrobiellen Lebensgemeinschaften in Biogasanlagen 13
1.8 Isolierung und Charakterisierung von Bakterien 16
1.9 Fragestellung und Ziel der Arbeit 17
2 Material und Methoden 19
2.1 Chemikalien und Verbrauchsmaterialien 19
2.2 Geräte 20
2.3 Biogasfermenter 21
2.4 Medien 22
2.4.1 Basismedien zur Gewinnung neuer Isolate 22
2.4.1.1 GS2-Medium (Johnson et al. 1981) 22
2.4.1.2 Mineralmedium (Komplex) (Widdel and Bak 1992) 23
2.4.1.3 Dilute Nutrient Broth (DNB) (Janssen et al. 2002). 24
2.4.1.4 Basal Salz Medium (Gupta et al. 2012) 24
2.4.2 Sonstige Medien 24
2.4.2.1 LB (lysogeny broth) Medium (pH 7,5) 24
2.4.2.2 SOC Medium (pH 7,5) 25
2.4.2.3 SOB-(Super Optimal Broth) Medium (pH 7,5) 25
2.4.2.4 ZYP5052-Medium (Autoinduktionsmedium) (Studier 2005) 25
2.4.3 Kohlenstoffquellen 26
2.4.4 Sonstige Additive/Zusätze 27
2.5 Bakterienstämme 28
2.6 Methoden zur Isolierung von Bakterien 29

2.6.1 Isolierungsstrategien ... 29
2.6.1.1 Isolierungstrategie 1: ... 29
2.6.1.2 Isolierungstrategie 2 (Schoenborn et al. 2004): ... 29
2.6.1.3 Isolierungsstrategie 3 (Zhang et al. 2003): ... 29
2.6.1.4 “In sacco” Anreicherung ... 29
2.6.2 Reinigung und Identifizierung der Isolate ... 30
2.7 Physiologische Charakterisierung bakterieller Isolate ... 33
2.7.1 Anthron Assay (Dische 1962) ... 33
2.7.2 Zellproteinbestimmung mit Chaps (Reveneau et al. 2003) ... 33
2.7.3 MALDI-TOF MS ... 33
2.7.4 GC ... 35
2.7.4.1 Probenvorbereitung und Auswertung ... 35
2.7.5 Analyse des Fettsäuremusters prokaryontischer Zellen ... 36
2.7.6 Ermittlung des optimalen pH-Wertes und Salzgehalts, sowie der optimalen Temperatur für die Kultivierung neuer Isolate ... 36
2.7.7 Substratspektrum ... 36
2.7.8 Abbau eines natürlichen Substrates (Maissilage) ... 37
2.7.9 Mikroskopie und Elektronenmikroskopie ... 37
2.7.10 Enterotube II ... 38
2.7.11 Gram-Färbung ... 38
2.8 Gentechnische Methoden ... 39
2.8.1 Primer ... 39
2.8.2 Vektoren ... 40
2.8.3 Herstellung Kompetenter Zellen ... 41
2.8.4 Isolierung genomischer Dna ... 41
2.8.5 Agarosegel-Elektrophorese ... 42
2.8.6 PCR ... 43
2.8.7 PCR basierte Genotypisierungsmethoden für Clostridium thermocellum Stämme ... 44
2.8.7.1 PCR-Ribotypisierung ... 44
2.8.7.2 VNTR (Variable number of tandem repeats) ... 44
2.8.7.3 RAPD (Random amplified polymorphic DNA) ... 45
2.8.7.4 RIME (Random inserted mobile element) ... 45
2.8.8 Aufreinigung und Modifikation von PCR-Produkten ... 46
2.8.9 Klonierung ... 47
2.8.10 Klonierung mittels Gibson-Assemblierung ... 47
2.8.11 Transformation chemisch kompetenter Zellen ... 49
2.8.12 Plasmid-Präparation ... 49
2.8.13 Enzymatischer Verdau von DNA ... 49
2.8.14 Sequenzierung ... 50
2.9 Proteinbiochemische Methoden ... 51
2.9.1 Fällung und Aufkonzentrierung von Proteinen ... 51
2.9.2 Bestimmung der Gesamtproteinkonzentration mittels Bradford ... 51

2.9.3 SDS Page ... 52
2.9.4 Western Blot ... 53
2.9.5 Cellulosompräparation ... 53
2.9.6 Affinitätsaufreinigung rekombinanter Proteine ... 55
2.9.6.1 Zellanzucht, -ernte und -Aufschluss ... 55
2.9.6.2 Native His_6-Affinitätsreinigung ... 55
2.9.6.3 Hitzedenaturierung ... 56
2.9.7 Enzymaktivitätsassay mit DNSA (3,5-Dinitrosalicylsäure) ... 56
2.9.8 Dünnschichtchromatographie ... 58
2.10 Bioinformatische Methoden/Software ... 59
2.10.1 Clustal Omega, Tree View ... 59
2.10.2 Blast ... 59
2.10.3 ARB und MEGA ... 59
2.10.4 Clone Manager ... 60
2.10.5 EDGAR ... 60
2.10.6 GEN DB ... 60
2.10.7 MGX ... 60
2.11 Statistische Auswertung ... 60
3 Ergebnisse ... 61
3.1 Ermittlung der genetischen Variation eines für den anaeroben Abbau in der Biogasanlage wichtigen Bakteriums ... 61
3.1.1 Auswahl repräsentativer Biogasanlagen ... 61
3.1.2 Isolierung ... 61
3.1.3 Etablierung der MALDI-TOF MS Analyse zur Stammidentifizierung ... 63
3.1.4 Genotypisierung verschiedener Clostridium thermocellum Isolate ... 64
3.1.4.1 PCR Ribotypisierung ... 64
3.1.4.2 RAPD (Random amplified polymorphic DNA) ... 65
3.1.4.3 VNTR (Variable Number Tandem Repeat) ... 66
3.1.4.4 RIME (Random Inserted Mobile Element) ... 67
3.1.4.5 Zusammenfassung ... 67
3.1.5 SARAMIS MALDI-TOF MS Clusteranalysen der Spektren unterschiedlicher *Clostridium thermocellum* Isolate ... 68
3.1.6 Physiologische Unterschiede der *Clostridium thermocellum* Isolate aus unterschiedlichen Biogasanlagen ... 69
3.1.7 Vergleich des Isolats *Clostridium thermocellum* BC1 mit *Clostridium thermocellum* ATCC 27405 und DSM 1313 auf enzymatischer und genomischer Ebene ... 73
3.2 Isolierung und Charakterisierung neuer Bakterien ... 79
3.2.1 Isolate aus der thermophilen Biogasanlage 4Biogas Nova Fermenter (H1) ... 79
3.2.2 Metagenomanalyse der thermophilen Biogasanlage 4Biogas Nova Fermenter (Viersen) ... 82
3.2.3 Charakterisierung ausgewählter thermophiler Isolate ... 85
3.2.3.1 Physiologische Charakterisierung ausgewählter thermophiler Isolate ... 85
3.2.3.2 Genomsequenzierung ausgewählter thermophiler Isolate ... 92

3.2.3.3 Charakterisierung des cellulolytischen Systems von Isolat T3/55 97
3.2.3.4 Charakterisierung des hemicellulolytischen Systems von Isolat T3/55 103
3.2.4 Isolate aus der mesophilen Biogasanlage Pellmeyer (T3) 104
3.2.5 Charakterisierung ausgewählter mesophiler Isolate 107
3.2.5.1 Physiologische Charakterisierung ausgewählter mesophiler Isolate 107
3.2.5.2 Genomsequenzierung ausgewählter mesophiler Isolate 110
3.3 „in sacco“ Versuche 112
3.3.1 Charakterisierung ausgewählter „in sacco“ Isolate 113
3.3.1.1 Physiologische Charakterisierung GGR1 113
3.3.1.2 Genomsequenzierung GGR1 115
3.3.2 Amplikonsequenzierung der „in sacco“ Anreicherung 116
3.3.2.1 Amplikonsequenzierung BM1T 116
3.3.2.2 Amplikonsequenzierung BM2T 118
3.3.2.3 Amplikonsequenzierung der Anreicherungen 120
3.3.3 Schlussfolgerungen 121
3.4 Vergleich der hydrolytischen Enzyme aller sequenzierten Genome neuer Isolate 123
4 Diskussion 125
4.1 Genotypisierung 125
4.2 Isolate vs. Metagenomanalyse 128
4.3 „In sacco“ Anreicherung vs. Amplikon 132
4.4 Cellulolytische Systeme vs. funktionale Einordnung der Isolate 134
4.5 Ausblick 137
5 Referenzen 138
VIII. Danksagung i
IX. Erklärung ii
X. Anhang iii
i. Eichgeraden iii
Bradford Assay iii
DNS Assay iii
Anthron Assay iii
ii. Genetisch veränderte Organismen und Vektoren iv
iii. 16S rRNA Gensequenzen mesophile Biogasanlage T3 v
iv. 16S rRNA Gensequenzen thermophile Biogasanlage H1 ix
v. Amplikonsequenzierung des BM1T Reaktors xxi
vi. Amplikonsequenzierung des BM2T Reaktor xxviii
vii. Lebenslauf xxxvi

II. ABBILDUNGSVERZEICHNIS

Abbildung 1: Entwicklung der Biogasanlagen in Deutschland - Anlagenzahl differenziert nach Leistungsklassen und installierter elektrischer Anlagenleistung MW_{el} ohne Biogasaufbereitungsanlagen, Deponie- und Klärgasanlagen (FNR 2013) 1

Abbildung 2: Schematische Darstellung des anaeroben Abbaus von lignocellulosehaltiger Biomasse bis zu Methan 2

Abbildung 3: a) Massebezogener Substrateinsatz in Biogasanlagen b) Massebezogener Substrateinsatz nachwachsender Rohstoffe in Biogasanlagen; Daten aus FNR Bericht (FNR 2013) 4

Abbildung 4: Struktur von Lignocellulose haltiger Pflanzenbiomasse (Ratanakhanokchai et al. 2013) 5

Abbildung 5: Klassisches Model für die anaerobe Hydrolyse von Cellulose. Endo-β-1,4-Glucanasen schneiden zufällig innerhalb des Cellulosestrangs. Cellobiohydrolasen (prozessive Exo-Glucanasen) setzen an den Enden prozessiv Cellobiose frei. Betaglucosidase scheidet anschließend die Cellobiose Einheiten. Entnommen aus Koeck et al (Koeck et al. 2014a). Modifiziert nach Watanabe und Tokuda (Watanabe and Tokuda 2010) 8

Abbildung 6: Enzymsysteme anaerober Bakterien für den Abbau lignocellulosehaltiger Biomasse. a) Cellulosom b) Einzelenzyme c) Multidomänenenzyme. CBM: Cellulose Bindungsmodul. SLH: S-layer Homologie. Gleiche Farben symbolisieren eine ursprüngliche Polypeptidkette 9

Abbildung 7:Taxonomische Profile der Bakterien aus verschiedenen mikrobiellen Gemeinschaften in Biogasanlagen basierend auf Metagenomsequenzen. Dargestellt ist die Zuordnung zu den einzelnen phylogenetischen Klassen. Die Autoren vergleichen entweder 16S rRNA Gensequenzen oder EGTs (Environmental Gene Tags) mit den Referenzdatenbanken Ribosomal Database Project Classifier oder ARB (Modifiziert nach Wirth et al. 2012; Kröber et al. 2009; Krause et al. 2008). 15

Abbildung 8: Gesamtansicht und schematische Darstellung des 4Biogas NOVA Fermenter in Viersen 21

Abbildung 9: links: Nylonsäckchen für die „in sacco“ Versuche, rechts: LfL Laborfermenter BM1T und BM2T 30

Abbildung 10: Zusammenfassung der Isolierungs , Reinigungs und Identifikationsschritte für ein neues Isolat. Eine gelbe Färbung ist jeweils gleichzusetzen mit Substrat bzw. Filterpapierabbau 32

Abbildung 11: Schematische Darstellung der Beschleunigung ionisierter Partikel im elektrischen Feld der MALDI-TOF MS (Quelle RIPAC GmbH) 34

Abbildung 12: Beispielspektrum des MALDI-TOF MS Proteinprofils von *Clostridium stercorarium* DSM 8532. 35

Abbildung 13: Auswertung des Enterotube II Testsystems (Quelle: BD Diagnostic Systems) 38

Abbildung 14: Plasmidkarte des pET24a(+) Vektors mit Schnittstellen der MCS 41

Abbildung 15: Schematische Darstellung des *C. thermocellum* (DSM 1313) rRNA Operon mit Bindungsstellen für die Primer zur Amplifikation der Spacerregion 44

Abbildung 16: Schematische Darstellung der RIME (Random inserted mobile element) Methode zur Genotypisierung 46

Abbildung 17: Überblick über die Gibson Assemblierungsmethode (New England Biolabs) 48

Abbildung 18: Typischer Verlauf der Gelfiltrationsreinigung des nativen Cellulosoms auf einer Superdex200 Säule. 54

Abbildung 19: Lysehöfe in PASC-Overlay mit *C. thermocellum* Kolonie im Zentrum, gefärbt mit Kongorot .. 61

Abbildung 20: SARAMIS MALDI-TOF MS Clusteranalysen der *C. thermocellum* und *C. stercorarium* Stämme (vgl. Tabelle 41)..... 63
Abbildung 21: (a) Multiples Sequenz Alignment des „Intergenic Spacer“ der vier rRNA Operone von C. thermocellum (DSM 1313). Lücken in dem unteren Balken symbolisieren Sequenzunterschiede oder Lücken. Basen mit Sequenzunterschieden sind schwarz hinterlegt. Basen, die in allen Sequenzen identisch sind, sind mit einem Stern markiert. (b) Agarosegel der PCR Produkte der Ribotypisierung aller C. thermocellum Isolate aus den verschiedenen BGA und der DSM Referenzstämme und C. stercorarium DSM 8532 (Koeck et al. 2014d) 64
Abbildung 22: RAPD Bandenmuster aller *C.thermocellum* Isolate und Referenz Stämme und *C. stercorarium* DSM 8532 (Cst) (Koeck et al. 2014d) 65
Abbildung 23: (a) Segment des multiplen Alignments des Clo1313_0090 Gens von den *C. thermocellum* Referenzstämmen (YS, DSM2360, DSM 1313; ATCC 27405, JW20) mit variabler Anzahl an repetitiven Sequenzen. Lücken in dem unteren Balken symbolisieren Sequenzunterschiede oder Lücken. Basen mit Sequenzunterschieden sind schwarz hinterlegt. Basen die in allen Sequenzen identisch sind, sind mit einem Stern markiert. (b) Ergebnisse der PCR Amplifikation dieser variablen Region von allen Isolierten Spezies und *C. stercorarium* (DSM8532) und *C. thermocellum* Referenzstämme (DSM2360, DSM4150, ATCC 27405) (Koeck et al. 2014d)..... 66
Abbildung 24: Ergebnisse der PCR Amplifikation der Regionen zwischen zwei Kopien das Mobilelements 1447 im Chromosom aller Isolierten Stämme und Referenzstämme DSM 8532 (*C. stercorarium*) und DSM 2360, ATCC 27405 (*C. thermocellum*) (Koeck et al. 2014d). 67
Abbildung 25: SARAMIS MALDI-TOF MS Clusteranalysen der Spektren unterschiedlicher *C. thermocellum* Isolate aus ausgewählten Biogasanlagen mit *C. stercorarium* Cluster 68
Abbildung 26: Cellulose Abbau (%) (y-Achse) aller Isolate und des *C. thermocellum* Typstammes ATCC 27405 bei verschiedenen pH Werten. Fehlerbalken repräsentieren die Standardabweichung 70
Abbildung 27: DNSA Assay zur Ermittlung der spezifischen Cellulase Aktivität des Kulturüberstandes aller Isolate und *C. thermocellum* ATCC 27405 auf Avicel unter Standardbedingungen. Ausgedrückt als µmol Glucose Äquivalent pro mg Protein aus Zellüberstand und Minute (U/mg). Von jeder Proteinprobe wurden drei Replikate gemessen und die numerische Standardabweichung mit Fehlerbalken dargestellt..... 72
Abbildung 28: Abbau von Maissilage aller Isolate und des *C. thermocellum* Typstamms ATCC 27405 nach Inkubation in GS2 Medium mit 0.5 % Maissilage (wt/vol) für sechs Tage..... 72
Abbildung 29: Spezifische Aktivität in Units pro Milligramm Enzym der Cellulosompräparationen von *C. thermocellum* BC1 (blau), DSM 1313 (rot), ATCC 27405 (grün) bei verschiedenen pH Werten auf Birkenxylan (a), Avicel (b) und Carboxymethylcellulose (c)..... 75
Abbildung 30 Zirkulärer Plot des *C. thermocellum* B1 Genom, von außen nach innen: annotierte Gene des Vorwärts-Strangs, Gene des Rückwärts-Strangs, GC Gehalt, GC Verhältnis (G - C)/(G + C)..... 75
Abbildung 31: Syntenie Plot zur Darstellung der relativen Position der Gene von *C. thermocellum* DSM 1313 (rot) und BC1 (grün) im Vergleich zu Referenzgenom von ATCC 27405. 76
Abbildung 32: Venn Diagramm zur Darstellung des Core Genoms und der „Singelton“ Gene zwischen den drei untersuchten *C. thermocellum* Stämmen (ATCC 27405 (1), DSM 1313 (2), BC1 (3)). 77
Abbildung 33: Phylogenetischer Baum basierend auf dem „Core“ Genom aller bisher Sequenzierten *C. thermocellum* Stämme berechnet mit dem Neigbour joining Algorithmus 78

Abbildung 34: Ungewurzelter phylogenetischer Baum der 16S rRNA Gen Sequenzen aller thermophilen Isolaten und ausgewählter Referenzstämme unter Verwendung eines Maximum likelihood-basierten Algorithmus. Auf den Ästen sind die einzelnen Bootstrap Werte angegeben. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in „Operational taxonomic units“ (OTU`s) zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Isolate, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede. 81
Abbildung 35: Verteilung der Metagenomsequenzen aus verschiedenen DNA Proben auf die bakteriellen Familien 83
Abbildung 36: Kronaplot aller Sequenzen der Metagenomanalyse der Biogasanlage Viersen 84
Abbildung 37: Radiale Darstellung der taxonomischen Hauptgruppen aus dem Metagenom der Biogasanlage Viersen 84
Abbildung 38: Abbau von Maissilage durch *Clostridium thermocellum* ATCC 27405, *Bacillus thermoamylovorans* 1A1 und einer Kokultur aus beiden. Fehlerbalken repräsentieren die Standardabweichung von drei Messwerten 87
Abbildung 39: Vergleich der Zellproteinkonzentration von T3/55 auf verschiedenen Substraten mit der Konzentration der Abbauprodukte nach sieben Tagen Inkubation bei 55°C in GS2 Medium 89
Abbildung 40: Spezifische Aktivität in Units pro Milligramm Protein im Kulturüberstand von C. stercorarium (DSM 8532) und *Herbinix hemicellulosilytica* (T3/55) Zellüberstand auf diversen hemicellulosehaltigen Substraten 90
Abbildung 41: Dünnschichtchromatographie von T3/55 Zellüberstand auf ausgewählten Hemicellulosen. T3 = T3/55, 0 = negativ Kontrolle, S = Standard 90
Abbildung 42: Messung des Abbaugrads von gewaschener Maissilage (~0,5 g in GS2 Medium); Cst = *Clostridium stercorarium;* Cth = *Clostridium thermocellum;* T3.55 = *Herbinix hemicellulosilytica* 91
Abbildung 43: Elektronenmikroskopische Aufnahmen von T3/55 Zellen nach einem (a), drei (b) und fünf (c) Tagen Kultivierung in GS2 Medium bei 55 °C (Koeck et al. 2015b) 92
Abbildung 44 Zirkulärer Plot des *C. cellulosi* DG5 Genoms, von außen nach innen: annotierte Gene des Leitstrangs, Gene des Gegenstrangs, GC Gehalt (schwarz), GC Verhältnis (G - C)/(G + C) (rot) 93
Abbildung 45: Vermeintliche Cellulosomale Komponenten von Isolat SD1D und deren modularer Aufbau, abgeleitet aus der Genomsequenz von SD1D. Doc = Dockerin, cel = Gene die für Cellulasen codieren, xyl = Gene die für Xylanasen codieren. 95
Abbildung 46: Potentielle Cellulasen im T3/55 Genom. Links: Protein-BLAST der Aminosäuresequenz und Lokalisation des Gens im Genom von T3/55. Rechts: Vorhergesagte Domänenstruktur der potentiellen Cellulasen. 97
Abbildung 47: Spezifische Aktivität in Units pro Milligramm Enzym der T3/55 Cellulasen. Links nach einer Stunde Inkubation bei 55 °C (1h). Rechts nach 24 Stunden Inkubation bei 55 °C (24h). 98
Abbildung 48: pH und Temperaturoptimum des GH9 (a), GH48 (b) und GH5 (c) Enzyms von Isolat T3/55 dargestellt als 3D Plot. Die Farbskale zeigt die spezifische Aktivität in U/mg an 99
Abbildung 49: Dünnschichtchromatographien zur Darstellung der Enzymaktivität auf ß-1-4-glycosidisch verknüpften Oligosacchariden nach 0, 0,5, 5 und 60 Minuten sowie über Nacht (ÜN). a) Enzymkinetik T3/55 GH5; b) Enzymkinetik T3/55 GH9; c) Enzymkinetik T3/55 GH48 101

Abbildung 50: Spezifische Aktivitäten in Units pro Milligramm Enzym von GH9 und GH48 Enzymmischungen und Einzelenzymen aus Isolat T3/55 auf PASC mit und ohne ß-Glucosidase BglT (ß-Glu). Messung nach 24 bzw. 48 Stunden (h) 102
Abbildung 51: Ungewurzelter phylogenetischer Baum der 16S rRNA Gen-Sequenzen aller mesophilen Isolate und ausgewählter Referenzstämme unter Verwendung eines Maximum likelihood-basierten Algorithmus. Auf den einzelnen Ästen sind die einzelnen Bootstrap Werte angegeben. 106
Abbildung 52: Zellen von Isolat Iso2/2 37 unter dem Phasenkontrastmikroskop (1000x Vergrößerung). Balken = 10µm 109
Abbildung 53: Zellen von Isolat GGR1 unter dem Phasenkontrastmikroskop (1000x Vergrößerung). Balken = 10 µm 114
Abbildung 54 links: Western Blot mit Antikörpern gegen Cohesine des CipA Protein von *C. thermocellum*. Rechts: SDS PAGE analog zu dem Western Blot links. Cth = *C. thermocellum.* F = Fraktion der Gelfiltration von GGR1 Kulturüberstand. KÜ = Kulturüberstände anderer Isolate (37/hoe, T3/55) als Negativkontrolle. Es wurden jeweils 2 µg Protein aufgetragen. 115
Abbildung 55: a) Phylogenetischer Baum der 16S rRNA Gensequenzen der Amplikonsequenzierung einer „in sacco“ Anreicherungen aus dem BM1T Reaktor unter Verwendung eines maximum likelihood-basierten Algorithmus. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in taxonomische Einheiten zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Sequenzen, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede. b) Prozentualler Anteil der wichtigsten taxonomischen Einheiten in der Amplikonsequenzierung der verschiedenen „in sacco“ Probenahmezeitpunkte von BM1T. Angegeben ist immer die nächste bekannte verwandte Spezies oder Gattung. N_{gesamt} = 70. 117
Abbildung 56: a) Phylogenetischer Baum der 16S rRNA Gen Sequenzen der Amplikonsequenzierung einer „in sacco“ Anreicherungen aus dem BM1T Reaktor unter Verwendung eines maximum likelihood-basierten Algorithmus. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in taxonomische Einheiten zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Sequenzen, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede. b) Prozentualler Anteil der wichtigsten taxonomischen Einheiten in der Amplikonsequenzierung der verschiedenen „in sacco“ Probenahmezeitpunkte von BM1T. Angegeben ist immer die nächste bekannte verwandte Spezies oder Gattung. N gesamt= 75. 119
Abbildung 57: Phylogenetische Analyse der Amplikonsequenzierung des zweiten Anreicherungsschrittes 120
Abbildung 58: Amplikonsequenzierung des fünften Anreicherungsschrittes 121
Abbildung 59: Vergleich der Abundanzen von Kohlenhydrataktiven Enzyme (CAZy), die in den Genomen der neuen Isolate codiert sind. GH = Glycosidhydrolasen, GT = Glucosyltransferasen, CE = Carbohydratesterasen, PL = Polysaccharidlyasen. 123
Abbildung 60: Glycosidhydrolase Profile der Genome aller sequenzierten Isolate. Gezeigt sind ausgewählte Familien welche typischerweise Xylanasen (GH5, GH10, GH11), Cellulasen (GH5, GH9, GH48), Mannanasen (GH5, GH26), β-Glucanasen/Xyloglucanasen (GH16, GH44, GH74), Arabinofuranosidasen/Xylosidasen (GH51, GH43), Glycosidasen (GH2, GH3), und Amylasen (GH13)

enthalten. Die funktionelle Zuordnung der GH Familien wurde aus der CAZy Website entnommen (http://www.cazy.org). 124
Abbildung 61: Kronaplotvergleich des Phylum Firmicutes zwischen der Metagenomanalyse der Biogasanlage H1 (a) und der Isolate aus derselben Biogasanlage (b) 130

III. TABELLENVERZEICHNIS

Tabelle 1: Liste der GHs (Glycosidhydrolasen) die in cellulolytischen Enzymen enthalten sind, basierend auf der CaZy Datenbank (Lombard et al. 2014)). Prozessive Exoglucanansen in kursiv 7
Tabelle 2: Spezifische Aktivität von Kulturüberstand auf Avicel, bestimmt unter Standardbedingungen (Koeck et al. 2014a) 12
Tabelle 3: Verwendete Kits, Enzyme und Längenstandards 19
Tabelle 4: Verwendete Geräte 20
Tabelle 5: Liste aller beprobten Biogasanlagen mit den wichtigsten Prozessparamentern (Temperatur, Substrat) 21
Tabelle 6: GS2 Medium 22
Tabelle 7: Mineralmedium-Stammlösung und Spurenelementlösung SL9 23
Tabelle 8: Selenit-Wolframat-Lösung 23
Tabelle 9: Vitaminlösungen 23
Tabelle 10: Basal Salz Medium 24
Tabelle 11: LB Medium 24
Tabelle 12: SOC Medium 25
Tabelle 13: SOB Medium 25
Tabelle 14: Einzelkomponenten des ZYP5052-Mediums 25
Tabelle 15: ZYP5052-Medium 25
Tabelle 16: Für Kultivierungsmedien und Enzymaktivitätsttests verwendete Kohlenstoffquellen 26
Tabelle 17: Sonstige Zusätze für Kultivierungsmedien 27
Tabelle 18: In dieser Arbeit verwendete Bakterienstämme, die nicht im Laufe dieser Arbeit isoliert wurden 28
Tabelle 19: Anthron Assay 33
Tabelle 20: Tabelle der Retentionszeiten verschiedener Substanzen im Gaschromatographen 35
Tabelle 21: Kakodylatpuffer 37
Tabelle 22: Liste aller in dieser Arbeit verwendeten Oligonucleotid-Primer. F steht jeweils für den Vorwärtsprimer und R für den Rückwärtsprimer. T_m bezeichnet die Schmelztemperatur 39
Tabelle 23: CCMB80 Puffer 41
Tabelle 24: Lysepuffer 42
Tabelle 25: TAE Puffer und 6x DNA Auftragspuffer 43
Tabelle 26: PCR Protokoll Phusion Polymerase 43
Tabelle 27: PCR Protokoll DreamTaq Polymerase. *Tabelle 22 43
Tabelle 28: Amplifikationsprogramm für die Poly-Adenylierung 47
Tabelle 29: Ansatz für die Gibson Assemblierung 48
Tabelle 30: Ansatz Restriktionsverdau 49
Tabelle 31: Ansatz Sequenzierung 50

Tabelle 32: Puffer und Ansätze für SDS Gele 52
Tabelle 33: 10 x SDS Elektrophorese-Laufpuffer 52
Tabelle 34: Färbe-/Entfärbe Lösung für SDS PAGE 53
Tabelle 35: Gelfiltrationspuffer 54
Tabelle 36: Puffer für Proteinaffinitätsaufreinigung 55
Tabelle 37: Reaktionspuffer für Enzym Assays und DNSA Lösung 57
Tabelle 38: DC Färbelösung 58
Tabelle 39: Ermittlung der cfu/ml aus der Anzahl der Lysehöfe nach Kongorotfärbung 61
Tabelle 40: 16S rRNA Sequenzanalyse der Isolate aus verschiedenen Biogasanlagen 62
Tabelle 41: *C. thermocellum* (*C. stercorarium*) Referenzstämme und Eigenisolate zur Etablierung der MALDI TOF MS Analyse 63
Tabelle 42: Zusammenfassung der Ergebnisse von RAPD (Random amplified polymorphic DNA), VNTR (variable number of tandem repeats), Ribotypisierung und RIME (Random inserted mobile element) Assays für die Genotypisierung der *C. thermocellum* Isolate aus verschiedenen Biogasanlagen 67
Tabelle 43: Zellproteinkonzentration von allen isolierten Stämme und *C. thermocellum* ATCC 27405 nach Inkubation für sechs Tage bei verschiedenen pH-Werten und Temperaturen in GS2 Medium mit 0.5 % Avicel (wt/vol). 69
Tabelle 44: Verwertung verschiedener Kohlenhydrate durch *C. thermocellum* Isolate in GS2 Medium. Bestimmung der Zellproteinkonzentration, um Wachstum nachzuweisen. Keines der Isolate konnte auf Xylan, Xylose, Stärke, Galactose, Mannose, Inulin, Sucrose, und Arabinose wachsen 71
Tabelle 45: Wichtige Merkmale der Genome von Clostridium thermocellum BC1, DSM 1313 und ATCC 27405 im direkten Vergleich. 75
Tabelle 46: Liste aller bisherigen Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten 16S rRNA Gen Sequenzähnlichkeit (laut NCBI Datenbankvergleich). Isolate wurden der gleichen Zeile zugeordnet, wenn ihre 16S rRNA Gen Sequenzähnlichkeit mindestens 98 % betrug. [1]Inc. sed. = incertae sedis 80
Tabelle 47: Zusammenfassung der bisherigen Isolate mit drei Isolierungsstrategien 82
Tabelle 48: DNA-Isolierungsmethoden für die Metagenomanalyse der Biogasanlage und Anzahl der generierten und bereinigten Einzelsequenzen 83
Tabelle 49: Liste der zur genaueren Charakterisierung ausgewählten thermophilen Isolate. BGA = Biogasanlage 85
Tabelle 50: Wichtige morphologische und physiologische Eigenschaften von den Isolaten T3/55, SD1D, 1A1, DG5. *Filterpapier (Whatman No. 1); n.b. = nicht bestimmt 86
Tabelle 51: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat T3/55 und SD1D auf verschiedenen Substraten mit Standardabweichung von drei Messwerten 89
Tabelle 52: Zelluläres Fettsäuremuster (> 1 %) von Herbinix hemicellulosilytica T3/55. *Gruppe 1 enthielt 13:1 CIS 12 FAME und 14:0 ALDE; Gruppe 4 enthielt eine unbekannte Fettsäure und 15:2 FAME; Gruppe 6 enthielt 15:0 anteiso 3OH FAME und 16:1 CIS 7 DMA; Gruppe 8 enthielt 17:1 CIS 9 FAME und 17:2 FAME; Gruppe 10 enthielt 18:1c11/t9/t6 FAME eine unbekannte Fettsäure. 91
Tabelle 53: Eigenschaften der von den Isolaten 1A1, T3/55, SD1D und DG5 sequenzierten Genome 92
Tabelle 54: „Readmapping“ der sequenzierten Genome auf den Metagenomdatensatz der Biogasanlage (BGA) Viersen (97% Sequenzübereinstimmung bei mindestens 40 Basenpaaren Überlappung) 94

Tabelle 55: Kohlenhydrat aktive Enzyme die in den Genomen von *Clostridium cellulosi DG5, Bacillus thermoamlylovorans 1A, Herbinix hemicellulosilytica T3/55* und *Herbinix sp. SD1D* codiert sind; CBM: Kohlenhydratbindemodule, CE: Carbohydratesterasen, GH: Glycosidhydrolasen, GT: Glycosyltransferasen. Potentielle Cellulasen sind grün hinterlegt 96
Tabelle 56: Potentielle Xylanasen im T3/55 Genom. GH Familie, Proteinblast (NCBI blastp) der Aminosäuresequenz und Lokalisation des Gens im Genom von T3/55. 103
Tabelle 57: DNSA Assay zur Ermittlung der spezifischen Aktivität der putativen Xylanasen von Isolat T3/55 auf diversen Substraten. Ausgedrückt als µmol Glucose Äquivalent pro mg Zellprotein und Minute (U/mg). Von jeder Proteinprobe wurden zwei Replikate gemessen. Auf den Substraten Arabinogalactan, Glucomannan, Galactomannan, Inulin, Laminarin, Lichenan, Pachyman, Pectin, Pullulan, Polygalacturonsäure, Xyloglucan war keines der Enzyme aktiv. 103
Tabelle 58: Liste aller mesophilen Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten Sequenzähnlichkeit (laut NCBI Datenbankvergleich). [1]Inc. sed. = incertae sedis 106
Tabelle 60: Liste der zur genaueren Charakterisierung ausgewählten mesophile Isolate. BGA = Biogasanlage 107
Tabelle 61: Wichtige morphologische und physiologische Eigenschaften von den Isolaten 37hoe/3, Iso2/2 37 107
Tabelle 62: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat 37hoe/3 und C. herbivorans DSM 14428 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten 108
Tabelle 63: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat Iso2/2 37 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten 109
Tabelle 64: Eigenschaften des von Isolat Iso2/2 37 sequenzierten Genoms 110
Tabelle 65: Kohlenhydrat aktive Enzyme die im Genom von Isolat Iso2/2 37 codiert sind; CBM: Kohlenhydratbindemodule, CE: Kohlenhydratesterasen, GH: Glycosidhydrolasen, GT: Glycosidtransferasen 111
Tabelle 66: Zusammenfassung und Ergebnis aller Medien die für die Gewinnung neuer Isolate aus den „in sacco“ Anreicherungen verwendet wurden. % Angaben in v/v. + Resultat positiv; - Resultat negativ 112
Tabelle 67: Liste aller „in sacco“ Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten Sequenzähnlichkeit (laut NCBI Datenbankvergleich). 112
Tabelle 68: Wichtige morphologische und physiologische Eigenschaften von Isolat GGR1. *Filterpapier (Whatman No. 1) 113
Tabelle 69: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat GGR1 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten 114
Tabelle 70: Bester NCBI Datenbanktreffer für kultivierte Bakterien der klonierten 16SrRNA Gene einer DNA Probe des „in sacco“ Ansatzes: BM1T Entnahmetag 7. Sequenzen die über 98 % Übereinstimmung aufweisen sind in gleicher Farbe hinterlegt. 116
Tabelle 71: Bester NCBI Datenbanktreffer für kultivierte Bakterien der klonierten 16SrRNA Gene einer DNA Probe des „in sacco“ Ansatzes: BM1T Entnahmetag 7. Sequenzen die über 98 % Übereinstimmung aufweisen, sind in gleicher Farbe hinterlegt. 118
Tabelle 72: Spez. Akt. (U/mg) der Zellüberstände ausgewählter Isolate auf PASC (Phosphoric acid swollen cellulose) 136

IV. ABBKÜRZUNGSVERZICHNIS

APS	Ammoniumpersulfat
ATCC	*American Type Culture Collection*
Bp	Basenpaare
BGA	Biogasanlage
°C	Grad Celsius
CBM	*carbohydrate binding module*
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonat
Cfu	Colony forming units (Lebendkeimzahl)
CipA	Scaffoldin Protein A
CMC	Carboxymethylcellulose
Cst	*Clostridium stercorarium*
Cth	*Clostridium thermocellum*
DNA	*Deoxyribonucleic acid*
ddH_2O	deionisiertes Wasser oder auch demineralisiertes Wasser
dest.	destilliert
g	Gramm
GH	Glycosidhydrolase
h	*hour*
H_2O	Wasser
kbp	Kilobasenpaare
kDa	Kilodalton
l	Liter
M	Molar
mM	Millimolar
ml	Milliliter
Min	Minuten
MALDI-TOF MS	*matrix-assisted laser desorption/ionization–time of flight/mass spectrometry*
nm	Nanometer
PCR	*polymerase chain reaction*
RT	Raumtemperatur
s	Sekunden
T	Temperatur
t	*time*
TEMED	N,N,N',N'-Tetramethylethylendiamin
U	*unit*
ÜN	über Nacht
vgl.	vergleiche
v/v	Volumenprozent
w/v	Massenprozent

V. ZUSAMMENFASSUNG

Nur ein kleiner Teil der in Biogasanlagen vorhandenen hydrolytischen, insbesondere Cellulose abbauenden Bakterien wurde bisher in Reinkultur isoliert. Davon ist *Clostridium thermocellum* eines der bekanntesten und am weitesten verbreiteten anaeroben cellulolytischen Bakterien. In dieser Studie wurden erstmals *C. thermocellum* Stämme aus verschiedenen Biogasanlagen isoliert und die genetische und phänotypische Variation zwischen den Stämmen bestimmt. Ferner wurden verschiedene Genotypisierungsmethoden für *C. thermocellum* Isolate evaluiert. Ein neuer Genotypisierungsansatz, die Amplifikation einer unbekannten Zielsequenz zwischen den Kopien eines mobilen Elements (IS1447), war das Genotypisierungsverfahren mit der höchsten Trennschärfe, bei dem jede Biogasanlage ein unterschiedliches Bandenmuster ergab. Unabhängig davon, welche Methode zur Genotypisierung gewählt wurde, zeigten die Isolate von jeder einzelnen Biogasanlage immer ein identisches Bandenmuster. Es scheint daher in jeder Biogasanlage ein optimal angepasster Stamm von *C. thermocellum* zu dominieren. Die physiologische Charakterisierung der Isolate spiegelt diese Ergebnisse wieder. In den meisten Fällen konnten zwei unabhängig gewonnene Isolate aus einer Biogasanlage die gleichen Substrate verstoffwechseln und diese zeigten in der Regel ähnliche spezifische Cellulaseaktivitäten. Auf der anderen Seite unterscheiden sich Stämme aus verschiedenen Biogasanlagen in ihren Genotypisierungsmuster und physiologischen Eigenschaften.

Das Hauptziel dieser Arbeit war die Isolierung neuer cellulolytischer Bakterien aus je einer definierten thermophilen und mesophilen Biogasanlage, um die „Black Box“ der Biogasmikrobiologie besser auszuleuchten. Dazu wurden verschiedene Strategien und Techniken zur Isolierung entwickelt und etabliert. Hierzu wurde auch die sogenannte „in sacco“ Methode, zur direkten Anreicherung von cellulolytischen Bakterien in Laborfermentern entwickelt. Zusätzlich wurde eine Metagenom-Analyse von einer Probe aus dem thermophilen Reaktor durchgeführt, welche zeigte, dass die bakteriellen Phyla Firmicutes, Thermotogae und Bacteroidetes diese Biogas-Mikrobengemeinschaft dominieren. Cellulolytische Bakterien konnten jedoch nur aus dem Phylum Firmicutes isoliert werden. Darunter waren einige mit bereits bekannten cellulolytischen Arten verwandte Isolate, wie *Clostridium cellulosi, Clostridium stercorarium, Clostridium clariflavum* und *Clostridium thermocellum*. Auch einige bis dahin unkultivierte, mesophile und thermophile cellulolytische Bakterien aus bisher unbekannten Linien der Familien *Lachnospiraceae* (SD1D, T3/55, 37hoe/1) oder *Ruminococcaceae* (GGR1) wurden isoliert. Besonders solche neuen Isolate liefern neue Genomdaten und damit den Zugang zu neuartigen Lignocellulose-abbauenden Enzymen für verschiedene industrielle Anwendungen. Ein Isolat aus jeder bisher unbekannten Gruppe wurde für weitere physiologische Untersuchungen, Genom-Sequenzierung und Charakterisierung der physiologischen Eigenschaften und der cellulolytischen und hemicellulolytischen Systeme ausgewählt.

VI. ABSTRACT

Only a small fraction of hydrolytic - especially cellulose-degrading - bacterial species present in a biogas fermenter has been isolated in pure culture. One of them is *Clostridium thermocellum,* the most prevalent anaerobic cellulolytic bacterium known so far. In this study, the genetic and phenotypic variations between *C. thermocellum* strains isolated from different biogas plants have been determined. Furthermore different genotyping methods for *C. thermocellum* isolates have been evaluated. One novel genotyping approach, the amplification of unknown target sequences between copies of an earlier discovered randomly inserted mobile element (IS1447), was the genotyping method with the highest discriminatory power, yielding a different pattern of bands for every biogas plant. Irrespective of the genotyping method the isolates from each individual biogas plant always showed an identical band pattern. Hence, it appears that one strain of *C. thermocellum* is dominating in each biogas plant. The genotypic groups reflect the results of the physiological characterization of the isolates. In most cases both strains of one biogas plant could grow on the same substrates and showed similar specific cellulase activity. On the other hand, strains isolated from different biogas plants differed in their genotyping results and physiological properties.

The second and major aim of this work was the isolation of new cellulolytic bacteria from sludge of defined thermophilic and mesophilic biogas plants in order to shed light into the black box of biogas microbiology. To this end different isolation strategies were examined, including one novel approach called “in sacco” for direct enrichment of cellulolytic bacteria in lab scale biogas plants. Additionally, metagenomic sequence analysis from a thermophilic reactor sample revealed that the bacterial phyla Firmicutes, Thermotogae and Bacteroidetes dominated the biogas community. Out of the phylum Firmicutes several isolates were retrieved, which were related to already characterized cellulolytic species like *Clostridium cellulosi, Clostridium stercorarium, Clostridium clariflavum* and *Clostridium thermocellum*. Also some previously uncultured cellulolytic mesophilic and thermophilic bacteria, forming hitherto unknown sublines within the families *Lachnospiraceae* (SD1D, T3/55, 37hoe/1) or *Ruminococcaceae* (GGR1) were isolated. These new isolates are pivotal to provide the access to genome data from novel cellulolytic bacteria as well as new lignocellulosic biomass-degrading enzymes for various industrial applications. One isolate of each group was chosen for further investigation such as genome sequencing, and characterization of their physiological traits and their cellulolytic and hemicellulolytic enzyme systems.

VII. PUBLIKATIONEN

Koeck, D. E., Maus, I., Wibberg, D., Winkler, A., Zverlov, V. V., Liebl, W., Puhler, A., Schwarz, W. H., Schlüter, **A.** (2015) Draft genome sequence of *Herbinix hemicellulosilytica* T3/55^T, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor. J. Biotechnol. 214: 59–60

Koeck, D. E., Koellmeier, T., Schiffner, S., Zverlov, V. V., Liebl, W., Schwarz, W. H. (2015). Differences in biomass degradation between newly isolated environmental strains of *Clostridium thermocellum* and heterogeneity in the size of the cellulosomal scaffoldin. Systematic and Applied Microbiology. 38: 424-432

Koeck, D.E., Ludwig, W., Wanner, G., Zverlov, V. V., Liebl, W., Schwarz, W.H. (2015). *Herbinix hemicellulosilytica*, gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. International Journal of Systematic and Evolutionary Microbiology. 65: 2365-2371

Zverlov, V. V., **Koeck, D. E**., Schwarz, W. H. (2015). The Role of Cellulose-Hydrolyzing Bacteria in the Production of Biogas from Plant Biomass. In B. Kamm (ed.), Microorganisms in Biorefineries, Microbiology Monographs 26: 335-361

Koeck, D.E., Zverlov, V.V., Liebl, W,, Schwarz, W.H. (2014). Comparative genotyping of *Clostridium thermocellum* strains isolated from biogas plants: genetic markers and characterization of cellulolytic potential. Syst Appl Microbiol. 37: 311-319.

Koeck, D.E., Pechtl, A., Zverlov, V.V., Schwarz, W.H. (2014). Genomics of cellulolytic bacteria. Curr Opin Biotechnol. 29: 171-183.

Koeck, D.E., Wibberg D., Maus I., Winkler A., Albersmeier, A., Zverlov V.V., Liebl W., Pühler A., Schwarz W.H., Schlüter A. (2014). Complete genome sequence of the cellulolytic thermophile *Ruminoclostridium cellulosi* wild-type strain DG5 isolated from a thermophilic biogas plant. J Biotechnol. 188: 136-137.

Koeck, D.E., Wibberg D., Maus I., Winkler A., Albersmeier, A., Zverlov V.V., Liebl W., Pühler A., Schwarz W.H., Schlüter A. (2014). First draft genome sequence of the amylolytic *Bacillus thermoamylovorans* wild-type strain 1A1 isolated from a thermophilic biogas plant. J Biotechnol. 192: 154-155

Koeck, D.E., Wibberg, D., Köllmeier, T., Blom, J., Jaenicke, S., Winkler, A., Albersmeier, A., Zverlov, V. V., Pühler, A., Schwarz, W. H., Schlüter, A. (2013). Draft genome sequence of the cellulolytic *Clostridium thermocellum* wild-type strain BC1 playing a role in cellulosic biomass degradation. Journal of Biotechnology 168: 62-63.

Zverlov, V. V., Hiegl, W., **Koeck, D.E.,** Kellermann, J., Köllmeier, T., Schwarz, W. H. (2010) Hydrolytic bacteria in mesophilic and thermophilic degradation of plant biomass. Engineering in Life Sciences 10: 528-536.

Publikationen in Vorbereitung

Koeck, D.E., Hahnke, S., Mechelke, M., Liebl, W., Zverlov, V.V., Schwarz, W.H. *Herbivorax saccocumulans* sp. nov., a cellulolytic, anaerobic bacterium isolated via in sacco enrichments from a lab scale biogas reactor. (in Vorbereitung)

Koeck, D.E., Hahnke, S., Zverlov, V.V. *Herbinix treslupi* sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. (in Vorbereitung)

Koeck, D.E., Maus, I., Wibberg, D., Winkler, A., Zverlov, V.V., Liebl, W., Pühler, A., Schwarz, W.H., Schlüter, A.. Draft genome sequence of *Herbinix* sp. SD1D, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor. (in Vorbereitung)

Koeck, D.E., Maus, I., Wibberg, D., Winkler, A., Zverlov, V.V., Liebl, W., Pühler, A., Schwarz, W.H., Schlüter, A.. Draft genome sequence of Isolate 2/2 37, a new mesophilic xylan degrading bacterium isolated from a mesophilic biogas reactor (In Vorbereitung)

Konferenzbeiträge

Koeck, D.E., Zverlov, V.V., Liebl, W., Schwarz, W.H. (2015). Isolation of novel cellulolytic bacteria from biogas plants and characterization of their cellulolytic and hemicellulolytic potential. Symposium on Biotechnology for Fuels and Chemicals. San Diego. USA

Koeck, D.E., Zverlov, V.V., Liebl, W., Schwarz, W.H. (2014). Isolation of cellulolytic bacteria from biogas plants and comparative genotyping of *Clostridium thermocellum* strains isolated: genetic markers and characterization of cellulolytic potential. Conference on Biogas Microbiology. Uppsala. Sweden

Schwarz, W. H., Zverlov, V. V., **Koeck, D. E.** (2014). Cellulolytic bacteria in industrial biogas-plants. Mie Bioforum in Nemuno Sato, Japan (eingeladener Vortrag)

Koeck, D.E., Zverlov, V.V., Liebl, W., Schwarz, W.H. (2015) Isolation of novel cellulolytic bacteria from biogas plants and characterization of their cellulolytic and hemicellulolytic potential. Gordon Research Conference "Cellulosomes, Cellulases & Other Carbohydrate Modifying Enzymes"(Poster)

1 EINLEITUNG

1.1 BIOGAS ALS REGENERATIVE ENERGIEQUELLE

Heutzutage werden immer noch etwa 80 % des globalen Energiebedarfs aus der Verbrennung fossiler Energieträger gewonnen. Der weltweite Energiebedarf wird in diesem Jahrhundert noch um das Zwei- bis Dreifache ansteigen. Ohne Gegenmaßnahmen wird dies in Zukunft durch die Anreicherung von CO_2 in der Atmosphäre zu drastischen klimatischen Veränderungen führen. In diesem Kontext wächst das Interesse an der Bereitstellung von Energie aus nachwachsenden Rohstoffen, die fossile Brennstoffe zum größten Teil ersetzen können (Kazamia and Smith 2014). Biogas-Erzeugung aus nachwachsenden Rohstoffen ist derzeit die attraktivste Technologie für eine nachhaltige und Kohlenstoffdioxid-neutrale Energieproduktion mit dem höchsten Energieertrag pro Hektar verglichen mit anderen Pflanzen basierten Bioenergiequellen (Antoni et al. 2007, Weiland 2006). Auch die energie- und umweltpolitischen Diskussionen in Deutschland werden maßgeblich durch die energiebedingten Umweltauswirkungen mitbestimmt [Fachagentur für nachwachsende Rohstoffe (FNR) 2013]. Seit Inkrafttreten des Erneuerbare-Energien-Gesetz (EEG) im Jahr 2000 hat die Erzeugung und Nutzung von Biogas erheblich zugenommen (Abbildung 1).

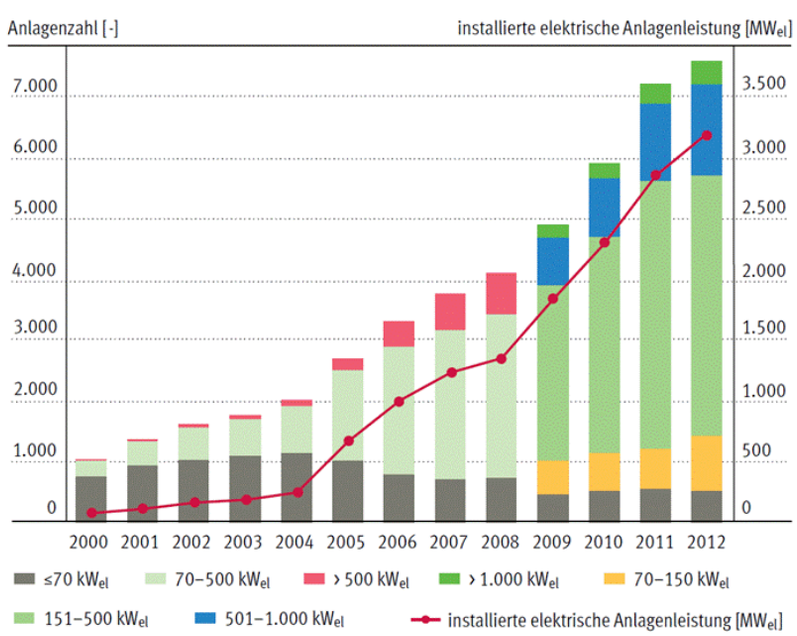


Abbildung 1: Entwicklung der Biogasanlagen in Deutschland - Anlagenzahl differenziert nach Leistungsklassen und installierter elektrischer Anlagenleistung MW_{el} ohne Biogasaufbereitungsanlagen, Deponie- und Klärgasanlagen (FNR 2013)

Das primäre Ziel von Biogasanlagen (BGA) ist es, organische Substanz, darunter zunehmend Cellulose-haltige Biomasse, unter anaeroben Bedingungen mikrobiell zu einem überwiegend methanhaltigen Gas abzubauen. Dieser natürliche Abbauprozess ist weitverbreitet und findet z.B. in Flüssen, Seen, Sumpfen, Mooren, schlecht durchlüfteten Böden und Klärgruben statt. Abhängig von der Substratzusammensetzung und dem Fermentationsverfahren besteht das gebildete Gasgemisch überwiegend aus Methan (50-75 Vol-%) und Kohlendioxid (FNR 2013). Aus diesem im Biogas enthaltenen Methan wird durch kontrollierte Verbrennung in Blockheizkraftwerken (BHKWs) elektrische und thermische Energie erzeugt (Görisch and Helm 2006). Der gesamte Abbauprozess im Biogasfermenter besteht im Wesentlichen aus vier Stufen, bei denen jeweils unterschiedliche Bakteriengruppen aktiv sind (Abbildung 2).

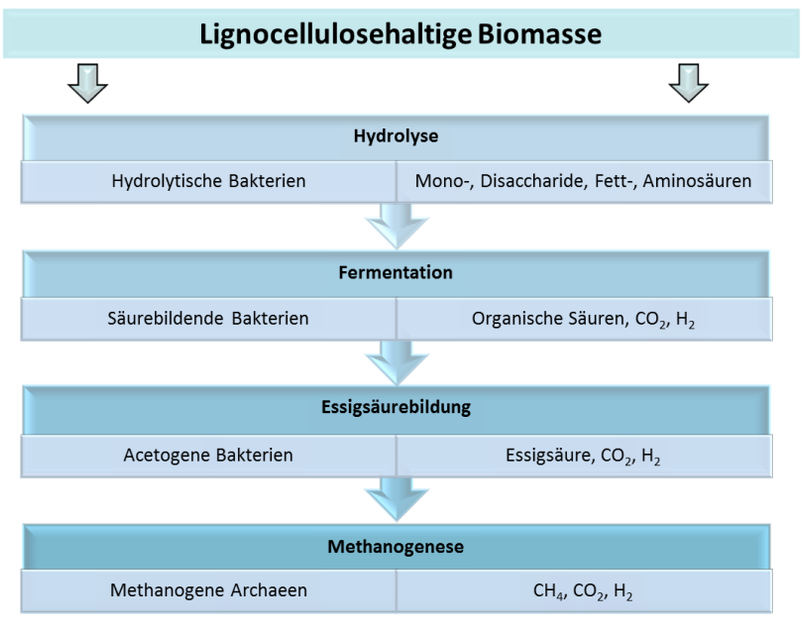


Abbildung 2: Schematische Darstellung des anaeroben Abbaus von lignocellulosehaltiger Biomasse bis zu Methan

In der ersten Stufe der **„Hydrolyse“**, bauen in erster Linie hydrolytische Bakterien die eingespeisten, hochmolekularen, organischen Subtanzen (Proteine, Kohlenhydrate, Fette) mit Hilfe hydrolytischer Exoenzyme in niedermolekulare Aminosäuren, Zucker und Fettsäuren ab. Die Geschwindigkeit der Hydrolyse ist dabei von der Aktivität der extrazellulären Enzyme wie Cellulasen, Hemicellulasen, Pectinasen, Amylasen, Proteasen, Xylanasen oder Lipasen abhängig (Eder and Schulz 2007). Bei mikrobiell schwer abbaubaren Stoffen (wie Cellulose) ist dieser Schritt im Gesamtprozess geschwindigkeitsbestimmend und daher für die Effizienz von Substratabbau und Biogasbildung von ganz besonderer Bedeutung; Clostridien sind in dieser ersten Stufe dann die dominierenden Organismen (Görisch and Helm 2006). Sie verwenden die gebildeten Stoffe zum Teil selbst und stellen sie zum Teil anderen Mikroorganismen zur Verfügung.

Die durch die hydrolytischen Bakterien gebildeten Zwischenprodukte werden in der zweiten Phase, der sogenannten **„Acidogenese“** oder Fermentation, durch säurebildende Mikroorganismen in kurzkettige Carbonsäuren (z.B. Butter-, Essig-, Ameisen-, Propionsäure), niedermolekulare Alkohole, Wasserstoff, CO_2 und Wasser umgewandelt. Dabei wird der pH-Wert abgesenkt und der noch vorhandene Sauerstoff verbraucht. Dadurch werden die anaeroben Bedingungen für die methanogenen Archaeen sichergestellt (FNR 2013).

Durch weitere biochemische Umsetzungen produzieren acetogene Bakterien aus den genannten Stoffen hauptsächlich Essigsäure, Wasserstoff und Kohlenstoffdioxid als Ausgangsprodukte für die anschließende Methanbildung. Dieser Vorgang wird auch als **„Acetogenese“** bezeichnet. Diese Wasserstoff bildenden Bakterien sind sehr temperaturempfindlich (Eder and Schulz 2007). Der Abbau von Butter- und Propionsäure zu Essigsäure ist dabei der geschwindigkeitsbestimmende und thermodynamisch schwierigste Schritt des Gesamtabbaus. Ein zu hoher Wasserstoffpartialdruck verhindert aus thermodynamischen Gründen die Umsetzung und führt zur Anreicherung der organischen Säuren, wodurch wiederum die Methanbildung gehemmt wird (Stams et al. 1993).

Im letzten Schritt der Biogasbildung, der **„Methanogenese“,** werden vor allem Essigsäure sowie Wasserstoff und Kohlendioxid von strikt anaeroben methanogenen Archaeen zu Methan umgewandelt. Man unterscheidet zwischen hydrogenotrophen Methanogenen, welche aus Wasserstoff und Kohlendioxid Methan produzieren, und acetoclastischen Methanogenen, welche durch Essigsäurespaltung Methan bilden. Unter den in landwirtschaftlichen Biogasanlagen vorherrschenden Bedingungen erfolgt die Methanbildung bei höheren Raumbelastungen vorwiegend über den Wasserstoff verwertenden Reaktionsweg und nur bei relativ geringer Raumbelastung über den Essigsäure spaltenden Reaktionsweg (Bauer et al. 2008).

Grundsätzlich finden die vier Phasen in einstufigen Reaktoren parallel statt. Die Abbaugeschwindigkeit der Substrate hat direkten Einfluss auf die notwendige Verweilzeit im Biogasfermenter. Beim sogenannten Batchverfahren wird ein Fermenter neu mit Substraten beschickt; dabei nimmt nach Durchlauf der einzelnen Phasen die Biogasmenge bis zu einem Scheitelpunkt zu und sinkt danach wieder, wenn nur noch die schwer abbaubaren Stoffe vergoren werden können, bis das Substrat nicht weiter abbaubar ist. Beim kontinuierlichen Verfahren kommt es durch laufende Zugabe von Substrat mehrmals täglich zu einer gleichmäßig hohen Gasbildung (FNR 2013).

Die Geschwindigkeit des Faulprozesses ist sehr stark temperaturabhängig. Je höher die Temperatur ist, umso schneller erfolgt der Abbau und umso höher ist die Gasproduktion. Aber mit steigender Temperatur sinkt der Methangehalt im Biogas, da bei hohen Temperaturen das im Substrat gelöste Kohlendioxid verstärkt in die gasförmige Phase übergeht, sodass der relative Anteil des Methans abnimmt. Die erzielbaren Gasausbeuten sind bei ausreichender Verweilzeit

gleich hoch. Je höher die Temperatur, desto empfindlicher sind die Bakterien gegenüber Schwankungen. Die meisten Anlagen werden daher im mesophilen Temperaturbereich zwischen 38 und 42 °C betrieben. Anlagen im thermophilen Bereich (55 °C) erfordern eine genauere Betriebsführung und –steuerung (Weiland 2001).

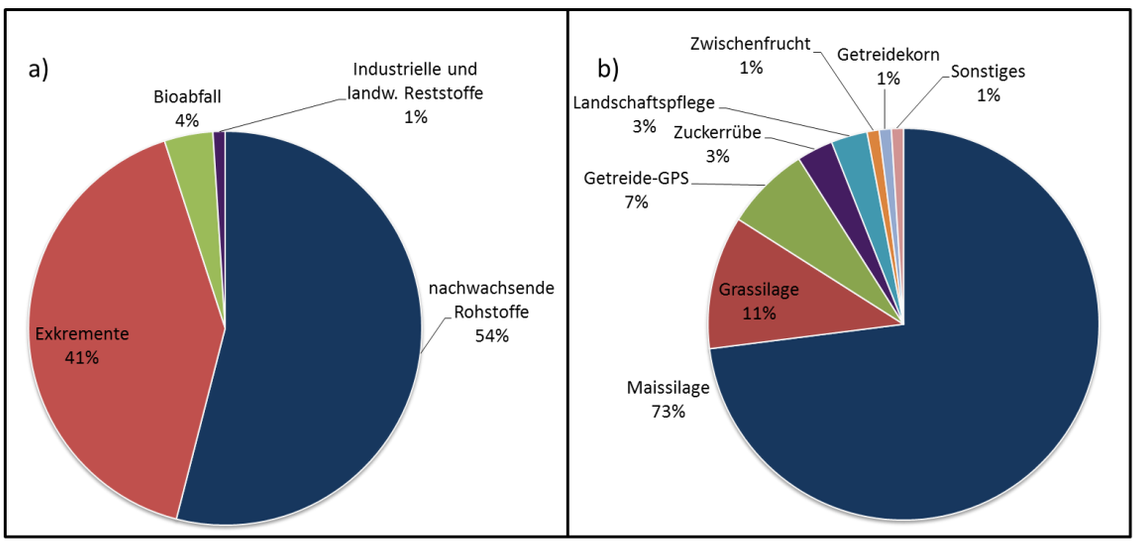


Abbildung 3: a) Massebezogener Substrateinsatz in Biogasanlagen b) Massebezogener Substrateinsatz nachwachsender Rohstoffe in Biogasanlagen; Daten aus FNR Bericht (FNR 2013)

Vor allem bei der Wahl des Substrats gibt es aber nach wie vor ungenutzte Potentiale, die zur Biogaserzeugung herangezogen werden können. In Deutschland werden als Basissubstrate vor allem Exkremente und nachwachsende Rohstoffe (NawaRo) eingesetzt (Abbildung 3a). Für Biogasanlagen relevante NawaRo wie Mais- oder Grassilage, liefern einen beträchtlichen Gasertrag (Abbildung 3b). Aber vor allem Substrate, für die keine zusätzlichen Aufwendungen erforderlich sind, sind ökologisch von Vorteil (FNR 2013). Biogasanlagen können mit fast jeder Form von Biomasse betrieben werden, solange das Material nicht einen zu hohen Anteil an Lignin besitzt, wie es beispielsweise bei Holz der Fall ist. Biogasanlagen bieten eine profitable Möglichkeit zur Nutzung von erneuerbaren, nachwachsenden, örtlich verfügbaren Rohstoffen unter Verwendung bisher ungenutzter oder speziell angebauter Pflanzen und Pflanzenteile. Statt des gezielten Anbaus von Energiepflanzen sollte vermehrt auf Reststoffe und Rückstände aus Landwirtschaft und Nahrungsmittelindustrie zurückgegriffen werden. Die Bereitstellung von Biogas aus Lignocellulose haltigen organischen Abfällen, also aus Bioenergieträgern der 2. Generation, ist dabei besonders ökonomisch und umweltfreundlich (Divya et al. 2015).

1.2 Lignocellulose als Substrat in Biogasanlagen

Lignocellulose bildet die Zellwand verholzter Pflanzen und ist ein Verbund aus Cellulose, Hemicellulose und Lignin (López Contreras 2003). Cellulose ist dabei der Hauptbestandteil von pflanzlichen Zellwänden (Massenanteil bis zu 50% des Trockengewichts) und damit die häufigste

organische Verbindung auf der Erde. Das Polysaccharid Cellulose besteht aus mehreren hundert bis zehntausend β-D-Glucose-Molekülen, die 1-4-β-glycosidisch zu einem linearen Strang verbunden sind. Die Wasserstoffbrückenbindungen zwischen den Makromolekülen verhindern eine spiralartige Drehung des Moleküls, weswegen das Makromolekül der Cellulose im Gegensatz zu Stärke linear angeordnet ist. Mehrere solcher linearer Stränge lagern sich bedingt durch die Wasserstoffbrückenbindungen zu Mikrofibrillen zusammen, wobei sich kristalline Bereiche mit nichtkristallinen abwechseln (Klemm et al. 2005). Hemicellulose ist das zweithäufigste Polymer und ist im Gegensatz zu Cellulose chemisch sehr heterogen. Hemicellulosen sind verzweigte Polymere aus Pentosen (Xylose, Arabinose), Hexosen (Mannose, Glucose, Galactose) und acetylierten Zuckern. Während die Hemicellulose in Stroh überwiegend aus Xylan besteht, findet man in Weichhölzern (z.B. Lärche) hauptsächlich Glucomannan. Xylane wiederum sind ebenfalls verzweigte Heteropolysaccharide mit einem Rückgrat aus 1-4-verknüpften β-D-Xylopyranose Einheiten. Hemicellulosen sind über Wasserstoffbrücken mit den Mikrofibrillen der Cellulose verknüpft (Ratanakhanokchai et al. 2013). Lignin ist das dritt-abundanteste Polymer in der Natur, jedoch kein Polysaccharid. Es besteht aus einem komplexen Netzwerk aus polymerisierten Phenylpropaneinheiten. Die drei monomeren Bestandteile von Lignin sind p-Coumaryl-, Coniferyl-, und Sinapylalkohol. Lignin kann auch mit einigen Hemicellulosen kovalente Bindungen eingehen, z.B. Benzylesterbindungen mit der Carboxylgruppe der 4-*O*-methyl-D-Glucuronsäure in Xylan (Perez et al. 2002, Ratanakhanokchai et al. 2013).

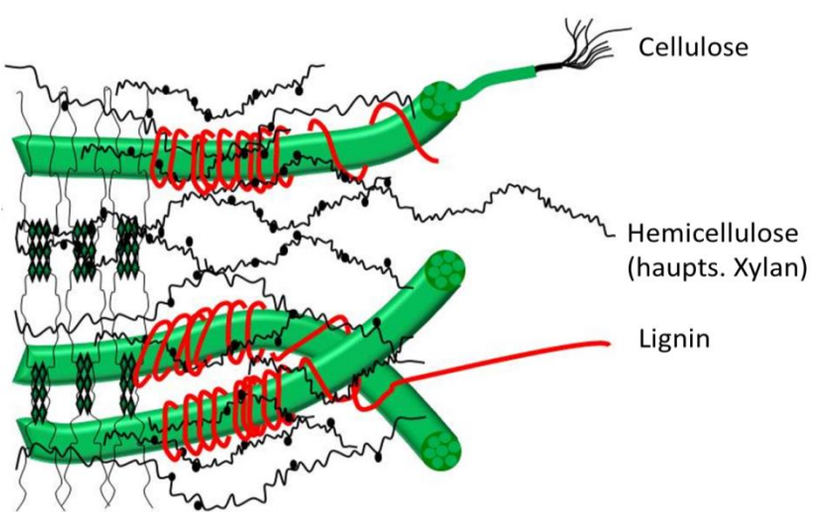


Abbildung 4: Struktur von Lignocellulose haltiger Pflanzenbiomasse. © 2013 Ratanakhanokchai K, Waeonukul R, Pason P, Tachaapaikoon C, Kyu KL, Sakka K, Kosugi A, Mori Y. veröffentlicht in Ratanakhanokchai et al. 2013 unter CC BY 3.0 Lizenz

Hemicellulose und Cellulose können durch anaerobe Mikroorganismen zu Biogas umgewandelt werden. Lignin kann von den anaeroben Bakterien jedoch nicht umgesetzt werden. Deshalb verbleiben in der Praxis bei der Vergärung lignocellulosehaltiger Rohstoffe größere Mengen nicht umgesetzter faserhaltiger Gärreste.

1.3 Cellulolytische Bakterien

Cellulose ist aufgrund der oben beschriebenen kristallinen Struktur sehr widerstandsfähig gegenüber Einzelenzymen. Der Abbau erfordert daher spezielle Enzymsysteme, die nur von einer limitierten Anzahl spezialisierter Mikroorgansimen produziert werden (Boone et al. 1993, Lynd et al. 2002, Noike et al. 1985). Bisher wurde nur eine relativ begrenzte Anzahl cellulolytischer Bakterien isoliert und charakterisiert (Schwarz et al. 2004). Eine vollständige Liste (Stand 2014) von Bakterien, die Cellulose in ihrer natürlichen Konfiguration (kristallin) als Kohlenstoff- und Energiequelle nutzen können, wurde kürzlich publiziert (Koeck et al. 2014a). Die meisten der bisher bekannten Bakterien, die effizient natürliche Cellulose abbauen, gehören zum Phylum Firmicutes bzw. zur Klasse der Clostridia.

Etwa die Hälfte der Bakterien mit Genen für Cellulasen und Hemicellulasen sind Saphrophyten, welche effizient abgestorbene Pflanzenreste mineralisieren (Medie et al. 2012), wovon aber nur ein geringer Prozentsatz mehr als zwei Gene für ß-1,4-Glucanasen (Cellulasen) besitzt. Bakterien werden oft in der Literatur schon als cellulolytisch bezeichnet, wenn sie extrazelluläre ß-Glucanasen produzieren, welche die ß-1,4-glycosidische Bindung in löslichen ß-Glucanen oder in künstlichen Cellulose ähnlichen Substraten, z.B. Carboxymethylcellulose (CMC), spalten können. Obwohl CMC aufgrund der ß-1,4-glucosidischen Verknüpfung chemisch zur Cellulose gezählt wird, ist dieses Substrat aufgrund der Substituenten im Gegensatz zu „natürlicher“ Cellulose nicht kristallin und in Wasser löslich. Im Gegensatz zu den echt cellulolytischen Bakterien sind diese sogenannten „cellulolytischen“ Bakterien nicht in der Lage kristalline Cellulose abzubauen und zu verwerten. Solche einfachen Cellulase-Systeme findet man in Pflanzenpathogenen oder Bakterien, die Cellulose synthetisieren (Medie et al. 2012).

Echte cellulolytische Mikroorganismen sind von großer Bedeutung für die Produktion von Energie und Gütern aus regenerativen Quellen: Sie bauen Pflanzenmaterial ab, produzieren eine große Bandbreite an wertvollen Ausgangsstoffen für die chemische Industrie und sie machen die genetische Information für die Produktion von cellulolytischen und hemicellulolytischen Enzymen zugänglich, welche wiederum für eine Vielzahl von industriellen und biotechnologischen Prozessen genutzt werden können (Schwarz et al. 2011).

1.4 Mechanismen und Enzyme zum Abbau von Cellulose

Aufgrund der variablen und oft unlöslichen (kristallinen) Struktur von Polysacchariden in Pflanzenbiomasse ist eine große Anzahl von Enzymen mit unterschiedlicher Substratspezifität nötig, um diese Art Substrat abzubauen (Claassen et al. 1999). Diese Gruppen von Enzymen bekannten Ursprungs sind in der CAZy Datenbank gesammelt (Lombard et al. 2014) und in funktionelle Gruppen eingeteilt, darunter z.B. Glycosidhydrolasen (GHs), Polysaccharidlyasen

(PLs), Kohlenhydratesterasen (CEs) und Kohlenhydratbindemodule (CBMs) (Henrissat and Davies 1997). Dabei sind die GHs katalytische Module von Enzymen, welche die glycosidische Bindung zwischen zwei Zuckern oder zwischen Zuckern und Nicht-Kohlenhydraten spalten (Henrissat and Davies 1997). Die CAZy Datenbank teilt diese auf Basis ihrer Aminosäuresequenz und den daraus resultierenden Enzymen in aktuell 133 Familien ein (Cantarel et al. 2009, Lombard et al. 2014). Die Zugehörigkeit eines Enzyms zu einer GH Familie liefert daher Einblicke in dessen strukturelle Merkmale, evolutionäre Verwandtschaft zu anderen Familienmitgliedern und den enzymatischen Mechanismus (Himmel et al. 2010).

Cellulasen sind wie oben bereits erwähnt Enzyme, die ß-1,4-glycosidische Bindungen durch einen hydrolytischen Mechanismus spalten. Hierbei wird die Konfiguration am anomeren C-Atom entweder invertiert oder beibehalten (Davies and Henrissat 1995). Bisher gibt es 6 GH-Familien (8 mit den Cellobiose spaltenden ß-Glucosidasen), die bekanntermaßen Cellulose hydrolysieren. Enzyme, die an der Cellulosehydrolyse von anaeroben oder fakultativ anaeroben Bakterien beteiligt sind, beinhalten Module der GH Familien 5, 6, 8, 9, 48 und 124 (Tabelle 1).

Tabelle 1: Liste der GHs (Glycosidhydrolasen) die in cellulolytischen Enzymen enthalten sind, basierend auf der CaZy Datenbank (Lombard et al. 2014)). Prozessive Exoglucanansen in kursiv

GH5	Endo-ß-1,4-Glucanase / Cellulase (EC 3.2.1.4), *cellulose ß-1,4-Cellobiosidase (EC 3.2.1.91), Exo-ß-1,4-Glucanase / Cellodextrinase (EC 3.2.1.74),* ß-Glucosidase (EC 3.2.1.21),
GH6	Endo-Glucanase (EC 3.2.1.4), *Cellobiohydrolase* (EC 3.2.1.91)
GH8	Cellulase (EC 3.2.1.4),
GH9	Endo-Glucanase (EC 3.2.1.4), ß-Glucosidase (EC 3.2.1.21), , *Exo-ß-1,4-Glucanase / Cellodextrinase (EC 3.2.1.74), Cellobiohydrolase (EC 3.2.1.91),*
GH48	*Cellobiohydrolase* (EC 3.2.1.176), Endo-ß-1,4-Glucanase (EC 3.2.1.4),
GH124	Endo-Glucanase (EC 3.2.1.4)

Neben der Enzymspezifität unterscheidet man außerdem zwischen drei verschiedenen Arten der Enzymaktivität: Enzyme mit Endo-Funktion (Endo-Glucanase) schneiden zufällig innerhalb der ß-1,4-Glucan-Moleküle, prozessive Exo-Glucanasen erkennen entweder das reduzierende oder das nicht reduzierende Ende des Cellulosestrangs. Prozessive Endo-Glucanasen spalten ebenfalls innerhalb des Moleküls, setzen an den neu entstandenen Enden jedoch prozessiv Cellobiose- oder Cellotetraose-Einheiten frei (Schwarz 2001). Die verschiedenen Typen von Cellulasen arbeiten im Synergismus bei der vollständigen Hydrolyse kristalliner Cellulose-Mikrofibrillen zu Cellobiose (Morais et al. 2012). Beta-Glucosidasen hydrolysieren anschließend die entstandene Cellobiose oder –tetraose zu Glucose (Abbildung 5). Im aeroben Bereich wird bei einigen Bakteriengruppen der Abbau zusätzlich durch die oxidative Spaltung der ß-1,4-glycosidische Bindungen mittels LPMOs (lytic polysaccharide monooxygenases) unterstützt (Horn et al. 2012).

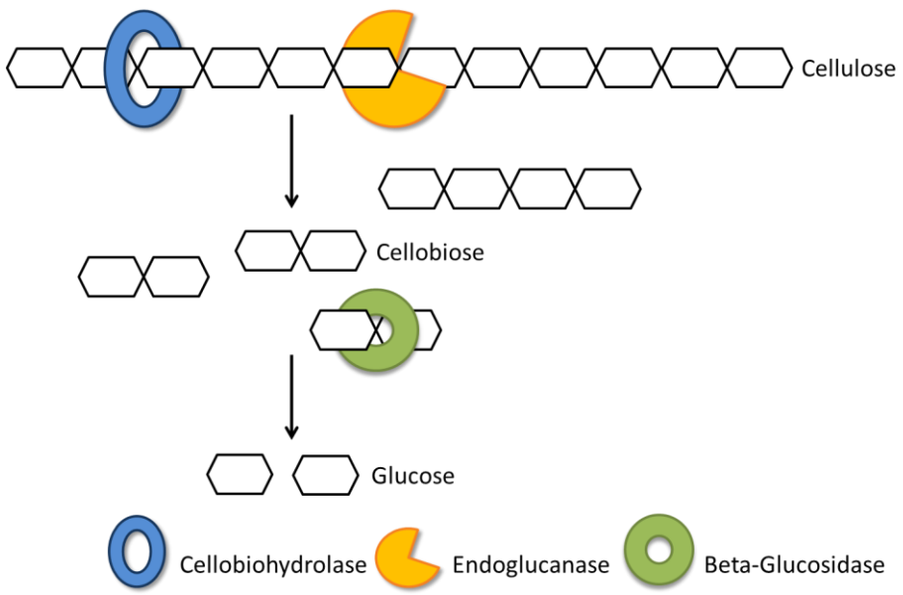


Abbildung 5: Klassisches Model für die anaerobe Hydrolyse von Cellulose. Endo-β-1,4-Glucanasen schneiden zufällig innerhalb des Cellulosestrangs. Cellobiohydrolasen (prozessive Exo-Glucanasen) setzen an den Enden prozessiv Cellobiose frei. Betaglucosidase scheidet anschließend die Cellobiose Einheiten. Entnommen aus Koeck et al (Koeck et al. 2014a). Modifiziert nach Watanabe und Tokuda (Watanabe and Tokuda 2010)

Die GH-Familien cellulolytischer Enzyme beinhalten nicht-prozessive Endo-Glucanasen, prozessive Exo-Glucanansen (Cellobiohydrolasen) oder prozessive Endo-Glucanasen (Zverlov and Schwarz 2008). Die Topologie des aktiven Zentrums unterscheidet dabei sich zwischen Endo- und Exo-Glucanasen. Endo-Glucanasen besitzen im aktiven Zentrum typischerweise eine spaltähnliche Topologie, welche es dem Enzym erlaubt, an irgendeiner Position entlang des Cellulosemoleküls anzugreifen. Die Form des aktiven Zentrums einer Exo-Glucanase ähnelt eher einem Tunnel, welcher durch lange Schleifen des Proteins geformt wird. Hierdurch wird erreicht, dass Cellodextrine einer definierten Größe am Ende des Substratstrangs abgespalten werden (Davies and Henrissat 1995).

Cellulasen und Hemicellulasen haben in der Regel einen modularen Aufbau, bei dem ein katalytisches Modul mit einem weiteren katalytischen oder nicht-katalytischen Modul/en, wie z.B. CBMs, kombiniert wird (Zverlov and Schwarz 2008). CBMs haben eine Affinität für kristalline und amorphe Cellulose und andere polymere Kohlenhydrate und bringen so das katalytische Modul in räumliche Nähe zum Substrat (Irwin et al. 1998). Einige CBMs haben eine einfache strukturelle Funktion und stabilisieren das katalytische Modul oder modifizieren seine Aktivität (Irwin et al. 1998).

Für den Abbau von kristalliner Cellulose ist mindestens eine Exo-Glucanase erforderlich. Bakterien mit der größten Anzahl an Cellulasen zeigen auch die höchste Aktivität auf kristalliner Cellulose (Himmel et al. 2010). Die meisten von ihnen gehören zum Phylum Firmicutes und besitzen ein Cellulosom.

1.5 Die Enzymsysteme cellulolytischer anaerober Bakterien

In der Natur wurden im anaeroben Bereich bisher drei dominante Typen mikrobieller Enzymsysteme zum Abbau von pflanzlichen Zellwänden gefunden. Dies sind entweder freie Einzelenzyme, multifunktionale Enzyme oder Multienzymkomplexe (Cellulosomen). Die Art des Systems kann bereits aus der Genominformation der betreffenden Bakterien abgeleitet werden. Sowohl cellulosomale als auch nicht-cellulosomale Enzyme können an der Zelloberfläche kovalent oder nicht kovalent gebunden sein (Himmel et al. 2010). In den nächsten Unterkapiteln werden die verschiedenen Systeme anhand von Beispielen genauer vorgestellt.

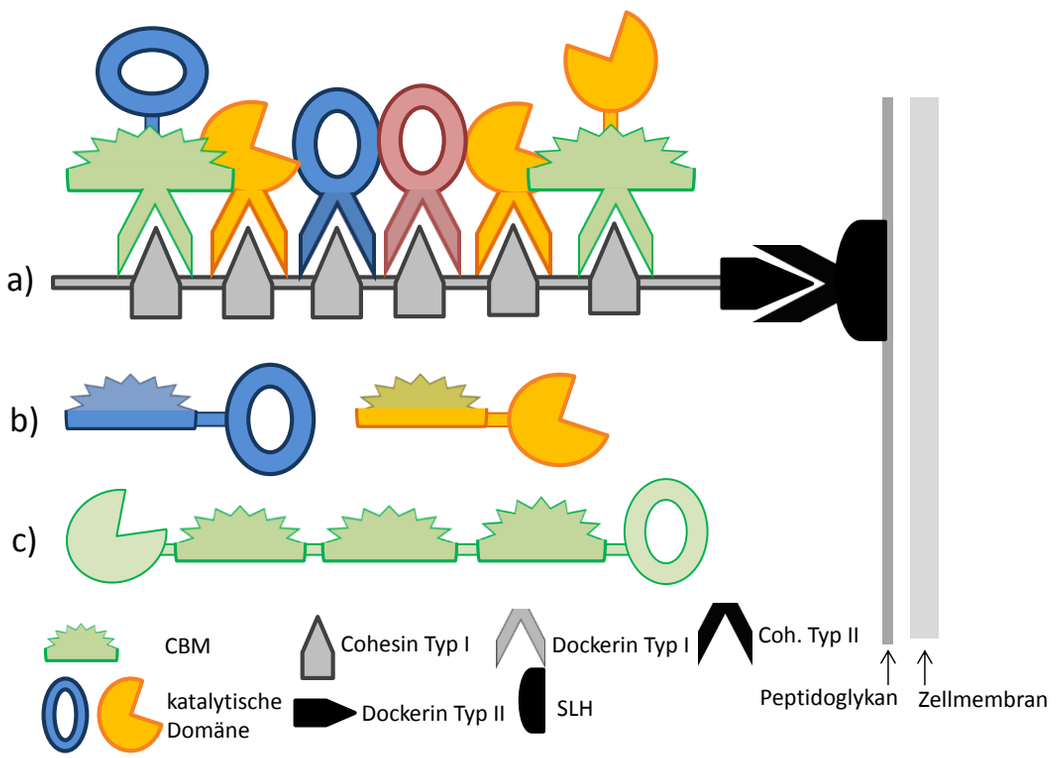


Abbildung 6: Enzymsysteme anaerober Bakterien für den Abbau lignocellulosehaltiger Biomasse. a) Cellulosom b) Einzelenzyme c) Multidomänenenzyme. CBM: Cellulose Bindungsmodul. SLH: S-layer Homologie. Gleiche Farben symbolisieren eine ursprüngliche Polypeptidkette

1.5.1 Das Cellulosom, ein effizienter Cellulase-Komplex

Das effizienteste Cellulasesystem, das bisher bekannt ist, stammt von *Clostridium thermocellum*. *C. thermocellum* kann fast überall nachgewiesen werden, wo anaerober Abbau von Biomasse stattfindet, z.B. in Kompost, Böden und Biogasanlagen (Ueno et al. 2001). Dieses anaerobe, thermophile Bakterium produziert einen extrazellulären, makromolekularen Komplex, welcher als *Cellulosom* bezeichnet wird (Schwarz 2001). Ähnliche Cellulase-Komplexe wie der von *C. thermocellum* wurden auch in anderen cellulolytischen Bakterien gefunden, darunter die mesophilen Spezies *Clostridium cellulovorans*, *Clostridium cellulolyticum*, *Clostridium papyrosolvens*, *Bacteroides cellulosolvens*, *Acetivibrio cellulolyticus*, *Ruminococcus albus*, und

Ruminococcus flavefaciens, und die thermophilen Spezies *Clostridium clariflavum* und *Clostridium josui (Bayer et al. 2013)*. Aus dieser Gruppe wurden *C. thermocellum*, *C. cellulolyticum* und *R. albus* als Bestandteil der mikrobiellen Gemeinschaft in Biogasanlagen nachgewiesen (Wirth et al. 2012). Durch die hohe Abbaueffizienz des Cellulosoms scheint *C. thermocellum* eine wichtige Rolle beim Celluloseabbau in Biogasanlagen zu spielen (Zverlov et al. 2010). Die verschiedenen Cellulosomen unterscheiden sich in erster Linie durch ihre Architektur (Bayer et al. 2013).

Besonders die Cellulosomen von *C. thermocellum*, *C. cellulolyticum* und *C. cellulovorans* wurden intensiv erforscht und sind immer noch Gegenstand einiger Studien, darunter Versuche den Komplex *in vitro* zu rekonstruieren (Krauss et al. 2012, Kruus et al. 1995, Zverlov et al. 2008, Zverlov et al. 2003, Zverlov et al. 2005). Diese Enzymkomplexe beinhalten ein nicht katalytisches, makromolekulares Scaffoldin-Protein, CipA, das die cellulosomalen Enzyme verankert. Das CipA-Rückgrat enthält bei *C. thermocellum* neun hochkonservierte Module, genannt Typ I Cohesine. Die katalytischen Komponenten sind über konservierte Typ I Dockerine mit diesen Typ I Cohesinen in CipA verbunden. Die C-terminale Region von CipA enthält zusätzlich ein Typ II Dockerin, welches mit einem Typ II Cohesin interagiert, das wiederum über einen Proteinanker mit der Zellwand des Bakteriums verbunden ist (Abbildung 6a). CipA enthält zusätzlich ein Kohlenhydratbindemodul (CBM), um das Substrat (Cellulosefasern) in die räumliche Nähe des Enzymkomplexes zu bringen und so die Abbaueffizienz zu verbessern (Schwarz 2001). Je nach Bakterium gibt es Spezies-spezifische Variationen in den Scaffoldin-Eigenschaften, dadurch variiert auch der Aufbau des Cellulosoms zwischen den Arten. Meist besitzen die Scaffoldine zwischen sechs und neun Cohesine, welche verschiedene cellulosomale Enzyme binden. Die Variation der Scaffoldine führt zu interspezifischen Unterschieden der Cellulosomen.

Die Variation der Enzymzusammensetzung bedingt Unterschiede innerhalb einer Spezies. Diese innerartlichen Abweichungen erlauben einer Spezies verschiedene Substrate effizienter abzubauen. Einige Spezies können außerdem verschiedene Typen von Scaffoldinen enthalten (Bayer et al. 2004). Alle cellulosomalen Enzyme besitzen Dockerin-Module, die mit jedem Cohesin am Scaffoldin speziesspezifisch interagieren können. Die Genomsequenz von *C. thermocellum* enthält 74 Leserahmen mit einem Dockerin Modul, ein Hinweis für eine Komponente des Cellulosoms (Zverlov et al. 2005). Die cellulosomalen Enzyme umfassen neben den Cellulasen auch Hemicellulasen, Pectinasen, Chitinasen und weitere Enzyme. Diese Enzyme entfernen offensichtlich die umgebenden Hemicellulosen von der Oberfläche der Cellulose-Kristalle und verbessern dadurch die Zugänglichkeit der Cellulasen zum Substrat (Zverlov and Schwarz 2008). Diese Komponentenanzahl ist in den Genomen verschiedener *C. thermocellum* Isolate ähnlich hoch, wobei die Anzahl der Cohesine des cipA Gens sich zwischen verschiedenen Stämmen unterscheiden kann (Koeck et al. 2015a). Der Effekt der Enzymkomplexierung auf die

Abbaueffizienz wurde durch cipA-Mutanten nachgewiesen, welche alle cellulosomalen Enzyme nur in Lösung produzieren. Die Aktivität auf kristalliner Cellulose ging durch die fehlende Komplexierung der Enzyme im Cellulosom um das 15fache zurück (Zverlov et al. 2008).

1.5.2 Nicht-cellulosomale freie Cellulasen

Das Zweikomponenten-Modell von *C. stercorarium* ist ein Beispiel für das einfachste anaerobe Cellulasesystem zum (allerdings langsamen und unvollständigen) Abbau kristalliner Cellulose (Zverlov and Schwarz 2008). Im Genom von *C. stercorarium* konnten nur zwei Gene identifiziert werden, deren Produkte auf Cellulose und β-1,4-Glucanen aktiv sind: Cel9Z (GH9-CBM3c-CBMX2-CBMX2-CBM3a) und Cel48Y (GH48-CBMX2-CBM3) (Poehlein et al. 2013). Diese beiden Enzyme zeigen einen starken Synergismus beim Abbau von Cellulose (Riedel et al. 1997). CelZ hat eine Endo-Glucanase Aktivität und CelY eine Exo-Glucanase Aktivität (Bronnenmeier et al. 1991, Bronnenmeier and Staudenbauer 1993). Wie CelZ und CelY besitzen freie Cellulasen meist ebenfalls CBMs, welche das katalytische Modul in räumliche Nähe zur Oberfläche des kristallinen Substrats bringen (Abbildung 6b) (Himmel et al. 2010). Neben dem cellulolytischen besitzt *C. stercorarium* vor allem ein großes hemicellulolytisches Potential. Aus dem Genom konnte ein Enzymsystem, bestehend aus einer Endo-Xylanase, einer Arabinofuranosidase und einer ß-Xylosidase, rekonstituiert werden, welches Arabinoxylan komplett abbaut (Adelsberger et al. 2004). Daher ist auch diese Spezies weitverbreitet bei der Zersetzung von Biomasse und wurde ebenfalls in Biogasanlagen nachgewiesen (Madden 1983, Schwarz et al. 2004, Zverlov et al. 2010).

1.5.3 Multifunktionale Enzyme

Das dritte System zum Abbau pflanzlicher Biomasse in anaeroben Ökosystemen sind multifunktionale Enzyme. Sie besitzen in der Regel einen modularen Aufbau aus ein oder mehreren katalytischen Modulen und zusätzlich mehrere CBMs und dadurch ein relativ hohes Molekulargewicht (Morais et al. 2012). Diese multifunktionalen Enzyme sind entweder freie Enzyme oder Bestandteil von Cellulosomen. Bisher wurden vier verschiedenen Typen von multifunktionalen Enzymen beschrieben: Cellulase-Cellulase, Hemicellulase-Hemicellulase, Hemicellulase-Cellulase und Hemicellulase-Kohlenhydratesterase Systeme (Himmel et al. 2010). Einige bekannte Cellulase-Cellulase Systeme beinhalten sowohl eine Endo-Glucanase als auch eine Exo-Glucanase auf derselben Polypeptidkette. Eines der am besten beschriebenen Cellulase-Cellulase Multienzyme ist CelA aus *Caldicellulosiruptor bescii,* einem extrem thermophilen, obligat anaeroben, chemoorganotrophen Bakterium, das aus heißen Quellen isoliert wurde (Svetlichnyi et al. 1990). Sequenzanalysen der wichtigsten Cellulase von *C. bescii* zeigten, dass CelA aus zwei katalytischen Modulen der Familie 9 und 48 und drei Cellulose-Binde-Modulen besteht (GH9-

CBM3-CBM3-CBM3-GH48) (Zverlov et al. 1998). Die drei CBMs sind vermutlich erforderlich, um das Substrat auch bei hyperthermophilen Bedingungen fest zu binden (Himmel et al. 2010). Durch die räumliche Nähe der beiden Enzymaktivitäten entsteht ein Synergismus, der die Effizienz des Abbaus gegenüber der Mischung zweier Einzelenzyme wie in *C. stercorarium* noch erhöht.

1.5.4 Vergleich der Cellulose-Abbaueffizienz

Aufgrund der hohen biotechnologischen Bedeutung waren die Vor- und Nachteile jeder der drei Abbausysteme Gegenstand einiger Forschungsprojekte (Caspi et al. 2008, Caspi et al. 2009, Morais et al. 2010). Es wurde nachgewiesen, dass die gleichen Enzyme in einem Cellulosom gebunden kristalline Cellulose effizienter abbauen als die Mischung der Einzelenzyme (Caspi et al. 2008). Morais et al. (2012) beschreiben die Anwendung des cellulolytischen Systems von *Thermobifida fusca* als Model, um die Effizienz der drei Systeme zu vergleichen. Dabei wurden die Konstrukte allerdings rekombinant aus Einzelcellulasen hergestellt. Eine andere Möglichkeit, um die natürlich vorkommenden Systeme zu vergleichen, ist die Bestimmung der Enzymaktivität auf kristalliner Cellulose (Avicel) im Zellüberstand verschiedener Wildtypsysteme (Tabelle 2).

Tabelle 2: Spezifische Aktivität von Kulturüberstand auf Avicel, bestimmt unter Standardbedingungen (Koeck et al. 2014a)

Stamm	*Clostridium thermocellum ATCC 27405*	*Caldicellulosiruptor bescii DSM 6725*	*Clostridium stercorarium DSM 8532*
Enzym System	cellulosomal	multifunctional	frei
Spez. Aktivität (mU/mg)	90 +/- 4,6	60	30 +/- 0,3

Aus Tabelle 2 ist ersichtlich, dass offensichtlich das System der freien Cellulasen von *C. stercorarium* die niedrigste Aktivität auf Avicel zeigt. Etwa doppelt so hoch ist die Aktivität der multifunktionalen Enzyme. Dieses System wird aber durch das Cellulosom von *C. thermocellum* nochmals deutlich übertroffen.

1.6 Cellulolytische Mikroorganismen in Biogasanlagen

Biogas in Biogasanlagen wird von einem natürlichem Konsortium aus Bakterien, Archaeen und anaeroben Pilzen produziert (Fliegerova et al. 2010). Für den primären Abbau der Biomasse, die dem Fermenter zugegeben wird, sind in erster Linie die Bakterien verantwortlich. Diese hydrolytische Gemeinschaft produziert extrazelluläre Enzyme, Zellbiomasse, sowie freie Zucker für andere Bakterien und konsumiert den restlichen Sauerstoff im Fermenter. Dies ermöglicht den strikt anaeroben Archaea erst die Vermehrung und Biogasproduktion. Daher ist dieser erste

Schritt des Biogasbildungsprozesses, die Hydrolyse, entscheidend für die Effizienz der Methanproduktion und wird als geschwindigkeitsbestimmender Schritt angesehen, weil alle weiteren Prozesse davon abhängen. Die Abbaurate hängt dabei stark von der Beschaffenheit des Substrats ab. Der Abbau von Cellulose und Hemicellulose geht generell langsamer und unvollständiger vonstatten als der Abbau von Proteinen. Es wurde aber gezeigt, dass die gezielte Zugabe von cellulolytischen Bakterien mit sehr guter Abbaueffizienz die Hydrolyse in Biogasanlagen beschleunigen kann und so die Biogasproduktion schneller und effizienter wird (Koeck et al. 2015a). Die Zugabe externer Enzyme hat dagegen einen geringen Effekt und ist sehr kostenintensiv. Durch die aktive Sekretion hydrolytischer Enzyme in das umgebende Medium werden Zucker aus der cellulosehaltigen Biomasse freigesetzt, die wiederum auch Nahrungsgrundlage für saccharolytische Bakterien darstellen. Diese stellen wiederum die Edukte für die Methanproduktion zu Verfügung. Durch wiederholte Zugabe von ausgewählten hydrolytischen Kulturen kann die Methanproduktion messbar erhöht werden (Köllmeier 2013).

C. thermocellum z.B. nutzt in erster Linie Polysaccharide als Energiequelle und nimmt Cellodextrine direkt in die Zellen auf; Glucose und andere Monosaccharide werden selten verstoffwechselt. Diese stehen anderen Bakterien des trophischen Netzwerks zu Verfügung. Nur ein kleiner Anteil der cellulolytischen Bakterien aus Biogasanlagen wurde bisher in Reinkultur isoliert, darunter *C. thermocellum* (Koeck et al. 2014d). Dagegen wurden diverse cellulolytische Bakterien, die aus anderen anaeroben Habitaten wie z.B. Pansen stammen, intensiv untersucht und charakterisiert (White et al. 2014). Obwohl sich bereits einige Studien mit der Biogasmikrobiologie befasst haben, herrscht generell noch ein Mangel an Informationen bezüglich der genauen Zusammensetzung der mikrobiologischen Gemeinschaft in allen Stufen des Biogasprozesses, insbesondere im ersten Schritt, der Hydrolyse (Krakat et al. 2010, Ziganshin et al. 2013).

1.7 Analyse der mikrobiellen Lebensgemeinschaften in Biogasanlagen

Zur spezifischen Detektion von Mikroorganismen stehen generell drei alternative Analyseverfahren zur Verfügung: der Nachweis der Mikroorganismen *in situ*, die Kultivierung der Mikroorganismen mittels spezieller Nähr- und Selektionsmedien und der kulturunabhängige Nachweis der Mikroorganismen auf Basis artspezifischer Moleküle. Die gewonnenen Isolate können durch spezifische genetische Marker, z.B. 16S rRNA Gene oder alternativ auch auf Basis der Erzeugung spezifischer Zellproteinprofile mittels MALDI-TOF MS identifiziert werden. Mit unterschiedlichen PCR-basierten Genotypisierungsmethoden können auch einzelne Stämme einer Art voneinander differenziert werden (Brazier 2001). Trotz einer Reihe von Studien muss die Biogas-Mikrobiologie im Allgemeinen noch als „*black box*“, also als weitestgehend unbekannter Mikrokosmos angesehen werden (Cirne et al. 2007, Ziganshin et al. 2013). Dennoch wird die Aufklärung der mikrobiellen Zusammenhänge allgemein als Schlüssel für die weitere

verfahrenstechnische Optimierung der Biogasfermentation betrachtet (Wietschel et al. 2010). Die mikrobiologische Analytik wird jedoch dadurch erschwert, dass die Mikrobengemeinschaft aus hunderten der unterschiedlichsten Arten von Mikroorganismen besteht, welche aufgrund der anaeroben Lebensweise und syntropher bzw. symbiotischer Wechselbeziehungen häufig (noch) nicht oder nur schwer kultivierbar sind. Viele Arten sind bislang weder taxonomisch noch physiologisch beschrieben. In den letzten Jahren konnten durch verschiedene kulturunabhängige Hochdurchsatzverfahren in der mikrobiologischen Analytik technologische Durchbrüche erzielt werden, welche es möglich erscheinen lassen, die „black box" der Biogas-Mikrobiologie zu entschlüsseln. Hierzu zählen insbesondere sog. OMIK-Technologien, wie Genomik (Analyse des gesamten Erbguts einer Zelle), Transkriptomik (Analyse der exprimierten Gene einer Zelle), Proteomik (Analyse der Proteine in einer Zelle) und Metabolomik (Analyse der Metabolite einer Zelle). Mit der Entwicklung neuer Sequenzierungstechnologien können ganze Metagenome mikrobieller Gemeinschaften schnell und kostengünstig entschlüsselt werden (Krause et al. 2008, Rademacher et al. 2012, Schlüter et al. 2008, Zakrzewski et al. 2013). Aufgrund mangelnder Referenzsequenzen von bereits beschriebenen Spezies und der hohen Anzahl an noch nicht beschriebenen Spezies, sowie der relativ kurzen DNA-Fragmentlängen sind die hiermit erhaltenen Datenmengen bislang jedoch nur ansatzweise auswertbar.

Nur ein geringer Prozentsatz der vorhandenen bakteriellen Spezies in Biogas Fermentern wurde bisher in Reinkultur isoliert. Die Kultivierung neuer Arten ist aber notwendig, um den Metabolismus und somit das Substrat- und Produktspektrum und die damit verbundene Rolle einer Art im Nahrungsnetz der Biogasanlage aufzuklären. Der überwiegende Anteil der zuordenbaren Sequenzen aus den Metagenomanalysen gehören zum Phylum Firmicutes und zur Klasse Clostridia (Krause et al. 2008, Wirth et al. 2012). Bezüglich der Sequenzen, die für Proteine kodieren, welche in den Abbau von Poly- und Oligosacchariden involviert sind, sind wiederum Firmicutes mit über 50% die dominierende phylogenetische Gruppe (Krause et al. 2008). Dementsprechend gehören auch die meisten identifizierten Spezies mit bekannten cellulolytischen Vertretern zur Klasse Clostridia. Cellulolytische Spezies, die in andere Phyla gehören, wurden nur selten identifiziert. Eine Auflösung der Metagenomdatensätze bis auf Speziesebene zeigt, dass meist *Clostridium thermocellum* die abundanteste cellulolytische Spezies ist (Krause et al. 2008, Wirth et al. 2012). Andere cellulolytische Bakterien mit hoher Abundanz (innerhalb der 40 häufigsten Spezies) sind *Clostridium cellulolyticum, Ruminococcus albus, Clostridium saccharolyticum,* und *Caldicellulosiruptor saccharolyticus* (Wirth et al. 2012). Die hohen Abundanzen cellulolytischer Bakterien aus der Klasse Clostridia sprechen für die große Bedeutung dieser Bakteriengruppe beim Abbau komplexer Substrate in der Biogasanlage. Die meisten Metagenom-Sequenzen, die von Clostridien stammen, können tatsächlich Genen bzw. Enzymen zugeordnet werden, die in den Abbau von Kohlenhydraten involviert sind (Schlüter et al., 2008; Rademacher et al., 2012). In der Tat weisen alle kulturunabhängigen Studien darauf hin,

dass vor allem die Clostridia wichtig für die Durchführung essentieller Stoffwechselwege im Prozess der Biogasproduktion sind. Die meisten Metagenom-Sequenzen können jedoch nur höheren taxonomischen Einheiten zugeordnet werden; für eine genauere Klassifizierung bis auf Artebene sind klassische mikrobiologische Ansätze unumgänglich. Generell werden durch Metagenom-Analysen auch zahlreiche Sequenzen (bis zu 74 % aller Einzelsequenzen) generiert, welche sich weder einer Familie noch einer Klasse zuordnen lassen (Abbildung 7) (Krober et al. 2009).

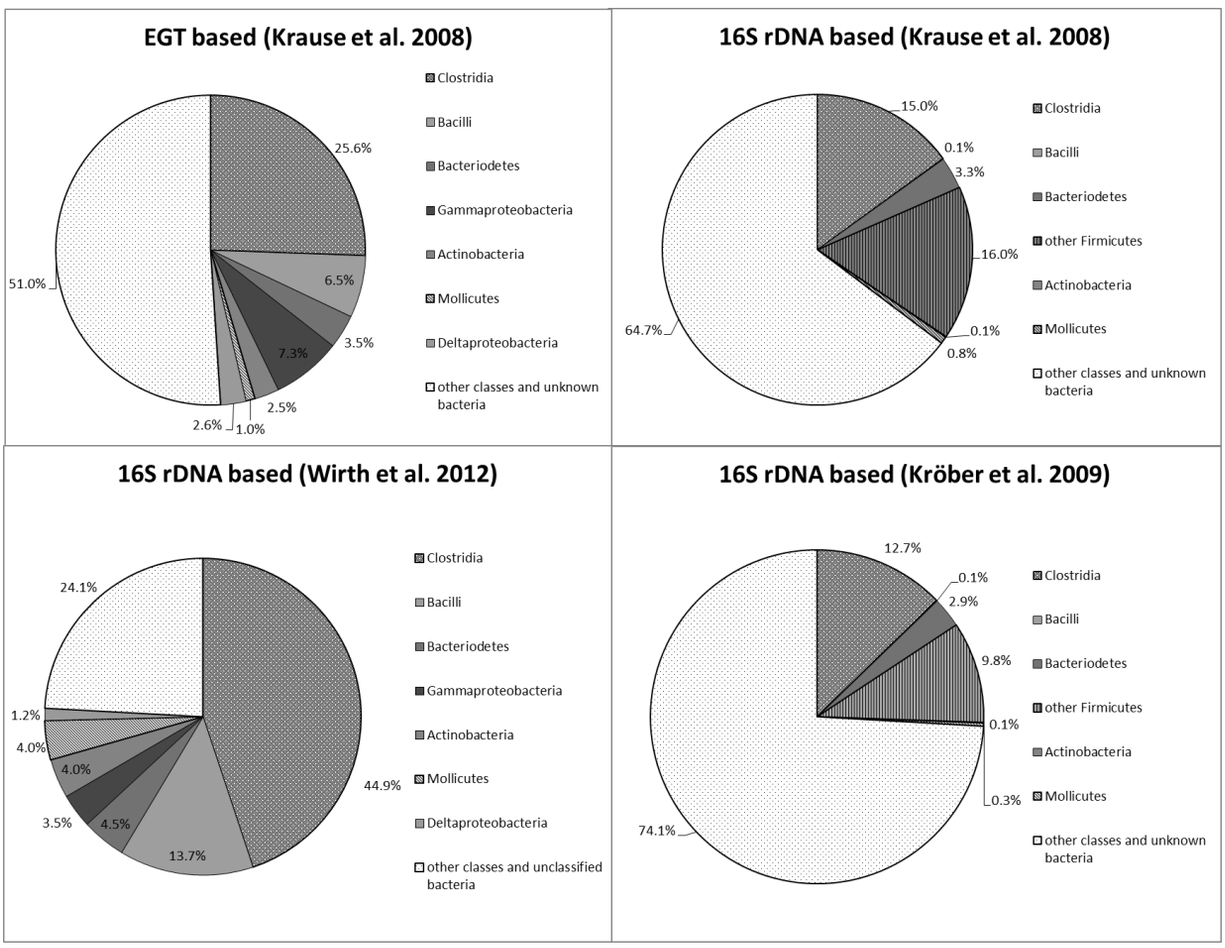


Abbildung 7:Taxonomische Profile der Bakterien aus verschiedenen mikrobiellen Gemeinschaften in Biogasanlagen basierend auf Metagenomsequenzen. Dargestellt ist die Zuordnung zu den einzelnen phylogenetischen Klassen. Die Autoren vergleichen entweder 16S rRNA Gensequenzen oder EGTs (Environmental Gene Tags) mit den Referenzdatenbanken Ribosomal Database Project Classifier oder ARB (Modifiziert nach Wirth et al. 2012; Kröber et al. 2009; Krause et al. 2008).

Dies weist darauf hin, dass aus Biogasanlagen ganze Mikroorganismengruppen noch nicht kultiviert und identifiziert wurden. Es ist vermutlich auch möglich, dass neuartige Bakterien, die maßgeblich am Celluloseabbau beteiligt sind, durch mangelnde Referenzsequenzen noch nicht identifiziert worden sind. Diese Wissenslücke kann nur durch Isolierung und Charakterisierung neuer cellulolytischer Bakterien geschlossen werden. Dies liefert wiederum einen wichtigen Beitrag, um die “Black Box” der Biogas-Mikrobiologie besser auszuleuchten und verborgene Potentiale zu nutzen. Durch Sequenzierung der Genome werden zum einen neue Referenzdaten

für Metagenomanalysen generiert, zum anderen wird der Zugang zu neuen Glycosidhydrolasen geschaffen. Diese Enzyme konnten durch die Umsetzung diverser Substrate für eine Vielzahl von industriellen Anwendungen genutzt werden.

1.8 Isolierung und Charakterisierung von Bakterien

Aktuelle Schätzungen gehen davon aus, dass 61 bakterielle Phyla existieren, von denen 31 bisher nicht kultiviert werden können (Vartoukian et al. 2010). Nach bisherigem Wissenstand ist es nicht möglich, die natürlichen Wachstumsbedingungen für alle Mikroorganismen bereit zu stellen. Mit künstlichen Medien schafft man es im Labor meist nicht, die abiotischen und biotischen Bedingungen, die ein bestimmtes Bakterium benötigt, exakt nachzustellen. Auch können sich Mikroorganismen oft nicht schnell genug an veränderte Bedingungen anpassen (Alain & Querellou, 2009). Dennoch wurden in den letzten 25 Jahren große Fortschritte bei der Entwicklung neuer Kultivierungstechniken erzielt, welche enorm zu einem besseren Verständnis der mikrobiellen Diversität beitragen.

Folgende Ansätze wurden dabei verfolgt: (i) verbesserte Medien, (ii) Veränderung der Wachstumsbedingungen, (iii) Kokultivierung, (iv) Einsatz von Transwellplatten mit Membranen, (v) optische Pinzetten und Mikromanipulatoren, (vi) Hochdurchsatz-Mikrobioreaktoren, (vii) Diffusionskammern, um das natürliche Milieu zu simulieren, und vieles mehr (Kaeberlein et al. 2002, Pham and Kim 2012). Die Gewinnung neuer Isolate hilft zu verstehen, welche Typen von Bakterien existieren, und deren Rolle in der Umwelt aufzuklären. Dies eröffnet neue Einblicke in die ökologische Bedeutung eines Bakteriums für die stoffliche Umsetzung von Biomasse und dessen potentielle Anwendung für industrielle Anforderungen. Besonders die vielfältigen Anwendungsbereiche neuer Isolate rechtfertigen eine stetige Weiterentwicklung der Kultivierungsmethoden. Zur Erreichung dieser Ziele ist ein umfassendes Verständnis der chemischen, physikalischen und biologischen Prozesse des Milieus erforderlich, in dem die Bakterien leben. Dieses Wissen muss dann in den entsprechenden Kultivierungsansatz einfließen. Die Hauptaufgabe für eine erfolgreiche Kultivierung ist die exakte Nachstellung des natürlichen Systems (Pham and Kim 2012).

Das Isolieren von Bakterien aus natürlichen Habitaten ist nicht trivial. Für eine erfolgreiche Kultivierung einzelner Mikroorganismen ist es erforderlich, die Kultivierungsbedingungen mit der Einstellung unterschiedlicher Parameter (Kultivierungsmedium, Temperatur, pH-Wert, Sauerstoffgehalt und andere) optimal und vor allem spezifisch an die Bedürfnisse des jeweiligen Zielorganismus anzupassen. Die Entwicklung angepasster und spezifischer Kultivierungsmedien ist aber ein zeitaufwendiger Prozess, der darüber hinaus für jede Organismengruppe von neuem wiederholt werden muss. Bei der Bei der meist Initial vorgenommenen Anreicherung von Bakterien kommt es außerdem oft zu einer Verschiebung der Populationszusammensetzung.

Dadurch werden Organismen, die auf dem eingesetzten Kultivierungsmedium schnell wachsen können, häufiger isoliert als langsam wachsende Organismen, auch wenn die langsam wachsenden Organismen zahlenmäßig häufiger in der Ausgangspopulation vertreten waren. Eine allein auf der Kultivierung mit Standardmedien beruhende Beschreibung der Biozönose ist also limitiert und repräsentiert nicht die tatsächlichen Verhältnisse im untersuchten mikrobiologischen Habitat (Wenzel 2002).

Eine weitere Herausforderung stellt die Reinigung eines neuen Isolates von allen Kontaminationen dar. Liegt das Isolat in Reinkultur vor, kann dieses zunächst anhand seiner 16S rRNA Gensequenz taxonomisch eingeordnet werden. Für die genaue Zuordnung zu einem Taxon muss beachtet werden, dass zwei Stämme, die weniger als 97 % 16S rRNA Gen Sequenzähnlichkeit aufweisen, nicht derselben Art zugeordnet werden können. Sequenzähnlichkeiten unter 95 % weisen auf unterschiedliche Gattungen hin (Tindall et al. 2010). Zu ihrer Charakterisierung müssen andere Merkmale, wie zum Beispiel die Physiologie der Organismen, herangezogen werden.

1.9 Fragestellung und Ziel der Arbeit

Die vorliegende Arbeit gliedert sich im Wesentlichen in zwei Teile. Das erste der Arbeitspakete beschäftigt sich mit der Ermittlung der genetischen Variation eines für den anaeroben Abbau in der Biogasanlage zentral wichtigen Bakteriums. Als Zielorganismus wurde *C. thermocellum* ausgewählt. Hierbei handelt es sich um ein sowohl auf genomischer Ebene als auch auf metabolischer Ebene sehr gut charakterisiertes cellulolytisches, thermophiles, anaerobes Bakterium, welches wie bereits erwähnt Bestandteil des hydrolytischen Konsortiums in Biogasanlagen ist. Dazu sollten verschiedene Genotypisierungsmethoden, die bereits für andere Bakterien etabliert wurden, auf *C. thermocellum* Stämme übertragen werden. Gegebenenfalls sollten neue Methoden zur Genotypisierung für *C. thermocellum* entwickelt und etabliert werden. Die Resultate der Genotypisierung sollten mit Unterschieden des Zellproteinprofils (vorrangig ribosomale Proteine) verglichen werden. Aus den Ergebnissen sollte abgeleitet werden, inwieweit sich die Stämme verschiedener Biogasanlagen unterscheiden. Daraus können wiederum Rückschlüsse auf die Variation des mikrobiellen Konsortiums zwischen verschiedenen Biogasanlagen gezogen werden. Zusätzlich sollte von allen untersuchten *C. thermocellum* Stämmen deren cellulolytisches Potential, Wachstumsbedingungen und Substratspektrum ermittelt werden, um Stämme zu identifizieren, welche in der Lage sind, besonders effektiv lignocellulosehaltige Biomasse in Biogasanlagen abzubauen.

Der zentrale Focus der Arbeit liegt auf dem zweiten Arbeitspaket: der Isolierung möglichst vieler neuer Arten/Stämme aus Modellbiogasanlagen. Dies soll die Diversität der Mikroorganismen im Biogasprozess abbilden, die “Black Box” der Biogas Mikrobiologie besser ausleuchten und als

Basis für weitere molekulargenetische Untersuchungen dienen. Dazu sollten zunächst geeignete Modellfermenter ausgewählt und neue Isolierungsstrategien für cellulolytische Bakterien entwickelt werden. Anschließend sollten neue Isolate gereinigt, identifiziert und charakterisiert werden. Das Hauptaugenmerk sollte hierbei auf das cellulolytische und hemicellulolytische Potential der Isolate gelegt werden. Von einigen Isolaten sollte zusätzlich zur physiologischen und molekularbiologischen Charakterisierung der Bakterienkulturen das Genom sequenziert werden. Dies liefert zum einen Referenzdaten für die Interpretation von Metagenomanalysen; zum anderen soll dadurch die Ausstattung der neuen Isolate mit Glycosidhydrolasen und die Art und Beschaffenheit des cellulolytischen Systems aufgeklärt werden, was zur Identifizierung von Enymen mit industriellem Anwendungspotential führen kann.

2 MATERIAL UND METHODEN

2.1 CHEMIKALIEN UND VERBRAUCHSMATERIALIEN

Die verwendeten Chemikalien und Verbrauchsmaterialien wurden, wenn nicht anders vermerkt, von den Firmen Amersham Pharmacia Biotech (Freiburg), Applied Biosystems (Weiterstadt), BD (Heidelberg), Biozym Diagnostik (Hessisch Oldendorf), Fluka (Buchs; Schweiz), J.T. Baker (Deventer, Holland), Megazyme (Wicklow, Ireland), Merck (Darmstadt), Promega (Mannheim), Qiagen (Hilden), Roth (Karlsruhe), Sigma-Aldrich (Taufkirchen), Thermo Fisher Scientific (Waltham, USA) bezogen.

Die verwendeten Kits, Enzyme und Längenstandards für Elektrophoresen sowie deren Bezugsquelle sind in folgender Tabelle zusammengefasst. Soweit nicht anders vermerkt, wurden alle Kits und Enzyme gemäß den Herstellerangaben verwendet.

Tabelle 3: Verwendete Kits, Enzyme und Längenstandards

Verwendung	Kits/Enzyme	Hersteller
Isolierung genomischer DNA	FastDNA™ SPIN Kit for Soil	MP-Biomedicals (Santa Ana, USA)
Isolierung genomischer DNA	PrestoSpin D Bug	Molzym (Bremen)
Isolierung genomischer DNA	innuPREP DNA Mini Kit	Analytic Jena (Jena)
Restriktionsverdau	Restriktionsenzyme NdeI, XhoI, EcoRI	Thermo (Waltham, USA)
Plasmidpräparation	QIAprep Spin Miniprep Kit	QIAGEN (Hilden)
Gelextraktion	QIAquick Gel Extraction Kit	QIAGEN
PCR Produkt Reinigung	QIAquick PCR Purification Kit	QIAGEN
PCR	DreamTaq DNA Polymerase	Thermo
PCR	Phusion DNA Polymerase	Thermo, NEB (Frankfurt am Main)
Klonierung	TOPO® TA Cloning® Kit	Thermo
Klonierung	TOPO® XL PCR Cloning Kit	Thermo
Klonierung	CloneJET PCR Cloning Kit	Thermo
Klonierung	Gibson Assembly® Master Mix	NEB
Längenstandard DNA Agarosegelelektrophorese	GeneRuler™ DNA Ladder Mix, 100-10000 bp	Thermo
Längenstandard SDS PAGE	PageRuler™ Plus Prestained Protein Ladder, 10-250 kD	Thermo

2.2 GERÄTE

Soweit nichts anderes erwähnt wurden alle Versuche mit den Geräten, die in folgender Tabelle gelistet sind, durchgeführt.

Tabelle 4: Verwendete Geräte

Gerät	Hersteller
FPLC	Pharmacia LKB (Uppsala, Schweden)
ÄKTA pure	GE Healthcare Life Sciences (Freiburg)
Type A Anaerobic Chamber	Coy Laboratory Product Inc. (Michigan, USA)
Model H3 (Agarose-Gelelektrophorese)	Bethesda Research Laboratories Inc. (Gaithersburg, USA)
Mini-Protean Tetra Cell (SDS-PAGE)	BIO-RAD (Hercules, USA)
EPS 600 (Agarose-Gelelektrophorese)	Pharmacia Biotech (Uppsala, Schweden)
2297 MACRODRIVE 5 (SDS-PAGE)	LKB Bromma (Stockholm, Schweden)
Mikroskop: Axio Skope.A1	Carl Zeiss Microscopy GmbH (Jena)
NanoDrop ND-1000 Spektrophometer	PEQLAB Biotechnologie GMBH (Erlangen)
PCR-Thermocycler: Primus 96 advanced	PEQLAB Biotechnologie GMBH (Erlangen)
Ultrospec PLUS	Pharmacia LKB (Uppsala, Schweden)
Ultrospec 2000	Pharmacia Biotech (Uppsala, Schweden)
Lambda 35	Perkin Elmer (Waltham, USA)
Sorvall RC 6+ Centrifuge	Thermo Scientific (Waltham, USA)
ROTINA 35R	Andreas Hettich GmbH & Co. KG (Tuttlingen)
Biofuge pico	Heraeus (Hanau)
MIKRO 200R	Andreas Hettich GmbH & Co. KG (Tuttlingen)
Sartorius 1204 MP Waage	Sartorius (Göttingen)
Precisa 125 A SCS Feinwaage	PESCA Waagen GmbH (Schweiz)

2.3 BIOGASFERMENTER

In folgender Tabelle sind alle Biogasfermenter gelistet, von denen im Laufe dieser Arbeit Gärrest, Inokulum oder Probenmaterial für die Versuche bezogen wurde. Auch Fermenter die ohne ständige Zugabe von Gülle betrieben werden müssen initial immer mit Gülle inokkuliert werden.

Tabelle 5: Liste aller beprobten Biogasanlagen mit den wichtigsten Prozessparamentern (Temperatur, Substrat)

Name	Biogasanlage	Temperatur	Substrat
T3 (TUM3)	Pellmeyer	38-40°C	Gülle, Stroh, Gras, Grassilage
A1 (ATB1)	BGA 03: Archaea	40°C	Maissilage, Roggen-GPS, Grassilage, Ferkelmist, Rindergülle
A2 (ATB2)	BGA 07: BiogasNord	40°C	Maissilage, Grünroggen, Rinder- und Schweinegülle
H1 (HAW1)	4Biogas NOVA Fermenter	54°C	Mais, Gülle, Gras
H2 (HAW2)	Beta-Biomethan Schuby GmbH	40°C	Zuckerrübensilage
H3 (HAW3)	Agraferm Technologies AG	42°C	Gras, Mais
J1 (JGU1)	Biogasanlage Bioenergie GbR	43/53°C	Maissilage, Schweinegülle
J2 (JGU2)	Biogasanlage Obere Blies GmbH & Co. KG	40°C	Maissilage, Grassilage, Festmist, Rindergülle
J3 (JGU3)	Biogasanlage Wagner	45/46°C	Maissilage
BM1T	Laborfermenter LfL	55°C	Grassilage, Rindergülle
BM2T	Laborfermenter LfL	55°C	Grassilage, Rindergülle

Die Biogasanlage H1 (4Biogas NOVA-Fermenter, Viersen) diente als wichtigste Bezugsquelle für die Isolierung cellulolytischer Bakterien im thermophilen Bereich (45-60 °C). Hierbei handelt es sich um zwei-phasige Fermenter mit integrierter Hydrolyse im äußeren Bereich (Abbildung 8). Jeder der drei Fermentertürme hat ein Volumen von ca. 115 m³ bei einer Höhe von 15 m und eine durchschnittliche Verweildauer von 21 Tagen.

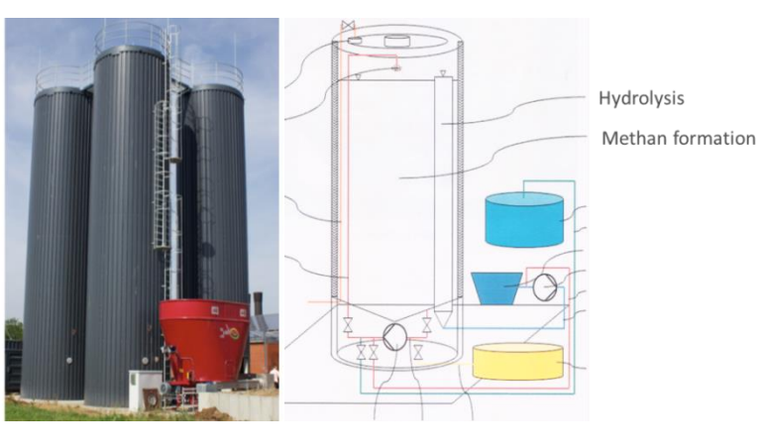


Abbildung 8: Gesamtansicht und schematische Darstellung des 4Biogas NOVA Fermenter in Viersen

2.4 MEDIEN

Die folgenden Angaben beziehen sich jeweils auf die Herstellung von einem Liter Medium. Vor dem Verschließen der Gefäße wurde der pH-Wert gemessen und abhängig von den jeweiligen Anforderungen mit 2 M HCl oder 1 M NaOH auf einen pH Wert zwischen 5,0 und 9,5 eingestellt. Medien für anaerobe Mikroorganismen wurden in N_2/H_2-Atmosphäre (98:2 %, v/v) unter anaeroben Bedingungen in einem Anaearobenzelt (Coy, Michigan, USA) mit einem Butylgummistopfen und einem Alu-Crimp-Verschluss verschlossen. Für Festmedien wurde 1,8 % Agar vor dem Autoklavieren zugegeben. Anaerobe Platten wurden während der Inkubation in AnaeroJars (Thermo Fisher Scientific, Waltham, USA) aufbewahrt. Alle Medien und Lösungen wurden 20 min bei 121 °C und 1,21 bar autoklaviert (Varoklav Dampfsterilisator 220T, H+P Labortechnik, Oberschleißheim). Zusätze, die nicht autoklaviert werden konnten, wurden steril filtriert (0,22 µm).

2.4.1 BASISMEDIEN ZUR GEWINNUNG NEUER ISOLATE

Für die unterschiedlichen Isolierungsversuche und Anzuchten wurde jeweils eines der Basismedien mit einer Kohlenstoffquelle und teilweise mit zusätzlichen Additiven versetzt.

2.4.1.1 GS2-Medium (Johnson et al. 1981)

Tabelle 6: GS2 Medium

GS2-Medium Stammlösung		GS2-Medium Salzlösung (100x)	
KH_2PO_4	1,5 g	$MgCl_2$ x 6 H_2O	10,0 g
K_2HPO_4	2,9 g	$CaCl_2$ x 2 H_2O	1,5 g
Harnstoff	2,1 g	$FeSO_4$ (1,25 %)	1 ml
Cystein	1 g	ddH_2O	100 ml
MOPS	10 g		
Hefe-Extrakt	6 g		
Na-Citrat	2,9 g		
Resazurin	2 mg		
ddH_2O	Ad 1000 ml		

Durch die Zugabe von Cystein wurde der im Medium vorhandene Restsauerstoff zu Wasser reduziert. Resazurin, welches in Kontakt mit Sauerstoff von farblos zu violett umschlägt, wurde als Redoxindikator eingesetzt. Die Salzlösung wurde sterilfiltriert und nach Abkühlen des Mediums mit einer Spritze unter sterilen Bedingungen zugegeben.

2.4.1.2 Mineralmedium (Komplex) (Widdel and Bak 1992)

Die Stammlösung für den Ansatz des Anaerobmediums wurde gemäß Tabelle 7 hergestellt. Die Spurenelementelösung wurde mit 1 M NaOH-Lösung auf pH 6,0 eingestellt. Beide Vitaminlösungen wurden sterilfiltriert und lichtgeschützt bei 4 °C gelagert.

Tabelle 7: Mineralmedium-Stammlösung und Spurenelementlösung SL9

Mineralmedium-Stammlösung		**Spurenelementlösung SL9** (Tschech and Pfennig 1984)	
KH_2PO_4	0,20 g	Nitrilotriessigsäure (NTA)	12,8 g
NH_4Cl	0,27 g	$FeCl_2$ x 4 H_2O	1,5 g
NaCl	1 g	$CoCl_2$ x 6 H_2O	190 mg
$MgCl_2$ x 6 H_2O	0,41 g	$MnCl_2$ x 2 H_2O	80 mg
KCl	0,52 g	$ZnCl_2$	70 mg
$CaCl_2$ x 2 H_2O	0,15 g	Na_2MoO_4 x 2 H_2O	36 mg
Cystein	1 g	$NiCl_2$ x 6 H_2O	24 mg
Resazurin	0,02 % (w/v)	H_3BO_3	6 mg
Add ddH_2O	1000 ml	$CuCl_2$ x 2 H_2O	2 mg
		ddH_2O	1000 ml

Tabelle 8: Selenit-Wolframat-Lösung

Selenit-Wolframat-Lösung (Tschech and Pfennig 1984)	
NaOH	200 mg
Na_2SeO_3 x 5 H_2O	6 mg
Na_2WO_4 x 2 H_2O	8 mg
ddH_2O	1000 ml

Tabelle 9: Vitaminlösungen

Vitaminlösungen (2000x) (Tschech and Pfennig 1984)			
4-Aminobenzoesäure	4 mg	Cyanocobalamin (B12)	5 mg
D(+)-Biotin	1 mg	ddH_2O	100 ml
Nikotinsäure	10 mg		
Ca-D(+)-Pantothenat	5 mg		
Pyridoxamindihydrochlorid	15 mg		
Thiaminchlorid-Dihydrochlorid	10 mg		
ddH_2O	100 ml		

Zur Herstellung des Bicarbonatpuffers wurden 7,06 g Natriumhydrogencarbonat in 84 ml Wasser in einer Schraubdeckelflasche (250 ml) gelöst.

Zu 993,5 ml Stammlösung wurden nach dem Autoklavieren folgende Lösungen anoxisch und steril zugegeben:

- Selenit-Wolframat-Lösung 1 ml
- Spurenelementlösung SL9 1 ml
- Bicarbonatlösung 3,05 ml
- Vitaminlösung 0,5 ml
- Vitamin B12-Lösung 1 ml

2.4.1.3 Dilute Nutrient Broth (DNB) (Janssen et al. 2002).

Zur Herstellung von „Dilute nutrient broth" (DNB) wurden 0,08 g Difco nutrient broth (BD Diagnostic Systems, Sparks, Md.) in einem Liter entsalztem Wasser gelöst.

2.4.1.4 Basal Salz Medium (Gupta et al. 2012)

Tabelle 10: Basal Salz Medium

$NaNO_3$	2.5 g
KH_2PO_4	2 g
$MgSO_4$	0,2 g
NaCl	0,2 g
$CaCl_2$ x 6 H_2O	0,1 g
ddH_2O	1000 ml

2.4.2 SONSTIGE MEDIEN

2.4.2.1 LB (lysogeny broth) Medium (pH 7,5)

Tabelle 11: LB Medium

Trypton	10 g
Hefe-Extrakt	5 g
NaCl	10 g
ddH_2O	1000 ml
pH	7,5

2.4.2.2 SOC Medium (pH 7,5)

Tabelle 12: SOC Medium

Trypton	20 g
Hefe-Extrakt	5 g
KCl	2,5 mM
Glucose	20 mM
ddH_2O	1000 ml

2.4.2.3 SOB-(Super Optimal Broth) Medium (pH 7,5)

Tabelle 13: SOB Medium

Trypton	20 g
Hefe-Extrakt	5 g
KCl	2,5 mM
NaCl	10 mM
$MgSO_4$	20 mM
ddH_2O	1000 ml

2.4.2.4 ZYP5052-Medium (Autoinduktionsmedium) (Studier 2005)

Tabelle 14: Einzelkomponenten des ZYP5052-Mediums

ZY		**1M $MgSO_4$**	
Trypton	10 g	$MgSO_4$ x $7H_2O$	24,65 g
Hefeextrakt	5 g	ddH_2O	ad 100 ml
ddH_2O	925 ml		
20x NPS		**50x 5052**	
$(NH_4)_2SO_4$	66 g	Lactose	100 g
KH_2PO_4	136 g	ddH_2O	730 ml
Na_2HPO_4	142 g	Glycerin	250 g
ddH_2O	900 ml	Glucose	25 g

Alle Komponenten wurden getrennt autoklaviert und vor der Nutzung, wie in Tabelle 15 angegeben, frisch gemischt.

Tabelle 15: ZYP5052-Medium

Komponenten		**Endkonzentration**
ZY	~928 ml	-
1 M $MgSO_4$	1 ml	1 mM
50x 5052	20 ml	1x
20x NPS	50 ml	1x

2.4.3 KOHLENSTOFFQUELLEN

Tabelle 16: Für Kultivierungsmedien und Enzymaktivitätsttests verwendete Kohlenstoffquellen

Substrat	Zusammensetzung und glycosidische Verknüpfung
Monosaccharide	
Glucose	
Fructose	
Mannose	
Galactose	
Arabinose	
Sorbitol	
Xylose	
Maltose	
Disaccharide	
Cellobiose	2x D-Glucose; β-1,4-glycosidische Bindung
Sucrose (Saccharose)	D-Glucose und D-Fructose; α,β-1,2
Polysaccharide/Hemicellulosen	
Cellulose (Filterpapier Whatman No. 1, Avicel)	β-D-Glucose-Moleküle, β-D-(1,4)
Phosphoric acid swollen cellulose (PASC)[1]	β-D-Glucose-Moleküle, β-D-(1,4)
Stärke	D-Glucose, α-D-(1,6), α-D-(1,4)
Alginsäure	α-L-Guluronsäure und β-D-Mannuronsäure, 1,4
Arabinogalactan (Lärche)	Galactose:Arabinose:Andere 81:14:5, ß-D-(1,3), ß-D-(1,6), ß-D-(1,5)
Chitosan	Glucosamin und N-Acetylglucosamin, ß-D-(1,4)
Gummi arabicum (Akazie)	L-Arabinose, D-Galactose, L-Rhamnose, D-Glucuronsäure
Inulin	Fructose, pro Molekül 1 endständige Glucose, β-(2,1)
Lichenan	β-D-Glucose, β-D-(1,3) & (1,4) [im Verhältnis 2:1]
Polygalacturonsäure	α-D-Galacturonsäure, α-D-(1,4)
Pullulan	Maltotriose (3 x α-D-Glucopyranose α-D-(1,4))-α-D-(1,6)- Maltotriose
Xylan (Birke, Hartholz)	β-D-Xylopyranose, β-D-(1,4)
Xylan (Lärche, Weichholz)	β-D-Xylopyranose, β-D-(1,4)
Pectin	Heterosaccharid, >75 % D-Galacturonsäure, α-D-(1,4), >70 % verestert
Komplexe Substrate	
Maissilage[2]	
Zuckerrübenschnitzel[3]	

[1] Durch Behandlung der Cellulose mit 85 %iger Phosphorsäure und anschließendes Waschen mit Wasser werden die Wasserstoffbrückenbindungen zwischen den einzelnen Cellulosemolekülen und damit zum Teil die kristalline Struktur der Cellulose aufgelöst. Die amorphe Cellulose verteilt sich homogener im Medium und ist für den enzymatischen Abbau leichter zugänglich (Wood 1988).

[2] Maissilage besteht überwiegend aus Stärke (> 30%) und pflanzlichen Zellwandbestandteilen (17 – 20%). Um leicht abbaubare, wasserlösliche Bestandteile zu entfernen, wurde die gemahlene Maissilage (maximale Faserlänge 5 mm) mit entsalztem Wasser gewaschen und anschließend bei 60 °C getrocknet.

[3] Zuckerrübenschnitzel bestehen vor allem aus Zellwandkohlenhydraten, darunter sind in erster Linie Pectine (ca. 25 % der Trockenmasse) und Hemicellulosen vertreten.

Soweit nichts anderes angegeben ist, wurden jeweils 0,5% (w/v) der entsprechenden Kohlenstoffquelle in das gewünschte Medium gegeben. Lösliche Substrate wurden steril filtriert und nach dem Autoklavieren zugegeben.

2.4.4 Sonstige Additive/Zusätze

Zusätzlich zu dem Basismedium und Kohlenstoffquelle wurde das Anzuchtmedium teilweise auch durch weitere Zusätze variiert. Alle Additive, die während dieser Arbeit den Medien zugesetzt wurden, sind in folgender Tabelle aufgelistet.

Tabelle 17: Sonstige Zusätze für Kultivierungsmedien

Medienzusätze	Endkonzentration
Zellfreier Extrakt von *Geobacillus toebii* DSM 1550 (Bae et al. 2005)	1% (v/v)
Gärrest	5-30% (v/v); steril filtriert 0,5-1% (v/v)
Rumensaft	5-30% (v/v); steril filtriert 0,5-1% (v/v)
Vancomycin	0,25 – 50 µg/ml
Chloramphenicol	0,25 – 50 µg/ml
Ampicillin	100 µg/ml
Kanamycin	25 µg/ml
IPTG	1 mM
X-Gal	40 µg/ml

Zur Präparation des zellfreien Extrakts wurden das Zellpellet einer 100 ml Kultur von *Geobacillus toebii* 2x mit 50 mM Phosphatpuffer gewaschen. Anschließend wurde das Pellet mit 1mg/ml

Lysozym verdaut und nach einen zusätzlichen Aufschluss mit Ultraschall (4 Minuten; 60/2,5) erneut abzentrifugiert. Der Überstand wurde schließlich sterilfiltriert (0,22 µm Porengröße).

Der Gärrest für die Kultivierungen stammte aus diversen Biogasanlagen, die während dieser Arbeit beprobt wurden (Tabelle 17). Der Rumensaft stammte aus fistulierten Kühen des Lehrstuhls für Tierernährung (TUM).

2.5 BAKTERIENSTÄMME

Die nachfolgende Liste umfasst nur Bakterienstämme, die von externen Quellen bezogen wurden oder nicht während dieser Arbeit isoliert wurden. Die Stämme, welche während dieser Arbeit isoliert wurden, werden im Ergebnisteil gesondert aufgeführt.

Tabelle 18: In dieser Arbeit verwendete Bakterienstämme, die nicht im Laufe dieser Arbeit isoliert wurden

Stamm	Quelle
Escherichia coli Top 10	Life Technologies
E. coli Top 10 F´	Life Technologies
E. coli DH10B	Life Technologies
E. coli BL21 Star	Life Technologies
Clostridium stercorarium DSM 8532	DSMZ
Clostridium thermocellum ATCC 27405	ATCC
C. thermocellum DSM 2360	DSMZ
C. thermocellum DSM 7072	DSMZ
C. thermocellum DSM 4150	DSMZ
C. thermocellum DSM 1313	DSMZ
C. thermocellum BC1	(Koeck et al. 2013)
C. thermocellum F7	Russische Stammsammlung VKPM
C. thermocellum 5G	(Koeck et al. 2015a)
C. thermocellum T2	(Koeck et al. 2015a)
C. thermocellum NB2	(Koeck et al. 2015a)
Bacillus thermoamylovorans DSM 13307	DSMZ
Geobacillus toebii DSM 1550	DSMZ

2.6 METHODEN ZUR ISOLIERUNG VON BAKTERIEN

2.6.1 ISOLIERUNGSSTRATEGIEN

2.6.1.1 Isolierungstrategie 1:

Bei Strategie 1 erfolgte eine einfache Anreicherung des Inokulums auf dem gewählten Substrat, bis dieses deutlich abgebaut bzw. angegriffen wurde. In der Regel handelte es sich hierbei um einen Filterpapierstreifen, dessen Abbau z.B. optisch verfolgt werden kann. Bei löslichen Substraten wurde eine deutliche Gasbildung oder Trübung des Mediums abgewartet. Anschließend wird aus der ersten Anreicherung frisches Medium (1:20) inokuliert. Dieser Vorgang wird 3-4-mal wiederholt.

2.6.1.2 Isolierungstrategie 2 (Schoenborn et al. 2004):

Ein zweiter Ansatz ist die Isolierung aus Flüssigmedium mittels mehrfacher Verdünnungsreihen. 1 ml Inokulum aus der ausgewählten Biogasanlage wurde in 50 ml NaCl Lösung (0,2 M) suspendiert und bis 10^{-8} verdünnt. Jede Verdünnung wurde in frisches Medium mit der entsprechenden Kohlenstoffquelle (in der Regel Filterpapier) überimpft. Die Kulturen wurden so lange inkubiert, bis das Filterpapier deutlich angegriffen war. Die höchste Verdünnungsstufe mit sichtbarem Filterpapierabbau wurde wiederum bis 10^{-8} in demselben Medium verdünnt. Nach vier Wiederholungen der Verdünnungsreihe wurde die resultierende Anreicherungskultur erneut verdünnt, in verschiedene Verdünnungsstufen ausplattiert und mit Cellulose- oder PASC-Topagar überschichtet.

2.6.1.3 Isolierungsstrategie 3 (Zhang et al. 2003):

Ein dritter Ansatz ist die direkte Isolierung von Agarplatten ohne vorherige Anreicherung in Flüssigmedium. Das Inokulum wurde dafür in sterilem Phosphatpuffer (20 mM, 1:100) ausgeschüttelt und anschließend bei 100 g für fünf Minuten zentrifugiert. Das Pellet wurde schließlich in steriler NaCl-Lösung (0,2 M) gewaschen und bis 10^{-8} verdünnt. Verschiedene Verdünnungsstufen wurden danach direkt ausplattiert. Nach Inkubation bei der gewählten Temperatur wurden Kolonien mit deutlichen Lysehöfen im Cellulose-Topagar erneut in verschiedenen Verdünnungsstufen ausplattiert oder direkt Verdünnungsausstriche angefertigt.

2.6.1.4 “In Sacco” Anreicherung

Nylonsäckchen mit 50 µm Porengröße wurden mit 2 g zerkleinertem Filterpapier (ca. 3 cm lang und 0,5 cm breit) gefüllt und mit Hilfe eines speziellen Keimträgers direkt in die Labor Fermenter BM1T und BM2T (Tabelle 5) eingebracht. Mit Hilfe der für Bakterien permeablen Nylonsäckchen sollte eine selektive Anreicherung „in situ“ von cellulolytischen Bakterien im anaeroben Milieu des Biomassefermenters erreicht werden. Diese Methode wurde ursprünglich zu Bestimmung der Verdaulichkeit von Futtermitteln direkt im Pansen von Wiederkäuern entwickelt (Mohamed and

Chaudhry 2008). Das Futtermittel wird zu diesem Zweck in einen durchlässigen unverdaulichen Beutel (Polyester oder Nylongewebe) mit definierter Porenweite eingeschlossen und zum Abbau in den gewünschten Darmabschnitt eingebracht. Die Säckchen zur Anreicherung cellulolytischer Bakterien wurden nach einer gewissen Inkubationszeit mit möglichst wenig Sauerstoffkontakt aus dem Fermenter entfernt und anschließend mit 50 ml isotonischer Lösung gespült, um Reste des Fermenterinhalts und unspezifische anhaftende Bakterien zu entfernen. Das „anverdaute“ Filterpapier konnte dann zur weiteren Anreicherung und Isolierung von cellulolytischen Bakterien in diverse Medien überführt werden.

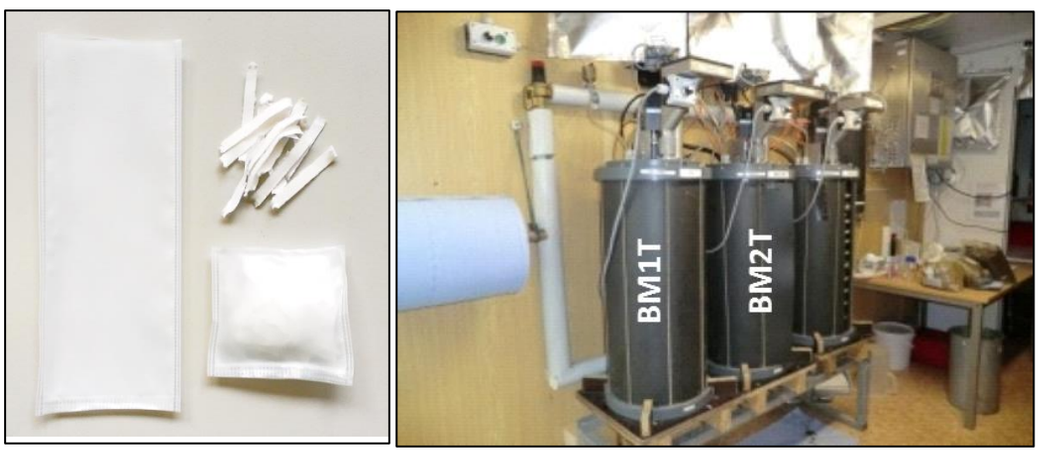


Abbildung 9: links: Nylonsäckchen für die „in sacco“ Versuche, rechts: LfL Laborfermenter BM1T und BM2T

Neben der Wahl der Isolierungsstrategie gibt es weitere Möglichkeiten, um die Diversität der Isolate zu erhöhen. Zum einen hängt das Isolat-Spektrum stark von der Zusammensetzung des Wachstumsmediums ab (Pham and Kim 2012). Das gewählte Grundmedium für die Zielorganismen wurde z.B. mit Zusätzen wie Gärrest oder Rumensaft versehen, um die Nährstoffbedingungen im natürlichen Habitat zu simulieren oder mit Antibiotika versehen, um bestimmte Organismengruppen am Wachstum zu hindern. Eine weitere Möglichkeit zur Diversifizierung bietet die Variation der Wachstumsbedingungen. In dieser Arbeit wurden vor allem die Temperatur, Inkubationszeit und der pH-Wert des Mediums variiert. Des Weiteren konnte durch Vorbehandlung das Inokulums, z.B. durch Pasteurisierung, eine weitere Variation erreicht werden.

2.6.2 Reinigung und Identifizierung der Isolate

Von den aus verschiedenen Strategien resultierenden Anreicherungen auf Agarplatten wurden Kolonien mit deutlichem Substratabbau und/oder unterschiedlicher Morphologie ausgewählt. Davon wurden mindestens drei Verdünnungsausstriche angefertigt. Die resultierenden Reinkulturen wurden erneut in Flüssigmedium überimpft, um den Substratabbau (meist Filterpapier) erneut zu überprüfen. Die nachfolgende Identifizierung und phylogenetische Zuordnung der Isolate erfolgte auf Basis der 16s rRNA Gensequenz, welche mit spezifischen

Oligonucleotid-Primern (Tabelle 22) von der genomischen DNA der Isolate durch PCR gewonnen wurde. Die Sequenzen werden mit der NCBI- und der RDP-Datenbank verglichen und bioinformatisch ausgewertet. Die entsprechenden Methoden sind in den Kapiteln 2.8 und 2.10 genauer beschrieben. In Abbildung 10 sind Isolierungs-, Reinigungs- und Identifikationsschritte für ein neues Isolat illustriert und zusammengefasst.

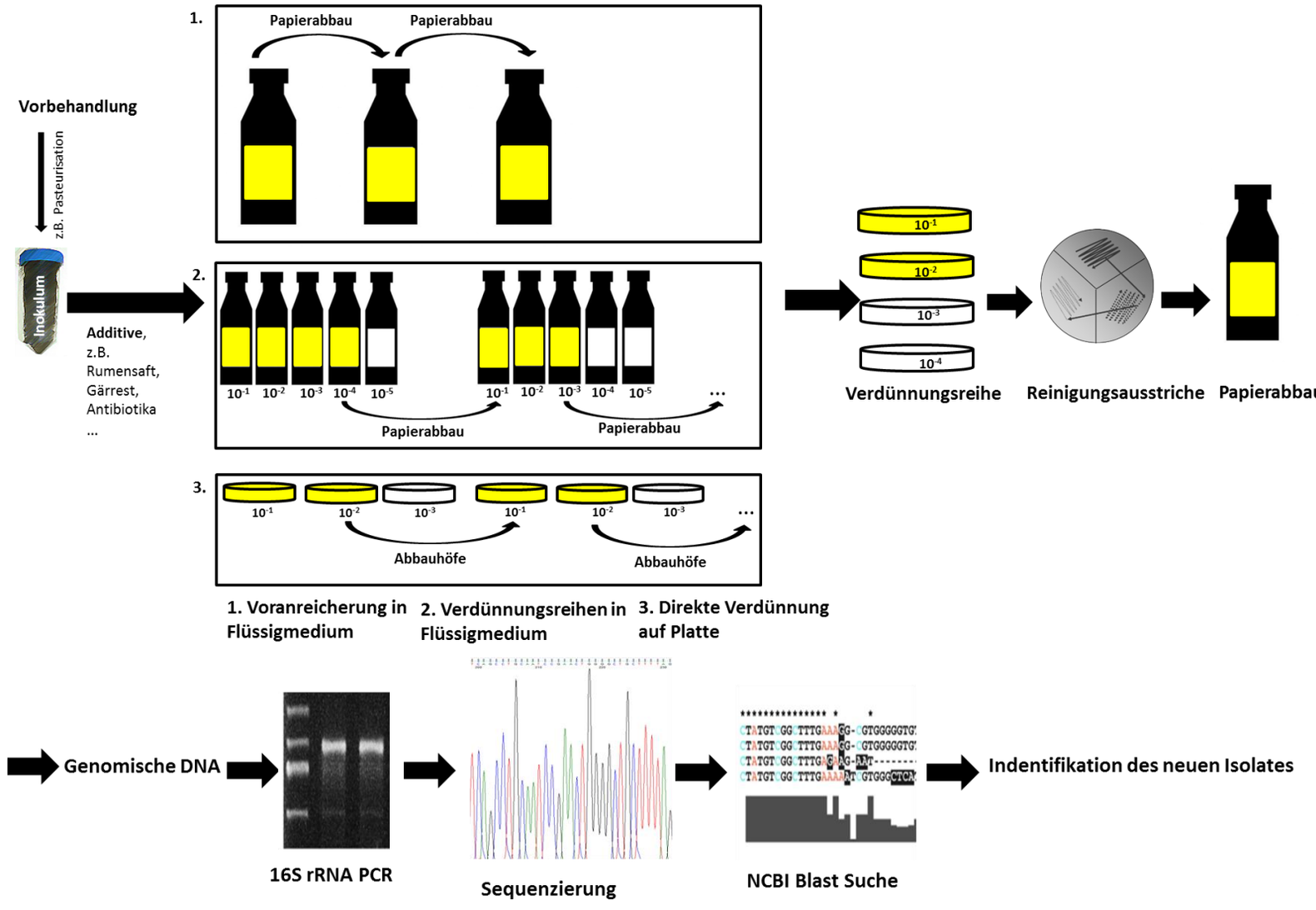


Abbildung 10: Zusammenfassung der Isolierungs, Reinigungs und Identifikationsschritte für ein neues Isolat. Eine gelbe Färbung ist jeweils gleichzusetzen mit Substrat bzw. Filterpapierabbau

2.7 Physiologische Charakterisierung bakterieller Isolate

2.7.1 Anthron Assay (Dische 1962)

Zur quantitativen Bestimmung des Celluloseabbaus einer Kultur wurden die nach der Inkubation im Medium verbliebenen Kohlenhydrate mittels Anthron Assay gemessen. Durch Aufkochen mit Schwefelsäure werden alle Kohlenhydrate hydrolysiert und zu Furfural (Pentosen) und Hydroxymethylfurfural (Hexosen) umgesetzt. Diese bilden bei Zugabe von Anthron Reagenz einen grünen Farbkomplex aus. Mit Hilfe der vorher erstellten Glucose-Kalibriergerade kann aus der Absorption bei 620 nm die Glucose-Konzentration berechnet werden. Für den Assay wurden 20 mg Avicel in 10 ml Medium eingewogen (0,2% (w/v)) und die frisch überimpfte Kultur für zwei bis sieben Tage inkubiert. Zur Messung wurde ein Aliquot (0,1 bis 0,5 ml) entnommen und abzentrifugiert. Das Pellet wurde mehrmals in sterilem Wasser ausgeschüttelt, um lösliche Zucker zu entfernen und schließlich in einem ml dest. Wasser aufgenommen. Auf die Probe wurden 4 ml Anthron Lösung und 2 ml 75%ige Schwefelsäure gegeben. Diese Mischung wurde anschließend für 15 min bei 100 °C aufgekocht und vor der Absorptionsmessung wieder abgekühlt.

Tabelle 19: Anthron Assay

Anthron	0,5 g
Ethanol (absolut)	10 ml
Ad H_2SO_4 (75% v/v)	250 ml

2.7.2 Zellproteinbestimmung mit CHAPS (Reveneau et al. 2003)

Die Bestimmung von Bakterienwachstum über die Gesamtzellproteinkonzentration wurde vor allem bei trüben Medien eingesetzt, bei denen eine visuelle Beurteilung des Wachstums nicht möglich war. Die Zellen wurden dazu abzentrifugiert, das Pellet zweimal mit isotonischer NaCl Lösung (0,3 M) gewaschen und in 1% (w/v) CHAPS (pH 11) für 20 min gekocht. Die Proteine im Überstand wurde mit Hilfe des Bradford Assays (2.9.2) bestimmt.

2.7.3 MALDI-TOF MS

Mittels MALDI-TOF MS wurden zur Identifizierung von neuen Isolaten spezifische Zellproteinprofile generiert. 1 ml frisch gewachsene Kultur wurde abzentrifugiert und das Pellet in 500 µl Ethanol resuspendiert und zur Analyse geschickt. Alle MALDI-TOF MS Analysen wurden durch die RIPAC-LABOR GmbH (Potsdam-Golm, Germany) durchgeführt. Hierbei wurde zunächst das zu untersuchende Isolat mit einer Kristallisationsmatrix vermischt. Mithilfe eines UV-Lasers wird diese Mischung dann verdampft, wobei ionisierte Partikel unterschiedlicher Größe und Ladungen entstehen. Die ionisierten Partikel, insbesondere ribosomale Proteine mit wachstums-

unabhängiger Zusammensetzung, werden in einem elektrischen Feld in eine vorgegebene Richtung beschleunigt. Abhängig von dem Masse/Ladungsverhältnis (*m/z*) bewegen sich die Partikel mit unterschiedlichen Geschwindigkeiten. Dies ermöglicht die getrennte Detektion dieser Partikel am Ende der Flugstrecke (Abbildung 11). Das Protein-Massen-Profil reicht generell von 1.000 *m/z* bis 20.000 *m/z* (Sauer and Kliem 2010). Am Detektor wird neben der Partikelgröße (x-Wert) auch die Häufigkeit eines Partikels (y-Wert) ermittelt (Abbildung 11). Das resultierende Spektrum ist für jeden Mikroorganismus einzigartig (Sauer et al. 2008). Über den Vergleich der erhaltenen Spektren mit Referenzspektren in der Datenbank SARAMI™ kann eine Identifizierung des Isolates erfolgen. Für die Klassifikation und den Vergleich verschiedener Isolate wurde eine Clusteranalyse durchgeführt und die Ergebnisse in einem Dendrogram visualisiert. Hierbei liegen unterschiedliche Spezies vor, wenn die Abzweigung zwischen den beiden Isolaten im Dendrogram bei > 70% liegt und unterschiedliche Gattungen, wenn die Abzweigung bei > 50% liegt.

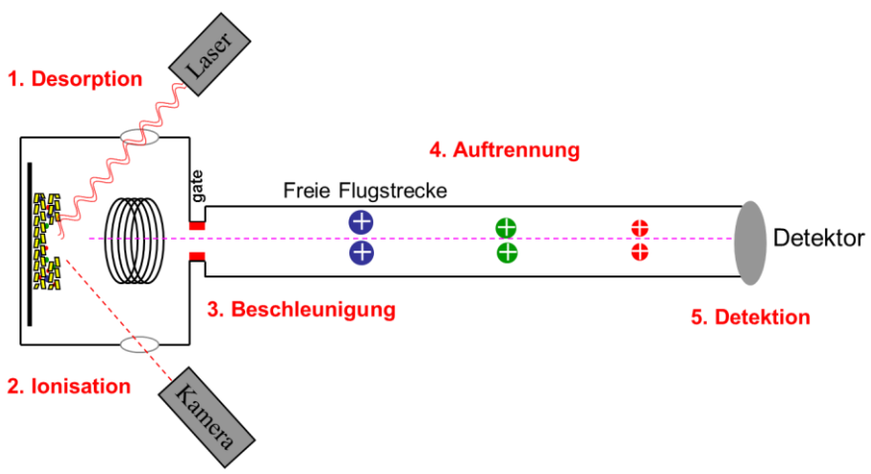


Abbildung 11: Schematische Darstellung der Beschleunigung ionisierter Partikel im elektrischen Feld der MALDI-TOF MS (Quelle RIPAC GmbH)

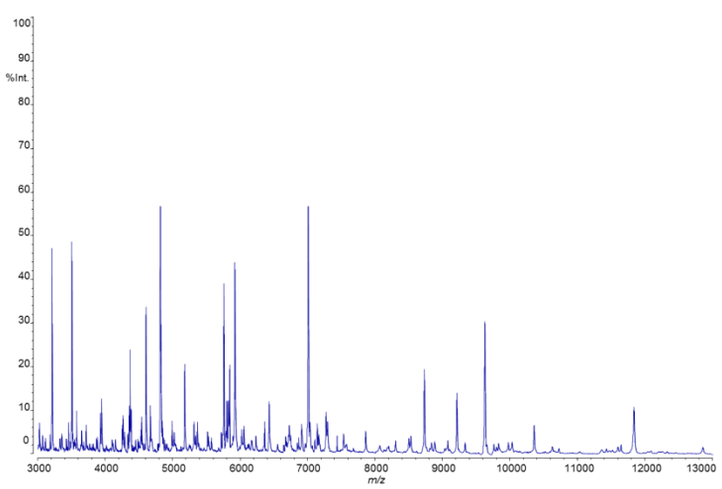


Abbildung 12: Beispielspektrum des MALDI-TOF MS Proteinprofils von Clostridium stercorarium DSM 8532

2.7.4 GASCHROMATOGRAPHIE

Zur Bestimmung der Stoffwechselprodukte neuer Isolate wurden flüchtige Säuren und Alkohole mit Hilfe eines Gaschromatographiesystems (GC-2010; Shimadzu, Kyoto, Japan) gemessen. Dieses Gerät war mit einen Split/Splitless Injektor (0,5 µl Injektionsvolumen) und einem Flammen-Ionisationsdetektor ausgestattet. Für die Auftrennung der Probe wurde eine Phese-Stabilwax-DA Kapillarsäule (60 m Länge, 0,32 mm Durchmesser, 0,5 µm Filmdicke; Firma RESTECK, Bad Homburg, Deutschland) verwendet. Als Trägergas kam Stickstoff mit einer Fließgeschwindigkeit von 50 cm/s zum Einsatz. Für eine bessere Auflösung wurde mit einem Temperaturgradienten gearbeitet, bei dem die Säulentemperatur innerhalb der ersten 12 min von 70 °C auf 260 °C erhitzt wird und dann zwei Minuten konstant bei 260 °C gehalten wird.

2.7.4.1 Probenvorbereitung und Auswertung

Aus frisch gewachsenen Kulturen wurden 2 ml entnommen und 20 min bei 13.000 rpm zentrifugiert. Um noch verbliebene Feststoffe zu entfernen wurde der Überstand über Nacht bei -20 °C gefroren und anschließend noch einmal zentrifugiert. 450 µl der Probe wurden mit 50 µl Isobutanol (0,5 %) als internem Standard gemischt. Zur Erstellung des Standardchromatogramms wurde bei jedem Lauf eine definierte Mischung aus Säuren und Alkoholen gemessen. Dies ermöglicht die Zuordnung der Peaks in den Proben zu bekannten Substanzen. In Tabelle 20 sind die Retentionszeiten der Säuren und Alkohole, die als externe Standards verwendet wurden, gelistet.

Tabelle 20: Tabelle der Retentionszeiten verschiedener Substanzen im Gaschromatographen

Substanz	Retentionszeit [min]	Substanz	Retentionszeit [min]
Aceton	4,5	Butanol	8,3
Methanol	5,2	Essigsäure	10,9
Ethanol	5,6	Propionsäure	11,5
Isopropanol	6,9	Buttersäure	12,2
Isobutanol	7,6		

Die Höhe und Fläche des Peaks ist der eingespritzten Stoffmenge proportional. Unbekannte Substanzmengen können durch Vergleich der Peakflächen mit bekannten Konzentrationen bestimmt werden. Nach erfolgter Analyse wird das Ergebnis des internen Standards mit der erwarteten Menge verglichen (Bestimmung der Wiederfindungsrate). Hat der interne Standard durch das Verfahren seine Konzentration verändert, wird angenommen, dass sich die Konzentration des Analyten in gleicher Weise verändert hat. Das Ergebnis des Analyten kann mit der Wiederfindungsrate des internen Standards mit folgender Formel korrigiert werden. C_A ist dabei die gesuchte Analytkonzentration, C_S ist die eingewogene Konzentration an Standard und A bezeichnet die Fläche des Peaks im Chromatogramm.

$$C_A = C_S \cdot (A_P / A_S)$$

Formel 1: Berechnung der Analytkonzentration

2.7.5 Analyse des Fettsäuremusters prokaryontischer Zellen

Die Analyse des Fettsäuremusters von prokaryontischen Zellen wurde vom Identifikationsservice der DSMZ (Braunschweig) durchgeführt. Hierzu wurden Zellen in der exponentiellen Wachstumsphase geerntet und als getrocknetes Pellet zur Analyse geschickt. Die Methylester wurden extrahiert und nach dem Standardprotokoll des "Sherlock Microbial Identification" Systems (Version 6.1; MIDI) analysiert.

2.7.6 Ermittlung des optimalen pH-Wertes und Salzgehalts, sowie der optimalen Temperatur für die Kultivierung neuer Isolate

Um die Optima von pH, Temperatur und Salzkonzentration zu bestimmen, wurden die Kulturen in frisches Medium überimpft und für sieben Tage unter den zu testenden Bedingungen inkubiert. Bei mesophilen Isolaten wurde der Temperaturbereich 25 – 45 °C getestet, bei thermophilen 37 – 70°C, wobei die Temperatur jeweils in 5 °C Schritten erhöht wurde. Die pH- und Salztoleranz wurde anschließend bei optimaler Temperatur getestet, dabei wurde der pH-Wert nach dem Autoklavieren mit Temperaturkompensation gemessen. Die Salztoleranz wurde mit 0,2 - 4 % NaCl (w/v) im Medium untersucht, wobei die Konzentration in 0,2 % Schritten erhöht wurde. Die pH Toleranz wurde zwischen pH 5,0 und 9,5 in Schritten von 0,5 Einheiten bestimmt. Zur Beurteilung des Wachstums der Kulturen wurde zum einen die Zellproteinkonzentration (vgl. 2.7.2) und zum anderen die optische Dichte der Kultur bei 600 nm bestimmt und jeweils mit der Negativkontrolle verglichen. Alle Werte wurden in Duplikaten gemessen.

2.7.7 Substratspektrum

Bei optimalen pH- und Temperatur-Werten wurde getestet welche Kohlenstoffquellen die Isolate verwerten können. Hierzu wurden jeweils 0,5 % (w/v) von diversen Mono- Di- und Polysacchariden (Tabelle 16) in GS2 Medium gegeben und dieses für sieben Tage mit dem entsprechenden Inokulum inkubiert. Positives Wachstum wurde visuell anhand der Trübung des Mediums im Vergleich zur nicht-inokulierten Negativkontrolle, mit derselben Kohlenstoffquelle, beurteilt. Zur Bestätigung des Ergebnisses wurde die Kultur nochmals in dasselbe Medium überimpft und die Trübung wiederum beurteilt. Zudem wurde zum Teil zur Überprüfung der Speziesidentität eine Sequenzanalyse des 16S rRNA Genes durchgeführt (vgl. 2.8). Außerdem wurde die Zellproteinkonzentration von positiven Wachstumsreaktionen bestimmt (vgl. 2.7.2).

2.7.8 *Abbau eines natürlichen Substrates (Maissilage)*

Als „natürliches“ Substrat der Biogasanlage wurde Maissilage gewählt. Um die Verfälschung der Abbauversuche durch lösliche Bestandteile der Maissilage (Proteine, Säuren etc.) zu vermeiden, wurde die Maissilage zunächst mit entsalztem Wasser gewaschen und anschließend wieder getrocknet. Etwa 0,5 g gewaschene, gemahlene (Faserlänge max. 5 mm) Maissilage wurde in 50 ml Medium dispergiert. Nach der Inokulation wurde der Ansatz für sieben Tage bei der optimalen Wachstumstemperatur der Kulturen inkubiert. Um den Biomasseabbau zu bestimmen, wurde das exakte Trockengewicht der Maissilage vor und nach der Inkubation bestimmt. Um die Zellen zu entfernen, wurde die Silage in Filterbeutel gefüllt und zweimal mit entsalztem Wasser gespült. Diese Filter wurden anschließend bei 70 °C getrocknet. Der Gewichtsverlust der Negativkontrolle durch den zweiten Waschvorgang wurde von dem Ergebnis subtrahiert.

2.7.9 *Mikroskopie und Elektronenmikroskopie*

Die Mikroskopie diente in erster Linie zur Visualisierung der Zellmorphologie neuer Isolate und zur Beobachtung von Zellmobilität, Zellgröße und Sporulation. Während dieser Arbeit wurde in erster Linie mit einem Phasenkontrastmikroskop der Firma Zeiss (Göttingen) gearbeitet. Für die hoch auflösende Feldemissions-Rasterelektronenmikroskopie, wurden frische Kulturen mit 2,5% (w/v) Glutardialdehyd in Kakodylatpuffer fixiert. Mehrere Tropfen der fixierten Probe wurden auf beschichtete Objektträger aufgebracht, mit einem Deckglas versehen und anschließend in flüssigem Stickstoff schock gefroren. Nach der Entfernung des Deckglases wird die Probe erneut in 2.5 %igem (w/v) Glutardialdehyd fixiert, gewaschen, mit 1% Osmiumtetroxid fixiert, mit Puffer und destilliertem Wasser gewaschen und stufenweise in Acetonlösungen dehydriert. Nach der Überführung in flüssiges CO_2 folgte eine „Kritische Punkt Trocknung“ Auf einer speziellen Halterung wurden die Proben schließlich mit Hilfe eines „Magnetron sputter coater“ mit 3-nm Platin beschichtet und mit einem Zeiss Auriga scanning electron microscope (Göttingen) untersucht (Wanner et al. 1989). Die Aufnahmen mit dem Rasterelektronenmikroskop wurden von Professor Dr. Gerhard Wanner (Biozentrum der LMU, Planegg) angefertigt.

Tabelle 21: Kakodylatpuffer

Kakodylat	75 mM
NaCl	25 mM
$MgCl_2 \times 2H_2O$	2 mM

2.7.10 *ENTEROTUBE II*

Der Enterotube II ist ein gebrauchsfertiges Testsystem von BD Diagnostic Systems (Heidelberg) zum gleichzeitigen Nachweis von 15 verschiedenen biochemischen Reaktionen der Bakterien. Alle Kammern werden in einem Arbeitsgang gleichzeitig beimpft und nach etwa 24-stündiger Inkubation bei optimaler Wachstumstemperatur ausgewertet. Die Auswertung erfolgte nach Herstellerangabe (Abbildung 13).

Abbildung 13: Auswertung des Enterotube II Testsystems (Quelle: BD Diagnostic Systems)

2.7.11 *GRAM-FÄRBUNG*

Die Gram-Färbung ermöglicht es, Bakterien anhand des unterschiedlichen Aufbaus ihrer Zellwände in zwei große Gruppen einzuteilen: Grampositive und gramnegative Bakterien. Die Färbung wurde wie von Johnson et al. beschrieben durchgeführt (Johnson et al. 1995).

2.8 GENTECHNISCHE METHODEN

2.8.1 PRIMER

Die verwendeten Primer wurden von der Firma Eurofins (Ebersberg) synthetisiert, in sterilem H_2O auf eine Konzentration von 100 pmol/µl verdünnt und bei -20 °C gelagert.

Tabelle 22: Liste aller in dieser Arbeit verwendeten Oligonucleotid-Primer. F steht jeweils für den Vorwärtsprimer und R für den Rückwärtsprimer. T_m bezeichnet die Schmelztemperatur

Nr.	Name	Sequenz 5´-> 3´	Verwendung	T_m °C
1	Cth f	TTTACCGGAAGTATATCCTAG	Spezifischer Nachweis von *C. thermocellum* 16S rRNA Genen	58
2	Cth r	AATCATCTGCCCCACCTTC	Spezifischer Nachweis von *C. thermocellum* 16S rRNA Genen	58
3	632F	GTKAAGTCGTAACAAGG	Spacer zwischen *C. thermocellum* 16S und 23S Gen	49,2
4	118R	GTBCCCCATTCGGA	Spacer zwischen *C. thermocellum* 16S und 23S Gen	49,7
5	S-layer F	GGAGTGAATGTCATAGCTCC	S-layer Protein von *C. thermocellum*	57,3
6	S-layer R	GAGTTGGTGTAGGCAATTGC	S-layer Protein von *C. thermocellum*	57,3
7	Bth r	GGATTCGCTTATGCGTCG	Spezifischer Primer für *Bacillus thermoamylovorans* 16S rRNA Gene	56,0
8	Bth f	GATTATCTTTCCGCCTGGAG	Spezifischer Primer für *Bacillus thermoamylovorans* 16S rRNA Gene	57,3
9	M13 f	GTAAAACGACGGCCAG	Sequenzierprimer für gängige Klonierungssystem (z.B. TOPO TA)	51,7
10	M13 r	CAGGAAACAGCTATGAC	Sequenzierprimer für gängige Klonierungssystem (z.B. TOPO TA)	50,4
11	Cth TP vorne f	CAGCCCAAATAAGTTAGAATG	Cth Transposon Nachweis	54
12	Cth TP hinten f	CAGAGTGCACAGTTGTAAAG	Cth Transposon Marker (Reverse komplementär zu 13)	55,3
13	Cth TP hinten r	CTTTACAACTGTGCACTCTG	Cth Transposon Nachweis	55,3
14	Cth TP vorne r	CATTCTAACTTATTTGGGCTG	Cth Transposon Marker ((Reverse komplementär zu 11)	54
15	616F	AGAGTTTGATYMTGGCTC	Eubakterielles 16Sr RNA Gene (Loy et al. 2005)	55
16	630R	CAKAAAGGAGGTGATCC	Eubakterielles 16Sr RNA Gene (Loy et al. 2005)	55
17	T3/55 GH9 f	CTTTAAGAAGGAGATATACATATGAA TGTGCCTAAAAAGGCGG	Klonierung T3/55 GH9	69,4
18	T3/55 GH9 r	CTCAGTGGTGGTGGTGGTGGTGCT CGAGCGGTTCTATTC	Klonierung T3/55 GH9	>75
19	T3/55 GH48 f	CTTTAAGAAGGAGATATACATATGCT TTCCCTGGGACAAAAGC	Klonierung T3/55 GH48	70,4
20	T3/55 GH48 r	CTCAGTGGTGGTGGTGGTGGTGCT CGAGCGGCTCAATACCGAAT	Klonierung T3/55 GH48	>75
21	T3/55 GH5 f	CTTTAAGAAGGAGATATACATATGAA AAAATTTGATGGATTTTTAGC	Klonierung T3/55 GH5	65,9
22	T3/55 GH5 r	CTCAGTGGTGGTGGTGGTGGTGCT CGAGATTATTTTTTTTACTTAC	Klonierung T3/55 GH5	73
23	pET24 f	TCAAGACCCGTTTAGAGGCC	Primer pET Vektoren	53
24	pET24 r	ACGACTCACTATAGGGGAAT	Primer pET Vektoren	53
25	T3.55 f	GTACGTAAAGCTCTATCAGCA	Spezifische Primer für T3/55 16S rRNA Gene	55,9
26	T3.55 r	GTCCGGTTAAGGACCGGT	Spezifische Primer für T3/55 16S	58,2

			rRNA Gene	
27	T355_1502 f	TTAAGAAGGAGATATACATATGAAG GCAAAACTGACTATAAATAAAG	Klonierung T355_1502 (X1)	66
28	T355_1502 r	CAGTGGTGGTGGTGGTGGTGCTCG AGTATTTGAAAACGTATCACATTCC	Klonierung T355_1502 (X1)	>75
29	T355_1501 f	TTAAGAAGGAGATATACATATGAAA AAGCAACTGCAATTATCAAG	Klonierung T355_1501 (X2)	66
30	T355_1501 r	CAGTGGTGGTGGTGGTGGTGCTCG AGCTCAAGACCGTAAAATTCTTTAG	Klonierung T355_1501 (X2)	>75
31	T355_1505 f	TTAAGAAGGAGATATACATATGAAA AAACAAGGATTTAATCCGTATC	Klonierung T355_1505 (X3)	66
32	T355_1505 r	CAGTGGTGGTGGTGGTGGTGCTCG AGTTCCAATGTAAATGAAGCCAGG	Klonierung T355_1505 (X3)	>75
33	T355_810 f	TTAAGAAGGAGATATACATATGAAG AAAGTATCGGTTTTAATTAATGC	Klonierung T355_810 (X4)	66
34	T355_810 r	GTGGTGGTGGTGGTGGTGCTCGAG ATTCAATTGAATTGATATTACCGAG	Klonierung T355_810 (X4)	>75
35	T355_1528 f	TTAAGAAGGAGATATACATATGGCC CAGACAATTTACAACAAT	Klonierung T355_1528 (X5)	67
36	T355_1528 r	CAGTGGTGGTGGTGGTGGTGCTCG AGATTGATTTCCAGATAATCTATG	Klonierung T355_1528 (X5)	>75
37	T355_1554 f	TTAAGAAGGAGATATACATATGGCG GAAGTATCCGCAG	Klonierung T355_1554 (X6)	69
38	T355_1554 r	CAGTGGTGGTGGTGGTGGTGCTCG AGTTGAACAAGTTGGCCGTC	Klonierung T355_1554 (X6)	>75
39	T355_2049 f	TTAAGAAGGAGATATACATATGGAA AAAAGGGTAAAAAATCCTA	Klonierung T355_2049 (X7)	65
40	T355_2049 r	CAGTGGTGGTGGTGGTGGTGCTCG AGATCAATGTTTTTGTAATATATCC	Klonierung T355_2049 (X7)	74
41	T355_2050 f	TTAAGAAGGAGATATACATATGGAT AATAAGTATAAACATCGCA	Klonierung T355_2050 (X8)	65
42	T355_2050 r	CAGTGGTGGTGGTGGTGGTGCTCG AGTTCAAGACATACGGTTACTG	Klonierung T355_2050 (X8)	>75
43	T355_269 f	TTAAGAAGGAGATATACATATGAATA CATACTGGCAGTATTGG	Klonierung T355_269 (X9)	67
44	T355_269 r	CAGTGGTGGTGGTGGTGGTGCTCG AGCCATACGGTTACGTTAGC	Klonierung T355_269 (X9)	>75
45	T355_698 f	TTAAGAAGGAGATATACATATGGTG CATGACATAACCCTTG	Klonierung T355_698 (X10)	68
46	T355_698 r	CAGTGGTGGTGGTGGTGGTGCTCG AGTTCAATACTGTAACCTGAATC	Klonierung T355_698 (X10)	>75
47	T355_905 f	TTAAGAAGGAGATATACATATGAAA GGAAAAAATATTTCAGACTT	Klonierung T355_905 (X11)	64
48	T355_905 r	CAGTGGTGGTGGTGGTGGTGCTCG AGATTAAATCCTGCATGATCATAA	Klonierung T355_905 (X11)	>75
49	T355_929 f	TTAAGAAGGAGATATACATATGAGC AAACAAGCAAAGCAAAATG	Klonierung T355_929 (X12)	67
50	T355_929 r	CAGTGGTGGTGGTGGTGGTGCTCG AGATTTTTATACGGATGAATTGTTTG	Klonierung T355_929 (X12)	>75

2.8.2 VEKTOREN

Die kommerziell erhältlichen Vektoren CloneJET™ und pCRII-Topo® (Thermo) wurden für die Klonierung von 16S rRNA Genfragmenten verwendet. Der Expressionsvektor pET24a(+) wurde zur heterologen Expression von Enzymen der Neuisolate genutzt. Dieser Vektor enthält ein Lac-Operon zur kontrollierten Induktion der Expression über IPTG und einen N-terminalen His-tag zur anschließenden Aufreinigung des Proteins mittels Affinitätschromatographie.

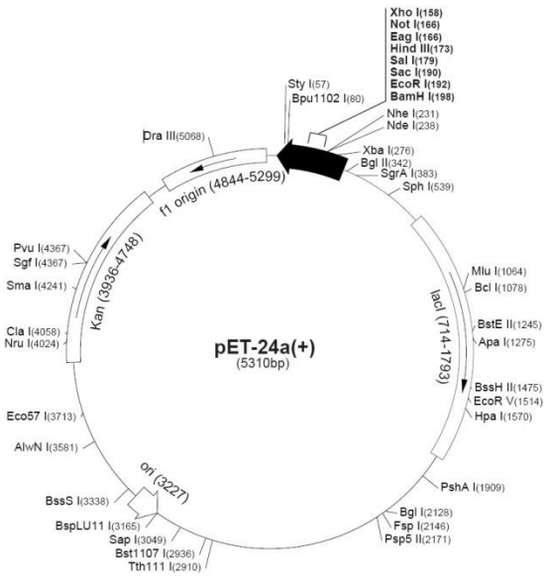


Abbildung 14: Plasmidkarte des pET24a(+) Vektors mit Schnittstellen der MCS

2.8.3 HERSTELLUNG KOMPETENTER ZELLEN

Zur Herstellung chemisch kompetenter *E. coli* DH10B, BL21 Star und TOP10 Zellen zur Transformation, wurde eine frische Vorkultur des gewünschten Stammes in LB Medium angezogen und anschließend 3 ml davon in 250 ml SOB Medium inokuliert. Die Zellen wurden bis zur einer optischen Dichte (OD_{600}) von 0,3 bei 37 °C geschüttelt und abzentrifugiert (3000 g, 4 °C, 10 min). Nach der Wiederaufnahme in 80 ml gekühltem CCMB80 Puffer (Tabelle 23) wurde 20 Minuten auf Eis inkubiert, wiederum abzentrifugiert und in 10 ml eisgekühltem CCMB80 Puffer resuspendiert. Die Zellen wurden auf Eis in 50 µl Aliquots aufgeteilt, sofort in flüssigem Stickstoff gefroren und bis zur weiteren Verwendung bei -80 °C gelagert.

Tabelle 23: CCMB80 Puffer

KOAc	10 mM
$CaCl_2$ x $2H_2O$	80 mM
$MgCl_2$ x $6H_2O$	10 mM
$MnCl_2$ x $4H_2O$	20 mM
10 % Glycerol (v/v)	
ddH_2O	1000 ml

2.8.4 ISOLIERUNG GENOMISCHER DNA

Zur klassischen Extraktion genomischer DNA wurde das Pellet von 50-100 ml einer frischen Kultur in 5 ml Lysepuffer (Tabelle 24) resuspendiert und mindestens eine Stunde bei 37 °C inkubiert. Nach der Zugabe von 500 µl einer 10 %igen SDS-Lösung und 1 mg/ml Proteinase K wurde weitere 2 Stunden bei RT inkubiert. Anschließend wurde 1/3 Volumen NaCl (5 M)

zugegeben. Zur Reinigung der DNA wurde 1/3 Volumen Chloroform zugegeben und 30 min ausgeschüttelt. Anschließend wurde für 10 min bei 5000 rpm zentrifugiert und die obere, wässrige Phase erneut mit Chloroform ausgeschüttelt. Dieser Schritt wurde so lange wiederholt, bis die wässrige Phase vollständig von der Proteinphase gereinigt war. Schließlich wurde die DNA durch Zugabe von 100 % Isopropanol gefällt und bei 5000 rpm und RT für 30 min abzentrifugiert. Danach wurde das Pellet mit eiskaltem 70 % Ethanol gewaschen und die DNA nach Verdunsten des Ethanols in sterilem, entsalzten H_2O gelöst. Die Lagerung erfolgte bei -20 °C.

Tabelle 24: Lysepuffer

TrisHCl	200 mM
EDTA	25 mM
NaCl	75 mM
Lysozym	100 mg
Ad ddH_2O	100 ml

Eine weitere Möglichkeit zur Gewinnung genomischer DNA stellt die Verwendung kommerzieller Kits dar. Genomische DNA aus Biogasfermenter- und Umweltproben wurde mit dem FastDNA® SPIN Kit for Soil (MP Biomedicals, Santa Ana, USA) nach Herstellerangaben extrahiert. Für gereinigte Isolate wurde der Presto Spin D Bug DNA Purification Kit (Molzym GmbH, Bremen) und der innuPREP DNA Mini Kit (AnalytikJena, Jena) nach Herstellerangaben verwendet. Für eine optimale Lyse grampositiver Bakterien wurde die Behandlung mit Lysepuffer auf eine Stunde verlängert. Der Vorteil dieser Methode ist die relativ einfache und schnelle Durchführung. Die DNA wird hochkonzentriert und hochmolekular (>50 kb) in geringem Volumen eluiert. Die abschließende Elution der DNA von den Säulchen wurde statt mit 75-150 µl Elutionspuffer mit 50 µl dest. H_2O durchgeführt. Die gereinigten DNA Proben wurden bis zur weiteren Verwendung bei -20°C gelagert.

2.8.5 AGAROSEGEL-ELEKTROPHORESE

Sowohl die analytische, als auch die präparative Auftrennung von DNA Proben erfolgte in Agarosegelen. Dazu wurden je nach zu erwartender Fragmentlänge 0,5 – 3 %ige (w/v) Agarosegele mit TAE-Puffer (Tabelle 25) hergestellt. Die Proben wurden vor dem Auftragen mit 6x DNA Auftragspuffer gemischt (Tabelle 25). Als Längenstandard wurden zusätzlich 5 µl DNA Leiter aufgetragen. Die Auftrennung erfolgte bei 80 mA und 120 V. Visualisiert wurden die Banden in den Gelen anschließend in einem Wasserbad mit 0,1 % Ethidiumbromid. Die Banden im Gel wurden mit einem UV-Transilluminator (AlphaImager MINI, Biozym) sichtbar gemacht und dokumentiert.

Tabelle 25: TAE Puffer und 6x DNA Auftragspuffer

50 x TAE-Puffer (pH 8,8)		**6 x DNA Auftragspuffer**	
Tris	2 M	Glycerin	50 % (v/v)
Essigsäure	1 M	EDTA	0,5 M
EDTA	0,1 M	Bromphenolblau	0,25 % (w/v)
Ad ddH_2O	1000 ml	Xylencyanol	0,25 % (w/v)

2.8.6 PCR

In dieser Arbeit wurde zur Amplifikation von DNA-Fragmenten entweder eine Phusion oder eine DreamTaq DNA Polymerase nach Herstellerangaben verwendet, dabei wurde immer 1 µl Template DNA (10 - 150 ng) eingesetzt. Der PCR-Prozess besteht aus einer Anzahl von 25 bis 30 Zyklen, die in einem Thermocycler (PTC-100™ Programmable Thermal Controller, MJ Research Inc., Waltham, USA) durchgeführt wurden.

Amplifikationsprogramm Phusion DNA Polymerase

*Tabelle 26: PCR Protokoll Phusion Polymerase *Tabelle 22*

Schritt	**Temperatur**	**Zeit**	**Zyklenzahl**
Initiale Denaturierung	98 °C	30 s	1
Denaturierung	98 °C	20 s	**25-35**
Primerbindung	T_M*	20 s	
Elongation	72 °C	30 s/kbp	
Finale Elongation	72 °C	10 min	1

Amplifikationsprogramm DreamTaq DNA Polymerase

*Tabelle 27: PCR Protokoll DreamTaq Polymerase. *Tabelle 22*

Schritt	**Temperatur**	**Zeit**	**Zyklenzahl**
Initiale Denaturierung	95 °C	3 min	1
Denaturierung	95 °C	30 s	**25-35**
Primerbindung	T_M* - 5 °C	30 s	
Elongation	72 °C	1 min/kbp	
Finale Elongation	72 °C	10 min	1

2.8.7 PCR BASIERTE GENOTYPISIERUNGSMETHODEN FÜR CLOSTRIDIUM THERMOCELLUM STÄMME

Im Rahmen dieser Arbeit wurden für *C. thermocellum* Isolate verschiedene DNA Banden basierte Typisierungstechniken wie z.B. PCR Ribotypisierung, RAPD (Random amplified polymorphic DNA) und VNTR (Variable Number Tandem Repeat) adaptiert. Diese sind für verwandte Bakterien wie *C. difficle* bereits erprobt (Benson 1999, Bidet et al. 1999, O'Neill et al. 1996, Zaiß et al. 2009). Während dieser Arbeit wurde zudem ein neuer Genotypisierungsansatz entwickelt, welcher auf der Amplifikation von Sequenzen zwischen Kopien eines kürzlich entdeckten Transposons (mobiles Element) basiert. Dieses kann sich zufällig an verschiedenen Stellen im *C. thermocellum* Genom integrieren (Zverlov et al. 2008). Die Methode wurde "randomly inserted mobile element" kurz RIME genannt.

2.8.7.1 PCR-Ribotypisierung

„PCR Ribotypisierung" nutzt spezifische Primer, welche komplementär zu Positionen innerhalb des rRNA operons sind, um die „Spacer"-Region zwischen dem 16S und 23S rRNA Gen zu amplifizieren. Für *C. difficle* wurde gezeigt, dass dieses Bakterium multiple Kopien des rRNA Operons besitzt, welche sich nicht nur in der Kopien Anzahl, sondern auch in der Größe des Spacers zwischen den Genen unterscheiden. Dies resultiert in hoch variablen Amplifikationsmustern zwischen den Stämmen (Brazier 2001). Basierend auf den verfügbaren *C. thermocellum* Genomsequenzen der Stämme ATCC 27405 und DSM 1313 wurden spezifische Primer zur Amplifikation des Spacers entwickelt (vgl Tabelle 22, Abbildung 15). Die PCR wurde mit der DreamTaq DNA Polymerase durchgeführt und die Banden in einem 2 %igen Agarosegel für zwei Stunden bei 80 mA aufgetrennt.

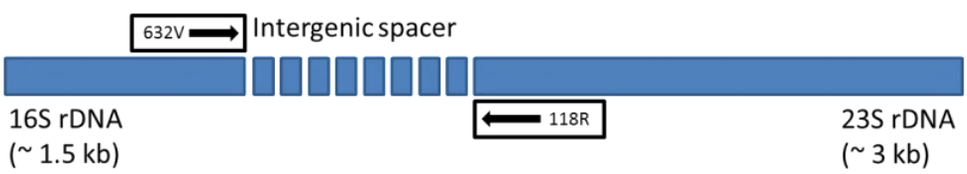


Abbildung 15: Schematische Darstellung des C. thermocellum (DSM 1313) rRNA Operon mit Bindungsstellen für die Primer zur Amplifikation der Spacerregion

2.8.7.2 VNTR (Variable number of tandem repeats)

Diese Methode nutzt das natürliche Auftreten von repetitiven Sequenzen in bakteriellen Genomen, die sich ebenfalls sowohl in Länge als auch in der Anzahl zwischen den verschiedenen Stämmen unterscheiden. Die Genomsequenz von *C. thermocellum* DSM 1313 wurde mit Hilfe des „Tandem Repeats Finder" Version 4.00 (Benson 1999, Zaiß et al. 2009) nach Regionen mit repetitiven Sequenzen durchsucht, welche den Selektionskriterien (Sequenzlänge 15-40 bp; Kopienzahl > 5; Konsensus Sequenz Ähnlichkeit < 90 %) entsprachen. Diese Kriterien wurden gewählt, damit die Sequenzvariationen der repetitiven Sequenzen hoch genug sind, um zwischen

einzelnen Stämmen zu diskriminieren. Das Gen Clo1313_0090 (Genomposition bp 101772 – 104951), welches für ein Protein mit einer S-layer Domäne codiert, lieferte den besten Treffer (Sequenzlänge: 36, Kopienzahl: 5.9, Konsensus: 36 %). Zur Überprüfung der Anwendbarkeit dieser Methode für die *C. thermocellum* Typisierung wurden PCR Primer komplementär zur flankierenden Region der repetitiven Sequenz erstellt (S-layer F; S-layer R; Tabelle 22). Die Amplifikation wurde wie in Kapitel 2.8.6 beschrieben mit der DreamTaq Polymerase bei einer Primerbindungstemperatur von 56 °C durchgeführt. Die PCR Produkte wurden in einem 3 %igen Agarosegel für 2,5 Stunden bei 80 mA aufgetrennt.

2.8.7.3 RAPD (Random amplified polymorphic DNA)

Ein M13 Standard Sequenzierungsprimer wurde für den RAPD Ansatz genutzt (Tabelle 22). Die Amplifikationsreaktionen wurden erneut mit der DreamTaq DNA Polymerase bei einer Primerbindungstemperatur von 35 °C und einer Elongationszeit von 2 Minuten durchgeführt. Die PCR Produkte wurden in einem 1 %igen Agarosegel aufgetrennt.

2.8.7.4 RIME (Random inserted mobile element)

Ein neuer Ansatz zur Genotypisierung, wurde im Rahmen dieser Arbeit entwickelt. Dieser basiert auf der Amplifikation von Sequenzen zwischen Kopien eines kürzlich entdeckten Transposons (mobiles Element), das sich zufällig an verschiedenen Stellen im *C. thermocellum* Genom integrieren kann (Zverlov et al. 2008). Die folgenden Oligonukleotidprimer, welche komplementär zu den Sequenzen des mobilen Elements IS1447 sind, wurden für die Experimente entwickelt: IS1447F, IS1447R, IS1447Fi, IS1447Ri (Tabelle 22). IS1447F und IS1447R wurde genutzt, um das Transposon in den verschiedenen *C. thermocellum* Genomen nachzuweisen. Die komplementären Primer wurden genutzt, um die Sequenzen zwischen den Transposons an verschiedenen Stellen zu amplifizieren (Abbildung 16). Die Amplifikationsreaktionen wurden wieder mit der DreamTaq DNA Polymerase bei einer Primerbindungstemperatur von 50 °C für die Amplifikation der mobilen Elemente bzw. von 40 °C für die komplementären Primer durchgeführt. Die PCR-Produkte wurden in einem 1,5 %igen Agarosegel aufgetrennt.

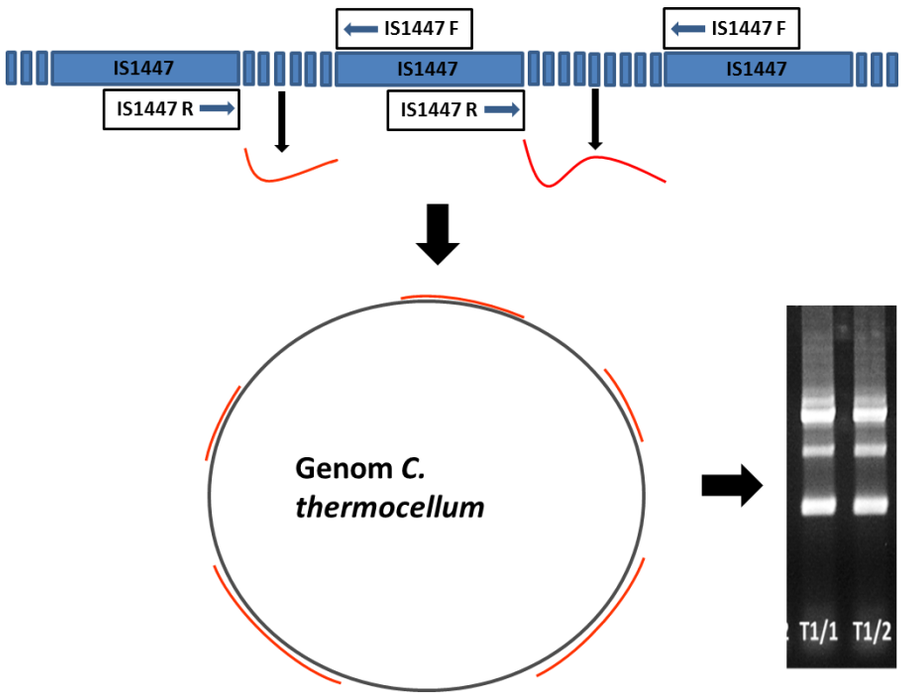


Abbildung 16: Schematische Darstellung der RIME (Random inserted mobile element) Methode zur Genotypisierung

Für jede Genotypisierungsmethode wurde die Anzahl der Banden, die für jeden Stamm erhalten wurden, und die Anzahl der gemeinsamen Banden von jedem Stamm mit den anderen Stämmen ermittelt. Daraus wurde die prozentuale Ähnlichkeit der Bandenmuster verschiedener Stämme mit Hilfe des Koinzidenz-Index ermittelt (Dice 1945).

$$\text{Koinzidenz-Index} = \frac{2 \times gemeinsame\ Banden\ von\ Isolat\ A\ und\ Isolat\ B}{(Anzahl\ der\ Banden\ von\ Isolat\ A + Anzahl\ der\ Banden\ von\ Isolat\ B)} \times 100$$

Formel 2: Berechnung Koinzidenz-Index

2.8.8 Aufreinigung und Modifikation von PCR-Produkten

Der QIAquick spin PCR Purification Kit wurde für die Reinigung der PCR-Produkte von Nukleotiden, Primern, Polymerasen und des PCR-Puffers durch Bindung der DNA Fragmente an die Silica-Gel-Membran der Säulchen verwendet. Lösungen und Arbeitsprotokoll sind im Kit enthalten. Für die Extraktion der Banden aus einem Agarosegel wurde der QIAquick Gel Extraction Kit nach Herstellerangaben verwendet. Die DNA wurde statt mit dem im Kit enthaltenen Elutionspuffern mit 20-50 µl steriles H_2O eluiert.

Für die nachträgliche Poly-Adenylierung von „blunt-end“ PCR Produkten zur anschließenden Klonierung in den pCR®2.1-TOPO® Plasmidvektor wurde folgender Reaktionsansatz für 20 Minuten bei 72 °C inkubiert:

Tabelle 28: Amplifikationsprogramm für die Poly-Adenylierung

	Konzentration	Volumen (µl)
Gereinigtes PCR Produkt	0,15 bis 1,5 pmol	variabel
dATP (10 mM)	0,2 mM	1
PCR Buffer (10x)	1x (1,5 mM $MgCl_2$)	5
DreamTaq DNA Polymerase (5 U/µl)	1U	0,2
ddH_2O	-	add 50 µl

2.8.9 KLONIERUNG

Die Klonierung in den pCR®2.1-TOPO® Plasmidvektor (TOPO TA Cloning Kit, Thermo) diente in dieser Arbeit in erster Linie dazu, die PCR-Amplifikate des 16S rRNA Gens einer Mischkultur zu separieren und einzelne Spezies zu identifizieren. Der Plasmidvektor pCR®2.1-TOPO® liegt zunächst linearisiert vor. Das TA Cloning® nutzt die Fähigkeit der Taq DNA Polymerase, einzelne Desoxyadenosine an die 3´-Enden jedes beliebigen PCR-Produkts anzuhängen. Der Plasmidvektor hat dazu komplementäre 3´-überhängende Desoxythymidin-Reste, wodurch das PCR-Produkt effizient in den Vektor ligieren kann. Das Enzym Topoisomerase I aus dem Vaccinia Virus, ist in der Lage, an spezifische Erkennungsstellen in der doppelsträngigen DNA zu binden, und spaltet die Phosphodiesterbindung nach den Basen 5´-CCCTT im jeweiligen Strang. Die Energie, die bei dieser Spaltung freigesetzt wird, wird verwendet, um eine kovalente Bindung zwischen dem 3`-Phosphat des gespalteten DNA-Stranges und einem Tyrosinrest der Topoisomerase I zu erzeugen. Alle nötigen Lösungen und das Arbeitsprotokoll sind im TOPO TA Cloning® Kit enthalten.

Zur Klonierung von PCR Fragmenten ohne überhängende Enden wurde in der Regel das CloneJET PCR Cloning Kit genutzt. Der Vektor pJET1.2/blunt kodiert ein Gen für ein letales Restriktionsenzym das durch Erfolgreiche Ligation eines DNA Fragments in die Klonierungsregion unterbrochen wird. Die Klonierung wurde nach Herstellerangaben durchgeführt.

2.8.10 KLONIERUNG MITTELS GIBSON-ASSEMBLIERUNG

Die Gibson Assemblierung liefert im Vergleich zu den klassischen Klonierungsstrategien (Restriktionsverdau, Ligation) gesteigerte Klonierungsausbeuten bei geringerem Arbeitsaufwand. Im Rahmen dieser Arbeit wurde das „Gibson Assembly-Verfahren“ (isothermale Assemblierung) ausschließlich zur Klonierung von Gensequenzen in Expressionsvektoren (beispielsweise in das pET24(+)-System) zur heterologen Expression von Enzymen genutzt. Die isothermale Assemblierungsreaktion benötigt drei verschiedene Enzyme, die zugleich bei 50 °C Inkubation aktiv sind. Der Assemblierungsmix (Gibson Assembly Master Mix, New England Biolabs) enthält eine T5 5‘-Exonuklease zur Erstellung einzelsträngiger Fragmentenden, eine Phusion DNA-

Polymerase, um die Lücken zwischen den angelagerten Fragmenten wieder aufzufüllen und die Taq DNA-Ligase, um die Enden zu assemblierten Fragmenten zu verbinden (Abbildung 17).

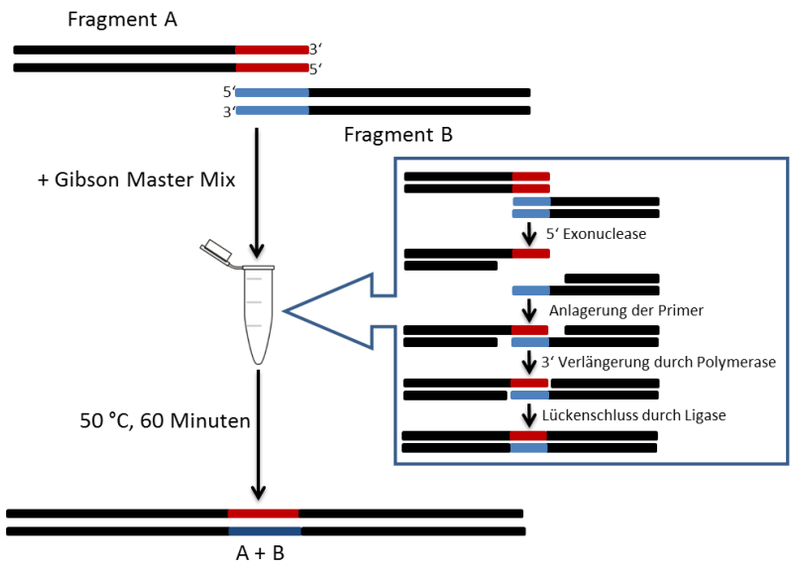


Abbildung 17: Überblick über die Gibson Assemblierungsmethode (New England Biolabs)

Dieser Mix ist kommerziell von New England Biolabs erhältlich und wurde nach Herstellerangaben mit dem Vektor und dem zu ligierenden PCR Fragment gemischt. Da für die Gibson-Assemblierung mit dem Vektor überlappende, endständige Sequenzbereiche an den zu klonierenden DNA Fragmenten notwendig sind, wurden diese, wie in der Anleitung empfohlen, mittels PCR-Primern an den Enden mit eingefügt. Insgesamt sollen 0,02 bis 0,5 pmol der DNA Fragmente pro Reaktionsansatz zur Assemblierung von einem bis zwei Fragmenten vorhanden sein. Die Konzentration der Edukte wurde dabei möglichst genau mittels Gel-Elektrophoretischer Quantifizierung abgeschätzt. Die Gibson-Reaktion wurde darauf wie folgt zusammengeführt und anschließend eine Stunde bei 50 °C inkubiert:

Tabelle 29: Ansatz für die Gibson Assemblierung

		Eingesetztes Volumen
Fragment A (pET24(+) Vektor	0,014 bis 0,028 pmol (Konzentration 50-100 ng/µl)	1 bis 2 µl
Fragment B (unaufger. PCR-Insert)	0,028 bis 0,084 pmol	variabel
Gibson Assembly Master Mix	2x konzentriert	5 µl
ddH_2O		Ad 10 µl Endvolumen

Der Mix wurde danach direkt transformiert (2.8.11).

2.8.11 *TRANSFORMATION CHEMISCH KOMPETENTER ZELLEN*

Zur Transformation wurden ca. 50 µl des gewünschten kompetenten *E. coli* Stammes mit 1-5 µl des zu transformierenden Konstrukts gemischt und 30 Minuten auf Eis inkubiert. Anschließend folgte ein Hitzeschock für 30 Sekunden bei 42 °C und danach die sofortige Aufnahme in 300 µl SOC Medium. Nach einer Stunde schütteln (500 – 700 rpm) bei 37 °C wurden die Zellen auf einem den Anforderungen entsprechenden Selektionsmedium ausplattiert und über Nacht bei 37 °C inkubiert.

2.8.12 *PLASMID-PRÄPARATION*

Der QIAprep Spin MiniPrep Kit (QUIAGEN, Hilden, Deutschland) eignet sich zur Isolierung von bis zu 20 µg Plasmid DNA aus Bakterienkulturen. Die Plasmide wurden, nachdem sie durch alkalische Lyse freigesetzt wurden, an eine Silica-Gel-Membran gebunden. Die so isolierten Plasmide weisen einen besonders hohen Reinheitsgrad und eine hohe Ausbeute auf. Alle benötigten Lösungen und das Arbeitsprotokoll sind im Kit enthalten. Die Plasmide wurden aus transformierten *E. coli* Kolonien auf Festmedium isoliert. Die abschließende Elution der DNA von den Säulchen wurde statt mit 50-100 µl Elutionspuffer mit 50 µl dest. H_2O durchgeführt.

2.8.13 *ENZYMATISCHER VERDAU VON DNA*

Zur Überprüfung des Klonierungserfolges wurden die Plasmide nach der Plasmid-Präparation mit verschiedenen Restriktionsenzymen (vgl. 2.1) auf beiden Seiten der Insert-Integrationsstelle geschnitten (Tabelle 30).

Tabelle 30: Ansatz Restriktionsverdau

10 x Puffer	2 µl
Enzym 1	1 µl
Enzym 2	1 µl
Plasmid-DNA	5 µl
ddH_2O	11 µl

Der Ansatz wurde für eine Stunde bei 37 °C im Heizblock inkubiert. Zur Visualisierung der Restriktionsfragmente wurden 10 µl des Ansatzes mit 6x DNA Auftragspuffer gemischt und in einem Agarosegel (1 %, w/v) bei 80 mA und 120 V aufgetrennt.

2.8.14 SEQUENZIERUNG

Gensequenzierungen wurden von der Firma Eurofins MWG GmbH (Ebersberg) von dem folgenden Ansatz durchgeführt:

Tabelle 31: Ansatz Sequenzierung

gereinigtes PCR-Produkt	10-20 ng/µl
bzw. gereinigte Plasmid DNA	50-100 ng/µl
Primer (10 pmol/µl)	2 µl
ddH_2O	ad 15 µl

Die Genomsequenzierungen wurden am Zentrum für Biotechnologie in Bielefeld durchgeführt und für die Daten für diese Arbeit freundlicherweise zur Verfügung gestellt. Dafür wurden 1-2 µg hochreine und hochmolekulare DNA isoliert. Die Genomrohdaten wurden davon in der Regel mit einem 8-K-mate-pair Sequenzierungslauf (2 x 300 bp) auf dem Illumina MiSeq System generiert. Die Rohdaten wurden mit der GS De Novo Assembler Software (Version 2.8) prozessiert und verbleibende Lücken im Genom gegebenenfalls *in silico* (mit Hilfe von Referenzsequenzen) oder mit der Sequenzierung von PCR Amplifikaten, die mit Hilfe flankierender Primer erstellt wurden, geschlossen.

2.9 PROTEINBIOCHEMISCHE METHODEN

2.9.1 FÄLLUNG UND AUFKONZENTRIERUNG VON PROTEINEN

Bei der Fällung mit Ammoniumsulfat ($(NH_4)_2SO_4$) entziehen die Ionen des Salzes den Proteinen ihre Hydrathülle, was zur Proteinaggregation und schließlich zum Ausfällen der Proteine führt (Lottspeich and Zorbas 1998). Eine frische Kultur (50 – 500 ml) wurde vor der Fällung auf Eis gekühlt und zweimal abzentrifugiert (5000 ×g, 30 min, 4 °C). Vorversuche haben gezeigt, dass die höchste Ausbeute bei gleichbleibender Aktivität der Proteine bei Endkonzentrationen von 60 % gesättigter Ammoniumsulfatlösung (v/v) erreicht wird. Daher wurde in den Kulturüberstand langsam 60 % (v/v) gesättigte Ammoniumsulfat Lösung getropft und über Nacht bei 4 °C gerührt. Die präzipitierten Proteine wurden durch Zentrifugation (5000 ×g, 30 min, 4 °C) geerntet und in 0,1 M MOPS Puffer (pH 6.5) gelöst. Gegebenenfalls kann die Proteinlösung noch mit Vivaspin Säulchen (Sartorius, Göttingen) aufkonzentriert werden.

2.9.2 BESTIMMUNG DER GESAMTPROTEINKONZENTRATION MITTELS BRADFORD

Der Bradford-Assay (Coomassie® Plus Protein Assay Reagent, Bio-Rad, München) dient zur photometrischen Bestimmung der Gesamtproteinkonzentration in einer Probe. Der Test beruht auf der Bindung des Farbstoffes Coomassie Brilliant Blue G-250 an die Proteine in saurer Lösung. Die Farbreaktion ist vor allem abhängig vom Gehalt an aromatischen und an basischen Aminosäuren (z.B. Arginin). Die Gesamtproteinkonzentration ist in einem bestimmen Messbereich mit der Intensität der Blaufärbung direkt proportional. Anhand der aus der Kalibriergerade (vgl. X) abgeleiteten Formel lassen sich so Proteinkonzentrationen von wenigen µg/ml in Flüssigkeiten bestimmen. Dazu wurden 5 - 50 µl der Proteinlösung mit 1 ml Färbelösung gemischt und nach fünf Minuten die Absorption bei 595 nm gegen eine Nullprobe mit gleicher Menge Wasser (statt Proteinlösung) gemessen. Mit Hilfe der folgenden Formel kann daraus die Proteinkonzentration errechnet werden, V ist dabei das Volume der eingesetzten Probe in µl:

$$Gesamtproteinkonz.(mg/ml) = (\frac{OD_{595} - 0{,}0319}{0{,}047}) \div V$$

Formel 3: Gesamtproteinkonzentration nach Bradford

Die Proteinlösungen wurden so verdünnt, dass die Messwerte im linearen Bereich der Kalibriergerade (Absorption < 1,0) lagen.

2.9.3 SDS PAGE

Die SDS-PAGE ist eine eindimensionale Poly-Acrylamid-Gel-Elektrophorese (PAGE), bei der das anionische Detergens Natriumlaurylsulfat (engl. sodium dodecyl sulfate, SDS) eingesetzt wird. Die Ladungen der Proteine werden durch das stark negativ geladene SDS maskiert, damit sich in der Regel alle Proteine im gleichen Ladungszustand befinden und nur bestimmt durch ihre Molmasse aufgetrennt werden. Je nach Konzentration des Acrylamids lassen sich unterschiedlich große Proteine trennen. In dieser Arbeit wurden 10 %ige Gele verwendet, welche Proteine von ca. 36-200 kDa auftrennen. Im Folgenden sind die Puffer und Lösungen zur Durchführung einer SDS-PAGE aufgelistet:

Tabelle 32: Puffer und Ansätze für SDS Gele

Trenngelpuffer (pH 8,8)		**Sammelgelpuffer (pH 6,8)**	
Tris	36,4 g	Tris	12,2 g
SDS 10%	8 ml	SDS 10%	8 ml
ddH_2O	ad 200 ml	ddH_2O	ad 200 ml
10 % Trenngel		**4 % Sammelgel**	
Trenngelpuffer	5 ml	Sammelgelpuffer	1,25 ml
Acrylamid (30% w/v)	6,7 ml	Acrylamid (30% w/v)	0,5 ml
ddH_2O	8,3 ml	ddH_2O	3,25 ml
APS (10% w/v)	100 µl	APS (10% w/v)	30 µl
TEMED	15 µl	TEMED	15 µl

30 µl der Probe werden mit 10 µl 4 x Auftragspuffer gemischt und bei 100 °C für fünf Minuten denaturiert. Die Auftrennung erfolgte für 1,5 Stunden bei 20 mA pro Gel, als Größenstandard wurden 5 µl PageRuler Protein Ladder (Tabelle 3) aufgetragen.

10 x SDS Elektrophorese-Laufpuffer

Tabelle 33: 10 x SDS Elektrophorese-Laufpuffer

Tris	30,3 g
Glycin	144,1 g
SDS	10,0 g
ddH_2O	ad 1000 ml

Die Proteine wurden mit dem blauen Farbstoff Coomassie Brillant Blue für 30 Minuten bei Raumtemperatur gefärbt und anschließend wurde das Gel mit Entfärbelösung behandelt, bis Banden zu erkennen waren. Anschließend wurden sie noch in ddH_2O von Entfärber Resten befreit.

Tabelle 34: Färbe-/Entfärbe Lösung für SDS PAGE

Färbelösung		**Entfärbelösung**	
Coomassie Brillant Blue (G 250)	1,5 g	Methanol	125 ml
Methanol	455 ml	Essigsäure (100%)	150 ml
Essigsäure (100%)	80 ml	ddH_2O	ad 2500 ml
ddH_2O	Ad 1000 ml		

2.9.4 WESTERN BLOT

Die Proteine wurden mittels SDS-PAGE getrennt und die Gele nach Lauf kurz mit Wasser gespült. Zum Blotten wurde ein Semi-Dry Blotter verwendet (LKB). Die PVDF-Membran wurde passgenau ausgeschnitten (Standardgelgröße etwa 7 x 8 cm) und vor Verwendung kurze Zeit in Methanol getaucht. Aus Whatman-Paper (Typ1) wurden ebenfalls 7 x 8 cm große Stücke ausgeschnitten und pro Blot 10 Stück dieser Formen in Western-Transfer Pufferlösung gebadet (5,8 g/l Tris, 2,9 g/l Glycin, 0,37 g/L SDS, Methanol 200 ml/l). Nach Anfeuchten der Graphitplatten des Blotters mit Wasser wurden 5 Whatman Paper aufeinandergelegt, gefolgt von der PVDF-Membran, darauf schließlich das SDS-Gel und weitere 5 Whatman Papers. Nach Abdecken des Blotters mit der zweiten Graphitplatte, wurde der Blot für 1 bis 1,5 Stunden bei 70 mA/Blot laufen gelassen. Der Erfolg des Transfers wird durch Übertragung der Markerbanden sichtbar.

Nach dem Blotting wurde die PVDF-Membran in 1x PBS-T (PBS Puffer mit 0,05% Tween) für 30 Minuten auf einem Wippschüttler bei RT gewaschen. Es folgte die Zugabe des primären Antikörper (Der Titer ist abhängig von vielen Faktoren, für den anti-Cohesin-Antikörper reicht eine 1:5000 Verdünnung), der Antikörper wurde in PBS-T + 5 % Milchpulver verdünnt (z.B. 5 µL AB in 5 mL PBS-T mit Milchpulver). Die Inkubation im Hybridisierungsofen fand bei 37 °C für eine Stunde statt. Nach Inkubation wurde der Blot dreimalig mit PBS-T bei RT gewaschen (etwa 15 Minuten), dann folgte der zweite Antikörper (Verdünnung 1:10000), welcher ein Konjugat aus IgG und AP (Alkaline Phosphatase) war. Der zweite Antikörper wurde ebenfalls in PBS-T mit 5 % Milchpulver für 30 Minuten inkubiert (RT). Dann wurde schließlich dreimalig für jeweils ca. 15 min mit PBS ohne Tween gewaschen. Dann folgte die Detektion des Blots über Inkubation eines Substrats (NBT BCIP für Alkaline Phosphatase) bei RT und lichtgeschützt, bis Banden sichtbar werden. Die Reaktion wurde durch Waschen der Membran in Wasser abgestoppt.

2.9.5 CELLULOSOMPRÄPARATION

Der für die Präparation gewünschte *C. thermocellum* Stamm wurde als Vorkultur in 50 ml GS2-Medium (mit Filterpapier) angezogen, anschließend 1:20 in 800 ml vorgewärmtes GS2 Medium überimpft und für 3-4 Tage bis zum vollständigen Abbau des Filterpapiers inkubiert. Die Kulturen

wurden nach der Inkubation im Kühlschrank abgekühlt, ehe die Zellen und nicht lösliche Medienbestandteile abzentrifugiert wurden. Von dem Kulturüberstand wurden wie unter 2.9.1 beschrieben, die Proteine mit Ammoniumsulfat gefällt und anschließend abzentrifugiert (20 min bei 8000 rpm und 4°C). Das Proteinpellet kann in gesättigter Ammoniumsulfat Lösung resuspendiert und vorübergehend im Kühlschrank gelagert werden. Die gefällten Proteine wurden anschließend bei maximaler Geschwindigkeit in einer Tischzentrifuge gekühlt pelletiert (4 °C, 13.000 rpm, 20 Minuten). Der Überstand wurde sorgfältig mit einer Pipette abgenommen und restliche Flüssigkeit möglichst genau entfernt. Das Pellet wurde in Gelfiltrationspuffer (Tabelle 35) resuspendiert und unlösliche Bestandteile ein weiteres Mal abzentrifugiert. Der Überstand wurde falls nötig mittels Vivaspin Konzentrator-Säulchen (100 kDa molecular weight cutoff) bei maximal 12.400 rpm (14.500 g), bis auf etwa 500 µL aufkonzentriert. Über einen Probenloop wird die Probe nach Herstellerangaben mit dem ÄKTApurifier System (GE Healthcare, Freiburg) auf eine Superdex 200 Gelfiltrationssäule (GE Healthcare, Freiburg) geladen und mit 1,5 Säulenvolumen Gelfiltrationspuffer eluiert. Der Hauptpeak wurde bei den 1 ml Fraktionen A6-A7 erwartet (siehe Abbildung 18). Die beiden Fraktionen wurden zusammengeführt und mit 1:100 Volumen an 2 %iger Natriumazid (NaN_3)-Lösung versetzt.

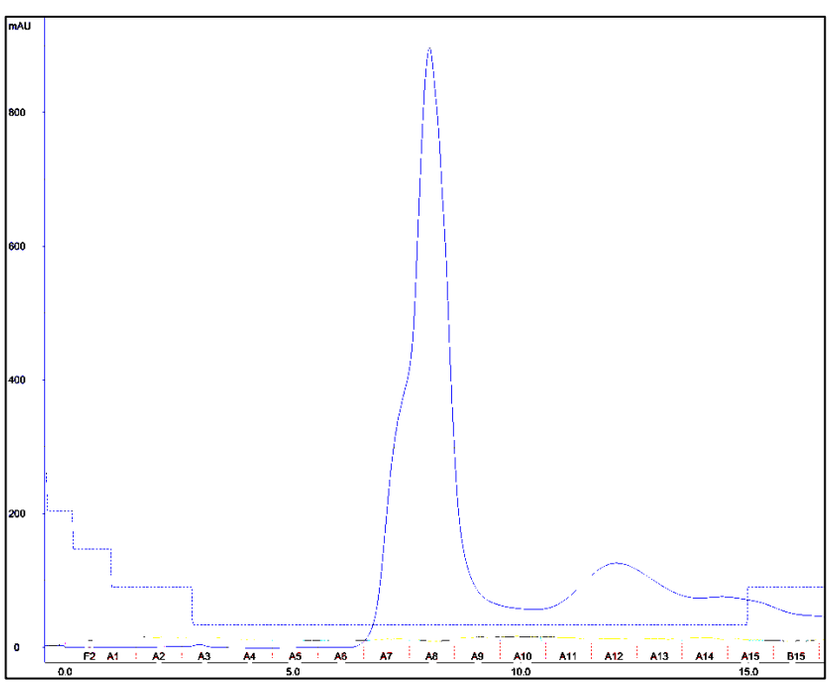


Abbildung 18: Typischer Verlauf der Gelfiltrationsreinigung des nativen Cellulosoms auf einer Superdex200 Säule.

Tabelle 35: Gelfiltrationspuffer

MOPS	50 mM
NaCl	0,5 M
$CaCl_2$	10 mM

2.9.6 AFFINITÄTSAUFREINIGUNG REKOMBINANTER PROTEINE

Im Folgenden wird die Vorgehensweise zur Expression von rekombinanten Proteinen in *E. coli,* Zellernte und -Aufschluss und die native His_6-Affinitätsaufreinigung des Zielproteins beschrieben.

2.9.6.1 Zellanzucht, -ernte und -Aufschluss

Zunächst wurde der mit dem entsprechenden Expressionsvektor erfolgreich transformierte *E. coli* BL21 Star Klon in 30 ml LB Medium als Vorkultur angezogen und anschließend 1:50 in 250-1000 ml ZY-5052 Autoinduktionsmedium überimpft. Dieses wurde über Nacht bei 37 °C geschüttelt. Das Zellpellet wurde auf Eis geerntet (10 min, 4500 rpm, 4°C) und mit einer ½ Tablette Complete Mini (Protease-Inhibitor, Roche Life Science) sowie je 20 ml Lysis-Wasch-Puffer resuspendiert. Nach Zugabe von 200 µl 100 mg/ml-Lysozym-Lösung wurden die Zellen 30 Minuten auf Eis inkubiert und anschließend auf Eis sonifiziert (Amplitude: 60%, Intervall: 0,25, 2x 4 min). Nach einer erneuten Zentrifugation (20 min, 18000 rpm, 4°C) konnte der Überstand für die His_6-Affinitätsreinigung verwendet werden.

Tabelle 36: Puffer für Proteinaffinitätsaufreinigung

Lysis-Wasch-Puffer (pH 7,3)		**Elutionspuffer (pH 7,3)**	
MOPS	50 mM	MOPS pH 7,3	50 mM
NaCl	0,1 M	Imidazol	0,5 M
Imidazol	20 mM	NaCl	0,1 M
$CaCl_2$	5 mM	$CaCl_2$	5 mM
Stripping-Puffer (pH 7,3)		**Regenerationspuffer**	
MOPS	50 mM	Ni_2SO_4	100 mM
NaCl	0,1 M		
EDTA	50 mM		

2.9.6.2 Native His_6-Affinitätsreinigung

Die Aufreinigung der rekombinanten Proteine wurde mit einer FPLC oder ÄKTA pure Anlage unter Verwendung des C-terminalen His-Tags, den die Proteine durch die Klonierung der Gene in den pET24-c (+) Vektor erhalten, nach Herstellerangaben durchgeführt. Über die sechs Histidin-Seitenketten kann das rekombinante Protein mittels Affinitätschromatographie von den *E. coli* Proteinen getrennt werden. Der His-Tag bindet reversibel an die Ni^{2+}-Ionen der HisTrap FF Säulen und bildet mit diesen einen Chelatkomplex. Die Elution des rekombinanten Proteins erfolgt mit Imidazol, welches das Protein aus dem Chelatkomplex verdrängt. Die Detektion der Proteine erfolgt durch Messung der Absorption bei 280 nm mit Hilfe eines UV-Detektors. Für die 1 ml Säulen wurde eine konstante Flussrate von 1 ml/min eingesetzt. Zunächst wurde die 1 ml HisTrap FF Säule mit 5 Säulenvolumen (SV) 20 % EtOH und 10 SV H_2O gewaschen und anschließend mit 10 Volumen Lysepuffer equilibriert. Der Proteinüberstand wurde mit einem

Superloop (FPLC) oder einer Probenpumpe (ÄKTA pure) auf die Säule aufgetragen. Der Hauptpeak des Durchflusses wurde für eine spätere SDS-PAGE gesammelt und die Säule so lange gewaschen bis die Absorption wieder das Basisniveau erreichte. Anschließend wurde das rekombinante Protein mit Elutionspuffer eluiert und separat aufgefangen. Die HisTrap FF Säule wurde erneut mit 5 SV Lysepuffer, 5 SV H_2O und 10 SV 20 % EtOH gewaschen und bis zur Regeneration bei 4°C gelagert. Zur Regeneration wurden die verwendeten HisTrap FF Säulen zunächst mit 10 SV H_2O gewaschen. Die Ni^{2+}-Ionen wurden mit 10 SV Stripping-Puffer von der Säule gewaschen. Dabei komplexiert EDTA die Ni^{2+}-Ionen und löst sie so aus der Sepharose-Matrix. Die Säule wurden erneut mit 10 SV H_2O gewaschen und durch 5 SV „Recharging" Lösung wieder mit Ni^{2+}-Ionen beladen. Anschließend wurde die Säule mit 5 SV H_2O und 10 SV 20 % EtOH gewaschen und bei 4°C aufbewahrt.

2.9.6.3 Hitzedenaturierung

Neben dem Zielprotein können auch andere Proteine mit Histidin-Resten oder negativen Ladungen an die Ni^{2+}-Ionen der HisTrap FF Säule binden. Bei Proteinen aus thermophilen Organismen kann zur weiteren Reinigung eine Hitzedenaturierung durchgeführt werden. Diese sind bei höheren Temperaturen noch stabil, während die *E. coli* Proteine denaturieren. Das Eluat wurde in 2 ml Reaktionsgefäße überführt und im Wasserbad bei 45 - 60 °C (abhängig von der optimalen Wachstumstemperatur des Spenderorganismus) für 15 min inkubiert. Anschließend wurde die Proteinlösung bei maximaler Geschwindigkeit für 15 min zentrifugiert und die Überstände vereinigt. Zur Stabilisierung und sicheren Lagerung wurde die Proteinlösung mit 20 % Glycerin und 1 % Natriumazid versetzt und bis zur weiteren Verwendung bei 4 °C oder -20 °C aufbewahrt.

2.9.7 Enzymaktivitätsassay mit DNSA (3,5-Dinitrosalicylsäure)

3,5-Dinitrosalicylsäure (DNSA) ist eine Phenolsäure, die von reduzierenden Zuckern oder anderen reduzierenden Molekülen wie z.B. Aldehyd-Gruppen zu 3-Amino-5-Nitrosalicylsäure reduziert wird, welche sichtbares Licht bei 540 nm stark absorbiert. Bei dieser Reaktion ändert sich die Farbe der Lösung von orangegelb nach rotbraun. Diese Farbreaktion wurde dazu verwendet, um den Abbau von polymeren Zuckern durch Cellulasen und Hemicellulasen mit Hilfe der dabei frei werdenden Aldehyd Gruppen der reduzierenden Zuckerenden zu quantifizieren (Miller 1959). Für die Enzymassays wurden 1 %ige (w/v) Lösungen bzw. Suspensionen der Polysaccharide in ddH_2O hergestellt (vgl. Tabelle 16). Die Enzymaktivität wurde im doppelten oder dreifachen Ansatz bestimmt. In einem Reaktionsansatz mit einem Gesamtvolumen von 500 µl wurden 250 µl der Polysaccharid-Substratlösung, mit 1x Reaktionspuffer und 0,1 bis 50 µg Enzym zusammen pipettiert. Die Wahl des Reaktionspuffers hing dabei vom gewünschten pH Bereich der Reaktion ab (Tabelle 37). Der gesamte Ansatz wurde mit ddH_2O auf 500 µl aufgefüllt, gut durchmischt und für 1 bis 72 h bei der für die Reaktion gewünschten Temperatur inkubiert Die Reaktion wurde

durch Zugabe von 750 µl DNSA-Lösung gestoppt. Unlösliche Substratbestandteile wurden bei 16000 g für 5 min abzentrifugiert, der Überstand in frische Reaktionsgefäße überführt und der Reaktionsansatz bei 100 °C für 5 min aufgekocht. Anschließend wurde der Reaktionsansatz auf Eis abgekühlt und die Absorption im Photometer bei 540 nm gemessen. Für die Berechnung der Enzymaktivität wurde anhand einer Kalibriergerade (siehe Anhang) zunächst die Glucose-Konzentration bestimmt, die der gemessenen Absorption bei 540 nm entspricht. Aus der Glucose-Konzentration kann die Enzymaktivität mit den folgenden Formeln berechnet werden:

$$Aktivität\ (U) = \frac{m_{Glu}}{M_{Glu} \cdot t}$$

$$Aktivität(U/ml) = \frac{m_{Glu} \cdot 1000}{M_{Glu} \cdot V_E \cdot t}$$

$$Spezifische\ Aktivität\ (U/mg) = \frac{m_{Glu} \cdot 1000}{M_{Glu} \cdot V_E \cdot c_E \cdot t}$$

$U = Units\ (\mu mol/min)$ $\quad m_{Glu} = Masse\ Glucose\ aus$ Kalibriergerade (μg)

$M_{Glu} = molare\ Masse\ Glucose\ (\mu g/\mu mol)$ $\quad t = Inkubationszeit\ (min)$

$V_E = eingesetztes\ Enzymvolumen\ (\mu l)$

$c_E = Enzymkonzentration\ (\mu g/\mu l)$

Formel 4: DNSA spezifische Enzymaktivität

Tabelle 37: Reaktionspuffer für Enzym Assays und DNSA Lösung

Acetat-Reaktionspuffer (2x) pH 3,5 – 5,5		**Tris-Reaktionspuffer (2x) 7,5 – 9,5**	
Acetat	0,1 M	Tris	50 mM
$CaCl_2$ x 2 H_2O	20 mM	$CaCl_2$ x 2 H_2O	20 mM
NaCl	0,1 M	NaCl	0,1 M
MOPS-Reaktionspuffer (10x) pH 5,5 – 7,5		**DNSA-Lösung**	
MOPS	1 M	3,5-Dinitrosalicylsäure	10 g
$CaCl_2$ x 2 H_2O	0,1 M	NaOH	10 g
NaCl	0,5 M	Kalium-Natrium-Tartrat	200 g
		Na_2SO_4	0,5 g
		Phenol	2 g
		H_2Odd	Ad 1000 ml

2.9.8 *DÜNNSCHICHTCHROMATOGRAPHIE*

Mit Hilfe der Dünnschichtchromatographie wurden die einzelnen Produkte der enzymatischen Hydrolyse von Polysacchariden qualitativ bestimmt (Zverlov et al. 2003). Dazu wurde zunächst ein entsprechender Reaktionsansatz aus Enzym, Reaktionspuffer und gewünschtem Substrat gemischt (vgl. 2.9.7) und im Wasserbad inkubiert. Um die Kinetik der Produktbildung zu erfassen, wurden in definierten Zeitabständen Proben mit einem Volumen von 50µl gezogen. Um eine fortlaufende Enzymaktivität zu vermeiden, wurden alle Proben unmittelbar nach Ende der Inkubationszeit für 15 Minuten bei 80 °C hitzedenaturiert. Danach folgte eine Zentrifugation für 10 Minuten bei 14.000 rpm, um die gebildeten kurzkettigen Zucker vom unlöslichen polymeren Substrat und den denaturierten Proteinen zu trennen. Die Überstände wurden eingedampft und anschließend zum Entfernen der Ionen bis etwa zur Hälfte des Volumens mit DOWEX-1 (Anionenaustauschharz, Dow Chemical) durchmischt. Jeweils 5 - 20 µl der Probe wurden auf Kieselgel-Platten (Silica Gel 60, Merck Millipore) als stationäre Phase aufgetragen und getrocknet. Als mobile Phase diente ein 80:20 Gemisch aus Acetonitril und Wasser. Als Referenzsubstanz wurde 1 µl eines 0,1 %igen (w/v) Gemisches aus unterschiedlichen Zuckern aufgetragen. Nach einer Auftrennungszeit von 40 bis 60 Minuten wurden die DC Platten mit einer Mischung aus 10 ml DC-Färbelösung und 1 ml 85 %iger (v/v) Phosphorsäure gleichmäßig eingesprüht. Die gefärbte DC-Platte wurde bei 120 °C für 10 min entwickelt.

Tabelle 38: DC Färbelösung

Anilin	1 ml
Diphenylamin	1 g
Aceton	100 ml

Um die Abbaubarkeit definierter Cellodextrine besser untersuchen zu können, wurden zusätzlich Dünnschichtchromatographien mit folgenden Substanzen durchgeführt: Cellotriose, Cellotetraose, Cellopentaose und Cellohexaose. Von den Reaktionsansätzen wurden nach Inkubationsbeginn in definierten Zeitabständen Proben von 5 µl entnommen.

2.10 Bioinformatische Methoden/Software

2.10.1 Clustal Omega, Tree View

Multiple Sequenzalignments von 16S rRNA Gensequenzen wurden in der Regel mit Clustal Omega berechnet (Sievers et al. 2011). Aus den multiplen Alignments wurde mit dem Programm „Tree View“ der resultierende phylogenetische Baum dargestellt (Zhai et al. 2002).

2.10.2 BLAST

Der 16S rRNA Gen-Sequenzvergleich selbst erfolgte durch einen Nukleotid-BLAST mit der Nukleotiddatenbank der NCBI–Homepage (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/). Der BLAST Algorithmus (Basic Local Alignment Search Tool) wird dazu verwendet, experimentell ermittelte DNA- oder Protein-Sequenzen mit bereits in einer Datenbank vorhandenen Sequenzen zu vergleichen. Als Ergebnis liefert das Programm eine Reihe lokaler Alignments, d.h. Gegenüberstellungen von Stücken der gesuchten Sequenz mit ähnlichen Stücken aus der Datenbank. Darüber hinaus gibt BLAST an, wie signifikant die gefundenen Treffer sind. Die Homologie der bearbeiteten Suchsequenz wird anhand von Score und E-Wert definiert. Der Score ist eine quantitative Bewertung der Ähnlichkeit der Suchsequenz mit einer bekannten Sequenz (je höher, desto höher ist auch die Identität der Sequenzen). Der E-Wert gibt an, mit welcher Wahrscheinlichkeit man Ergebnisse mit gleichen Scores in einer Datenbank, in welcher sich zufällig generierte Sequenzen befinden, erzielen konnte (je kleiner, desto besser). Bei einer Übereinstimmung der Sequenzen von 98 -100 % handelt es sich sehr wahrscheinlich um einen Organismus der gleichen Spezies (Tindall et al. 2010).

2.10.3 ARB und MEGA

Die phylogenetische Einordnung der Sequenzdaten wurde neben dem NCBI-Datenbankvergleich mit Hilfe der ARB-Software durchgeführt (Ludwig et al. 2004). Mit den 16S rRNA-Sequenzen wurde ein automatischer Sequenzvergleich (Alignment) mit einem Datensatz nicht redundanter, qualitativ hochwertiger Primärsequenzen aus der SILVA Datenbank durchgeführt (Quast et al. 2013). Dabei wurden basierend auf dem „Living Tree Project“ (LTP) (Munoz et al. 2011) und der „List of Prokaryotic Names“ (LPSN) (Parte 2014) alle 16S rRNA Sequenzen von validen Typstämmen ausgewählt. Phylogenetische Bäume wurden mit Hilfe der ARB- oder MEGA6-Software unter Verwendung des Maximum-Likelihood-Algorithmus aus multiplen Alignments berechnet (Stamatakis 2006, Tamura et al. 2013).

2.10.4 *CLONE MANAGER*

Mit der Software Clone Manager 9 (http://www.scied.com) wurden alle *in silico* Klonierungen und Modifikationen von DNA- und Aminosäuresequenzen durchgeführt.

2.10.5 *EDGAR*

EDGAR ermöglicht eine vereinfachte komparative Analyse verwandter Genome. Die Software gibt einen schnellen Überblick über die evolutionäre Verwandtschaft und zeigt die Anzahl und Funktionalität von sich unterscheidenden Genen zwischen verwandten Genomen. Die Resultate können unter anderem mit Syntenyplots oder Venndiagrammen webbasiert visualisiert werden (http://edgar.cebitec.uni-bielefeld.de) (Blom et al. 2009).

2.10.6 *GEN DB*

Mit Hilfe der Annotierunsplatform GenDB wurden die Gene in den neu sequenzierten Genomen zunächst automatisch und anschließend noch einmal manuell annotiert (Meyer et al. 2003). Informationen aus den sequenzierten Genomen können direkt als Nukleinsäuresequenz oder Aminosäuresequenz ausgelesen werden.

2.10.7 *MGX*

Das MGX Projekt am CeBiTec (Universität Bielefeld) beinhaltet die Entwicklung einer neuen Analyseplattform zur Auswertung der großen Datensätze aus Metagenomanalysen. Eine Vorversion dieses Softwarepakets wurde für diese Arbeit zur Auswertung von Metagenomdaten zur Verfügung gestellt.

2.11 STATISTISCHE AUSWERTUNG

Für statistische Analysen wurde ein ungepaarter T-test verwendet, um verschiedene Datensätze paarweise zu vergleichen. Signifikanzwerte (P-Werte) < 0.05 wurden als statistisch signifikant definiert. Zur Berechnung der Diversität der Isolate wurde der Shannon-Wiener-Index (Hs) verwendet (Spellerberg and Fedor 2003).

$$\mathrm{Hs} = \sum_i \; p_i \times \ln p_i \qquad \text{mit} \qquad p_i = \frac{n_i}{N}$$

$$n_i = Individuen\ einer\ Spezies\ N = Summe\ aller\ Individuen$$

Formel 5: Shannon-Wiener-Index

3 ERGEBNISSE

3.1 ERMITTLUNG DER GENETISCHEN VARIATION EINES FÜR DEN ANAEROBEN ABBAU IN DER BIOGASANLAGE WICHTIGEN BAKTERIUMS

3.1.1 AUSWAHL REPRÄSENTATIVER BIOGASANLAGEN

Als Zielorganismus wurde *Clostridium thermocellum* ausgewählt. Hierbei handelt es sich wie eingangs bereits erwähnt um ein sowohl auf genomischer Ebene als auch auf metabolischer Ebene sehr gut charakterisiertes cellulolytisches, thermophiles, anaerobes Bakterium. *C. thermocellum* in ein wichtiger Bestandteil des hydrolytischen Konsortiums in Biogasanlagen. Das Probenmaterial sollte aus einer Auswahl an verschiedenen mesophilen und thermophilen Praxisanlagen mit konstanter Fütterung stammen. Bei der Wahl des Substrates wurden vorrangig NawaRo-Anlagen aus verschiedenen Regionen in Deutschland berücksichtigt. Die neun letztendlich ausgewählten Anlagen, sowie deren wichtigste Prozess Parameter sind in Tabelle 5 zusammengefasst.

3.1.2 ISOLIERUNG

Zur Überprüfung und Etablierung der Isolierungsmethode wurde zunächst ein Vorversuch mit dem *Clostridium thermocellum* Typstamm (ATCC 27405) durchgeführt. Dazu wurde eine 10^{-3}, 10^{-5} und 10^{-7} Verdünnung des Stammes auf GS2 Agarplatten einmal auf dem Topagar, einmal unter dem Topagar und einmal zusammen mit dem Topagar gemischt ausplattiert. Der Topagar bestand aus GS2 Medium mit 0,5 % PASC. Nach einer Woche Inkubation bei 60 °C wurde die Cellulose im Topagar mit Kongorot angefärbt und die farblosen Lysehöfe ausgezählt und die cfu/ml ermittelt (Tabelle 39).

Tabelle 39: Ermittlung der cfu/ml aus der Anzahl der Lysehöfe nach Kongorotfärbung

Ausstrich	Verdünnung	Lysehöfe	cfu/ml
Unter Topagar	10^{-3}	200	$2*10^{5}$
	10^{-5}	20	$2*10^{6}$
	10^{-7}	6	$6*10^{7}$
Über Topagar	10^{-3}	3	$3*10^{3}$
	10^{-5}	22	$2{,}2*10^{6}$
	10^{-7}	0	0
In Topagar	10^{-5}	150	$1{,}5*10^{7}$
	10^{-7}	13	$1{,}3*10^{8}$

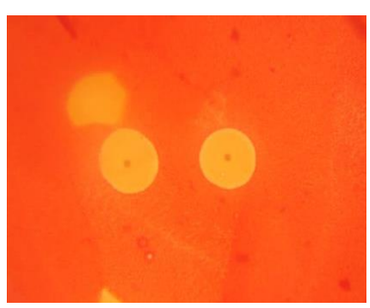


Abbildung 19: Lysehöfe in PASC-Overlay mit C. thermocellum Kolonie im Zentrum, gefärbt mit Kongorot

Die Lysehöfe waren mit PASC deutlich sichtbar (Abbildung 19) und es wurden signifikant mehr Kolonien erhalten, wenn die Bakterien von Agar umgeben waren. Für die anschließende Isolierung von *C. thermocellum* aus den Anreicherungen der Biogasanlagenproben wurde deshalb immer unter Topagar gearbeitet. Nach 2 Wochen Inkubation der GS2 Platten bei 60 °C mit MN301 Cellulose im Topagar wurden schließlich deutliche Lysehöfe auf den Platten aller Biogasanlagen erhalten. Von jeder Biogasanlage wurden mindestens zwei Kolonien gepickt und gereinigt. Die Ergebnisse der 16S rRNA Sequenzanalyse sind in Tabelle 40 zusammengefasst.

Tabelle 40: 16S rRNA Sequenzanalyse der Isolate aus verschiedenen Biogasanlagen

Biogasanlage	Anzahl der Isolate	16S rRNA Sequenzidentität (> 99 %)	Bezeichnung
T1	3	*Clostridium thermocellum*	T1/1; T1/2; T1/3
T3	2	*C. thermocellum*	T3/1; T3/2
J1	2	*C. thermocellum*	J1/1; J1/2
J2	2	*C. thermocellum*	J2/1; J2/2
J3	2	*C. thermocellum*	J3/1; J3/2
H1	2	*C. thermocellum*	H1/1; H1/2
H2	2	*C. thermocellum*	H2/1; H2/2
H3	2	*C. thermocellum*	H3/1; H3/2
A2	2	*C. thermocellum*	A2/1, A2/2

Insgesamt wurden 19 Isolate mit einer 16S rRNA Sequenzidentität > 99 % als *C. thermocellum* identifiziert. Aus jeder Anlage konnten mindestens zwei Stämme isoliert werden, welche für die weiteren Analysen verwendet wurden.

3.1.3 *Etablierung der MALDI-TOF MS Analyse zur Stammidentifizierung*

Um die Stabilität der MALDI-TOF MS Analyse mit dem zur Isolierung verwendeten Medium zu gewährleisten und um *C. thermocellum* Referenzspektren für die Datenbank des RIPAC Labor zu generieren, wurde zunächst eine MALDI-TOF MS Analyse mit einigen *C. thermocellum* Referenzstämmen und Eigenisolaten, sowie eines *C. stercorarium* Stammes durchgeführt. Diese sind in folgender Tabelle aufgeführt.

Tabelle 41: C. thermocellum (C. stercorarium) Referenzstämme und Eigenisolate zur Etablierung der MALDI TOF MS Analyse

Nummer	Bezeichnung	Stamm
001	Medium	
002	Medium	
056	*Clostridium stercorarium*	DSM 8532
079	*C. thermocellum*	DSM 2360
080	*C. thermocellum*	F7
081	*C. thermocellum*	DSM 7072
082	*C. thermocellum*	DSM 4150
093	*C. thermocellum*	NB2
094	*C. thermocellum*	5G
095	*C. thermocellum*	T2
096	*C. thermocellum*	BC
097	*C. thermocellum*	ATCC 27405

Die Anzucht erfolgte auf GS2 Medium mit 0,5 % Cellobiose für 48 h bei 55 °C. Die Zellen wurden nach Protokoll geerntet und in 70 % Ethanol konserviert an das RIPAC Labor verschickt. Die folgende Abbildung zeigt die Auswertung der aufgenommenen MALDI TOF MS Spektren in Form einer SARAMIS MALDI-TOF MS Clusteranalyse.

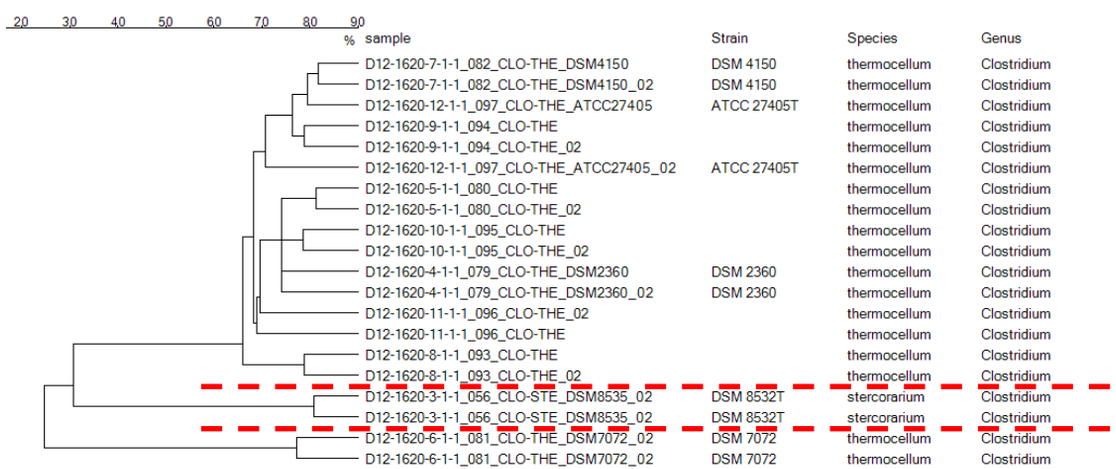


Abbildung 20: SARAMIS MALDI-TOF MS Clusteranalysen der C. thermocellum und C. stercorarium Stämme (vgl. Tabelle 41)

Aufgrund einer Verunreinigung der Ausgangskultur weicht Stamm DSM 7072 *C. thermocellum* von den anderen Stämmen ab. Alle anderen Stämme von *C. thermocellum* zeigen untereinander ähnliche Spektren und bis auf ATCC 27405 ein gutes Subcluster (jeweils zwei Replikate). Der Stamm *Clostridium stercorarium* DSM 8532 zeigt eine eindeutige Trennung von *C. thermocellum* (rot markiert). Von den *C. thermocellum* Spektren konnte ein Subspektrum erstellt werden. Somit eignet sich die Methode sowohl zur Identifizierung als auch zur Feindifferenzierung verschiedener *C. thermocellum* Laborstämme. Zusätzlich wurde noch der Einfluss von Inkubationstemperatur und –zeit der Kulturen auf die Analyse bestimmt. Es konnte festgestellt werden, dass auch längere (bis zu sechs Tage) Inkubationszeiten und höhere Inkubationstemperaturen die MALDI TOF MS Analyse nicht negativ beeinflussen.

3.1.4 GENOTYPISIERUNG VERSCHIEDENER CLOSTRIDIUM THERMOCELLUM ISOLATE

Im Folgenden wird die Etablierung verschiedener Methoden zur Genotypisierung von *C. thermocellum* beschrieben. Verschiedene auf Bandenmustern basierte Methoden, wie PCR Ribotypisierung, RAPD (Random amplified polymorphic DNA) und VNTR (Variable Number Tandem Repeat) wurden für alle *C. thermocellum* Isolate getestet. Zudem wurde eine neue Methode, basierend auf zufällig ins Genom integrierten Mobilen Elementen (**R**andom **I**nserted **M**obile **E**lement), entwickelt.

3.1.4.1 PCR Ribotypisierung

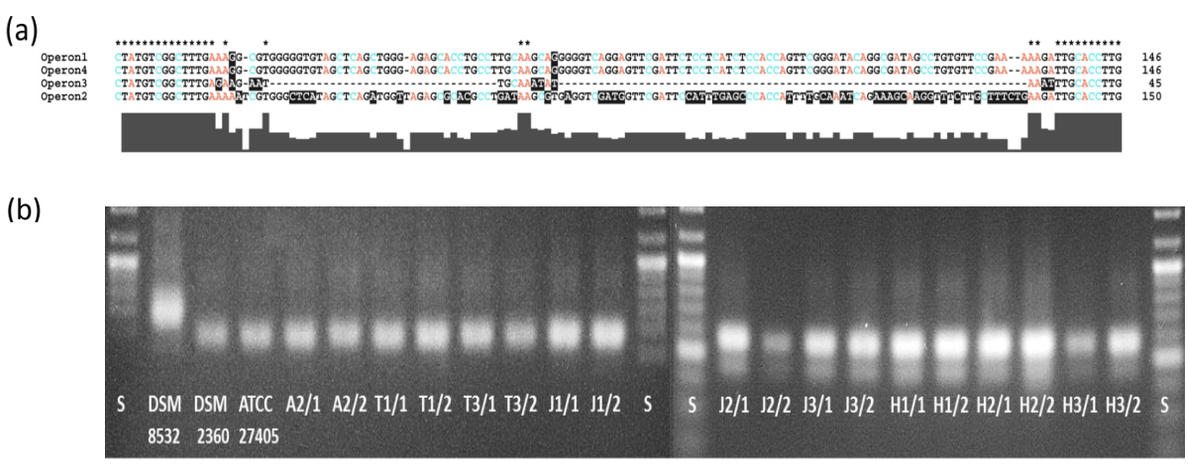


Abbildung 21: (a) Multiples Sequenz Alignment des „Intergenic Spacer" der vier rRNA Operone von C. thermocellum (DSM 1313). Lücken in dem unteren Balken symbolisieren Sequenzunterschiede oder Lücken. Basen mit Sequenzunterschieden sind schwarz hinterlegt. Basen, die in allen Sequenzen identisch sind, sind mit einem Stern markiert. (b) Agarosegel der PCR Produkte der Ribotypisierung aller C. thermocellum Isolate aus den verschiedenen BGA und der DSM Referenzstämme und C. stercorarium DSM 8532 (Koeck et al. 2014d). S = Molekulargewichtstandard

Die bisher sequenzierten Genome von *C. thermocellum* besitzen nur vier Kopien des rRNA Operons, bei welchen sich innerhalb eines Stammes die „Spacer" in Sequenz und Länge unterscheiden (Abbildung 21a). Die Resultate der Ribotypisierung der verschiedenen Isolate

zeigten jedoch identische, überlappende Bandenmuster (Abbildung 21b). Eine genauere Analyse der Operone von bereits sequenzierten Stämmen zeigte, dass Unterschiede nur zwischen den vier Operonen eines Stammes auftauchen, aber die einzelnen Operone sich nicht zwischen den Stämmen unterscheiden. Daher ist im Unterschied zu pathogenen Clostridien bei *C. thermocellum* die PCR-Ribotypisierung nicht geeignet, um verschiedene *C. thermocellum* Isolate auf DNA-Ebene zu unterscheiden. Anhand des höheren Molekulargewichtes konnte aber das Amplifikat von *C. stercorarium* (DSM 8532) von den übrigen Banden differenziert werden.

3.1.4.2 RAPD (Random amplified polymorphic DNA)

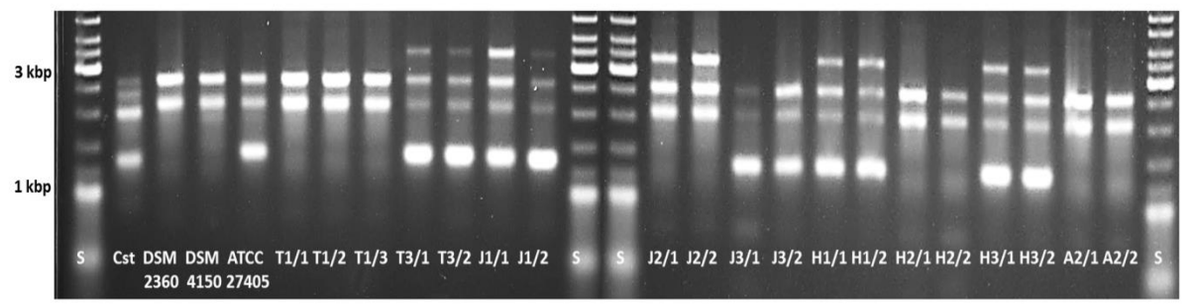


Abbildung 22: RAPD Bandenmuster aller C.thermocellum Isolate und Referenz Stämme und C. stercorarium DSM 8532 (Cst) (Koeck et al. 2014d). S = Molekulargewichtstandard

Mit der RADP Methode wurden von 19 Isolaten nur vier verschiedene Typen von Bandenmustern erhalten (Abbildung 22). Zwei Banden tauchen, außer bei *C. stercorarium* DSM 8532 (Cst), in jedem Muster auf. Die durchschnittliche Ähnlichkeit zwischen zwei Bandenmustern unterschiedlicher Biogasanlagen lag bei 77 %. Zwischen den Bandenmustern der zwei Isolate einer Biogasanlage gab es keine Unterschiede. Aber in einigen Fällen gab es auch keine Unterschiede zwischen den Bandenmustern von Isolaten verschiedener Biogasanlagen (z.B. T3 und J1). Die Ergebnisse, die mit zwei Primern (M13 forward und M13 reverse) erzielt wurden, konnten die Diskriminierung zwischen den Stämmen verschiedener Biogasanlagen nicht erhöhen.

3.1.4.3 VNTR (Variable Number Tandem Repeat)

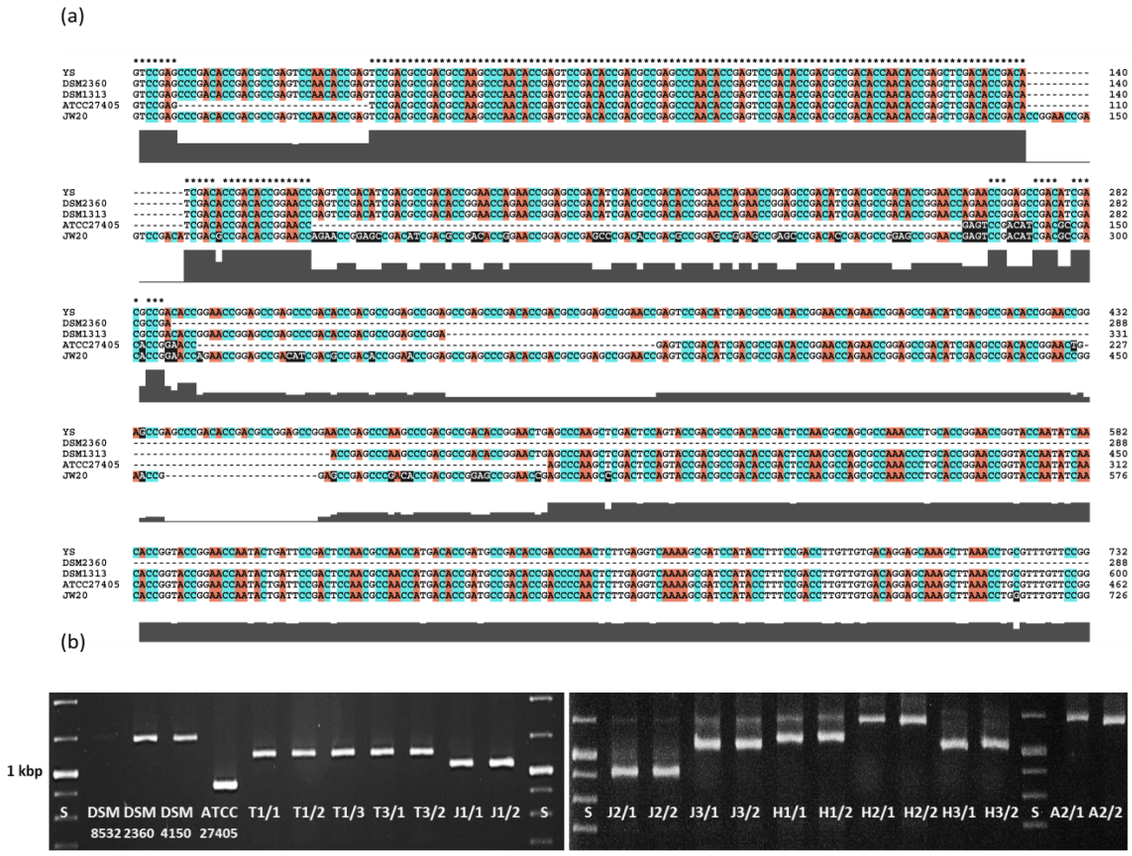


Abbildung 23: (a) Segment des multiplen Alignments des Clo1313_0090 Gens von den C. thermocellum Referenzstämmen (YS, DSM2360, DSM 1313; ATCC 27405, JW20) mit variabler Anzahl an repetitiven Sequenzen. Lücken in dem unteren Balken symbolisieren Sequenzunterschiede oder Lücken. Basen mit Sequenzunterschieden sind schwarz hinterlegt. Basen die in allen Sequenzen identisch sind, sind mit einem Stern markiert. (b) Ergebnisse der PCR Amplifikation dieser variablen Region von allen Isolierten Spezies und C. stercorarium (DSM8532) und C. thermocellum Referenzstämme (DSM2360, DSM4150, ATCC 27405) (Koeck et al. 2014d). S = Molekulargewichtstandard

Ein multiples Alignment des Zielgens Clo1313_0090 (vgl. 2.8.7.2) von verschiedenen *C. thermocellum* Stämmen (ATCC 27405, DSM 1313, JW20, DSM2360) zeigte eine variable Anzahl von repetitiven Sequenzen innerhalb der variablen Region. Aufgrund der unterschiedlichen Kopienzahl wies der Lokus einen ausgeprägten Längenpolymorphismus auf. Die Sequenzen der einzelnen Wiederholungen unterschieden sich nur geringfügig (Abbildung 23a). Mit PCR Primern komplementär zu den flankierenden Regionen des repetitiven Bereichs, konnten reproduzierbare Produkte von allen Isolaten und Referenzsequenzen generiert werden. Von der genomischen DNA der 22 Isolate konnten sechs verschiedene Längen des Gens amplifiziert werden. Aufgrund der hohen Primer-Spezifität wurde von der *C. stercorarium* (DSM 8532) DNA kein Amplifikat erhalten. Alle Isolate einer Biogasanlage weisen dieselbe Amplifikatlänge auf, aber Isolate unterschiedlicher Biogasanlage ergeben meist unterschiedliche Längen. Durch den Längenpolymorphismus des Clo1313_0090 Gens konnten somit die meisten *C. thermocellum* Stämme aus verschiedenen Biogasanlagen voneinander abgegrenzt werden (Abbildung 23b).

3.1.4.4 RIME (Random Inserted Mobile Element)

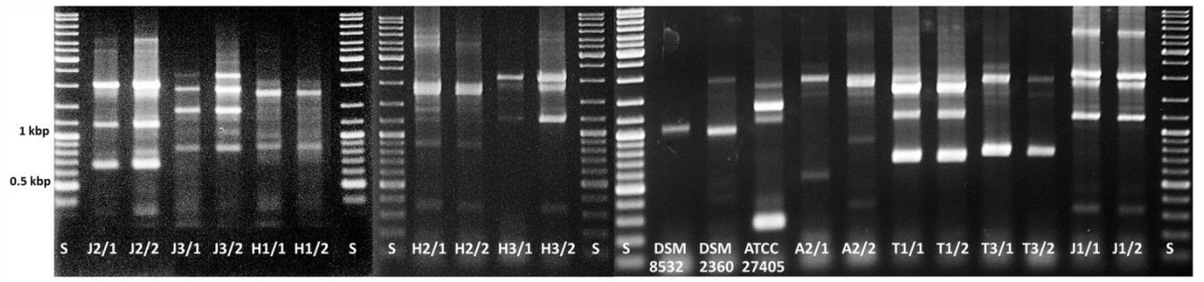


Abbildung 24: Ergebnisse der PCR Amplifikation der Regionen zwischen zwei Kopien das Mobilelements 1447 im Chromosom aller Isolierten Stämme und Referenzstämme DSM 8532 (C. stercorarium) und DSM 2360, ATCC 27405 (C. thermocellum) (Koeck et al. 2014d). *S = Molekulargewichtstandard*

In silico konnten 18 Kopien des mobilen Elements IS1447 in der Genomsequenz von *C. thermocellum* ATCC 27405 identifiziert werden (Zverlov et al. 2008), sowie 15 vollständige und zwei fragmentierte Mobilelemente im Genom von Stamm DSM 1313. Mit den PCR Primern IS1447F und IS1447R, komplementär zu den gegenläufigen Wiederholungssequenzen im Randbereich des mobilen Elements, konnte das Transposon in der genomischen DNA aller Isolate nachgewiesen werden. Die Amplifikation der Regionen zwischen zwei Mobilelementen lieferte sehr variable komplexe Bandenmuster (Abbildung 24). Die beiden Isolate einer Biogasanlage zeigen dabei bis auf Biogasanlage A2 das gleiche oder sehr ähnliche Muster, die sich nur in der Intensität der Banden unterscheiden. Da auch die aus der DNA von *C. stercorarium* eine schwache Bande amplifiziert werden konnte, besitzt auch dieser Stamm vermutlich mindestens eine Kopie von IS1447, welche auch mit 87 % Sequenzübereinstimmung in der *C. stercorarium* DSM 8532 Genomsequenz zu finden ist.

3.1.4.5 Zusammenfassung

Tabelle 42: Zusammenfassung der Ergebnisse von RAPD (Random amplified polymorphic DNA), VNTR (variable number of tandem repeats), Ribotypisierung und RIME (Random inserted mobile element) Assays für die Genotypisierung der C. thermocellum Isolate aus verschiedenen Biogasanlagen

Assay	Anzahl Isolate[a]	Anzahl Banden-muster	Anzahl Banden		Banden-größe (kb)	% Ähnlichkeit[c]	
			Total	Gemeinsam[b]		Mittel	Bereich
Ribotyp	11	1	1	1	~ 0.6	100	100
RAPD	12	4	2-4	2	1 – 3.5	77	66 -100
VNTR	12	6	1	0	0.9 – 1.2	-	0 - 100
RIME	11	11	2-7	0	0.3 - 5	24	0 – 66

[a] *Nur ein Stamm jeder BGA wurde mit einbezogen*

[b] *Gemeinsame Banden aller Bandenmuster einer Methode*

[c] *Ähnlichkeiten wurden mit dem „Coincidence index“ berechnet*

Die höchste Diskriminierung zwischen den Isolaten verschiedener Biogasanlagen lieferte die Amplifikation der Sequenzen zwischen den Mobilelementen IS1447. Zwei Bandenmuster verschiedener Biogasanlagen weisen im Schnitt nur eine Ähnlichkeit von 24 % auf (Tabelle 42).

3.1.5 SARAMIS MALDI-TOF MS CLUSTERANALYSEN DER SPEKTREN UNTERSCHIEDLICHER CLOSTRIDIUM THERMOCELLUM ISOLATE

Bei einem Vergleich der Zellproteinprofile durch eine Samaris MALDI-TOF MS Analyse clustern die beiden Isolate einer Biogasanlage nur zum Teil. Die beiden Proben der Anlage H3 liegen relativ weit voneinander entfernt. Somit ist die MADLI TOF MS Methode bei Umweltisolaten nur für eine Auflösung bis auf Artebene sehr gut geeignet (Abbildung 25). Ähnlich wie bei der Ribotypisierung, konnte die MALDI-TOF MS Analyse die Isolate nicht eindeutig bis auf Subspezieslevel auflösen, aber alle *C. thermocellum* Isolate eindeutig von *C. stercorarium* abgrenzen. Ebenso wie die ribosomalen rRNA Operone scheinen auch die ribosomalen Proteine zwischen zwei *C. thermocellum* Stämmen hoch konserviert zu sein.

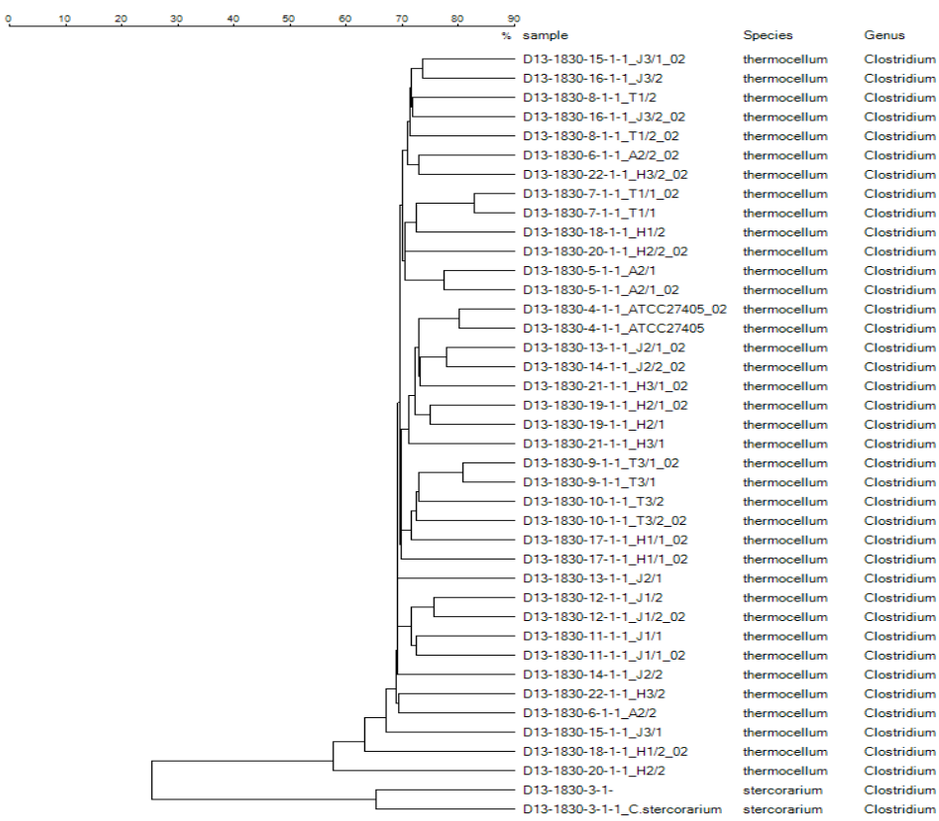


Abbildung 25: SARAMIS MALDI-TOF MS Clusteranalysen der Spektren unterschiedlicher C. thermocellum Isolate aus ausgewählten Biogasanlagen mit C. stercorarium Cluster

3.1.6 *Physiologische Unterschiede der Clostridium thermocellum Isolate aus unterschiedlichen Biogasanlagen*

Zur Ermittlung der physiologischen Unterschiede zwischen den Isolaten verschiedener Biogasanlagen wurden optimale Wachstumsbedingungen, Substratspektrum, cellulolytische Aktivität etc. bestimmt.

Tabelle 43: Zellproteinkonzentration von allen isolierten Stämme und C. thermocellum ATCC 27405 nach Inkubation für sechs Tage bei verschiedenen pH-Werten und Temperaturen in GS2 Medium mit 0.5 % Avicel (wt/vol).

	Temperatur °C					**pH**					
	45	50	55	60	65	6	6,4	6,8	7,2	8,3	9,2
A2/1	0	0,06	0,07	0,14	0,04	0,03	0,09	0,12	0,14	0,14	0
A2/2	0	0,05	0,06	0,11	0,04	0,01	0,07	0,11	0,11	0,16	0
T1/1	0	0,05	0,04	0,19	0,04	0	0,07	0,12	0,19	0,17	0
T1/2	0	0,06	0,05	0,18	0,05	0	0,05	0,12	0,18	0,18	0
T3/1	0	0,07	0,06	0,06	0,07	0,02	0,02	0,05	0,06	0,10	0
T3/2	0	0,06	0,05	0,12	0,04	0,01	0,05	0,09	0,12	0,09	0
J1/1	0	0,06	0,09	0,18	0,06	0	0,06	0,16	0,18	0,19	0
J1/2	0	0,05	0,09	0,17	0,04	0	0,07	0,13	0,17	0,15	0
J2/1	0,02	0,08	0,09	0,07	0,02	0	0,06	0,05	0,03	0,01	0,07
J2/2	0	0,05	0,09	0,09	0,03	0,02	0,07	0,09	0,09	0,13	0
J3/1	0	0,05	0,07	0,11	0,03	0	0,08	0,08	0,11	0,14	0,02
J3/2	0	0,05	0,07	0,13	0	0	0,07	0,14	0,13	0,17	0,10
H1/1	0	0,07	0,09	0,11	0,03	0,03	0,15	0,10	0,11	0,12	0,08
H1/2	0	0,07	0,06	0,18	0,04	0,06	0,11	0,17	0,18	0,17	0,08
H2/1	0	0,07	0,10	0,10	0,03	0,06	0,07	0,08	0,06	0,05	0,02
H2/2	0	0,06	0,13	0,13	0,03	0,08	0,10	0,04	0,04	0,04	0
H3/1	0	0,07	0,07	0,12	0,05	0,05	0,05	0,18	0,12	0,08	0,07
H3/2	0,01	0,07	0,11	0,17	0,05	0,04	0,11	0,14	0,17	0,13	0,06
ATCC	0	0,05	0,04	0,11	0,05	0	0	0,08	0,11	0,12	0

Alle Stämme hatten die höchste Zellproteinkonzentration (und damit das optimale Wachstum) bei 50 bis 60°C Wachstumstemperatur. Bei Temperaturen über 60 °C sank die Zellproteinkonzentration wieder und bei 70 °C konnte keines der Isolate wachsen. Nur die Isolate J2/1 und H3/2 konnten bei 45 °C auf Cellulose wachsen (Tabelle 43). Der optimale pH Wert lag für alle Isolate, außer J2/1, H2/1 und H2/2, bei pH 7,2 und pH 8,3. Die Zellproteinkonzentrationen waren zwischen pH 7,2 und 8,3 signifikant höher ($p = 0,0001$) als bei pH 6. Die Isolate der Biogasanlage H2 scheinen niedrigere pH Werte zu tolerieren als die Isolate der anderen Biogasanlagen. Obwohl einige der Isolate (A2/1, A2/2, T3/1, T3/2, J2/2, H1/1, H1/2, H3/1, H3/2) auch noch bei pH 6,0 langsam wachsen konnten, waren die gemessenen Zellproteinkonzentration der Stämme aus Biogasanlage H2, mit 0,06 mg/ml für H2/1 bzw. 0,08 mg/ml für H2/2, signifikant höher ($p = 0,00112$).

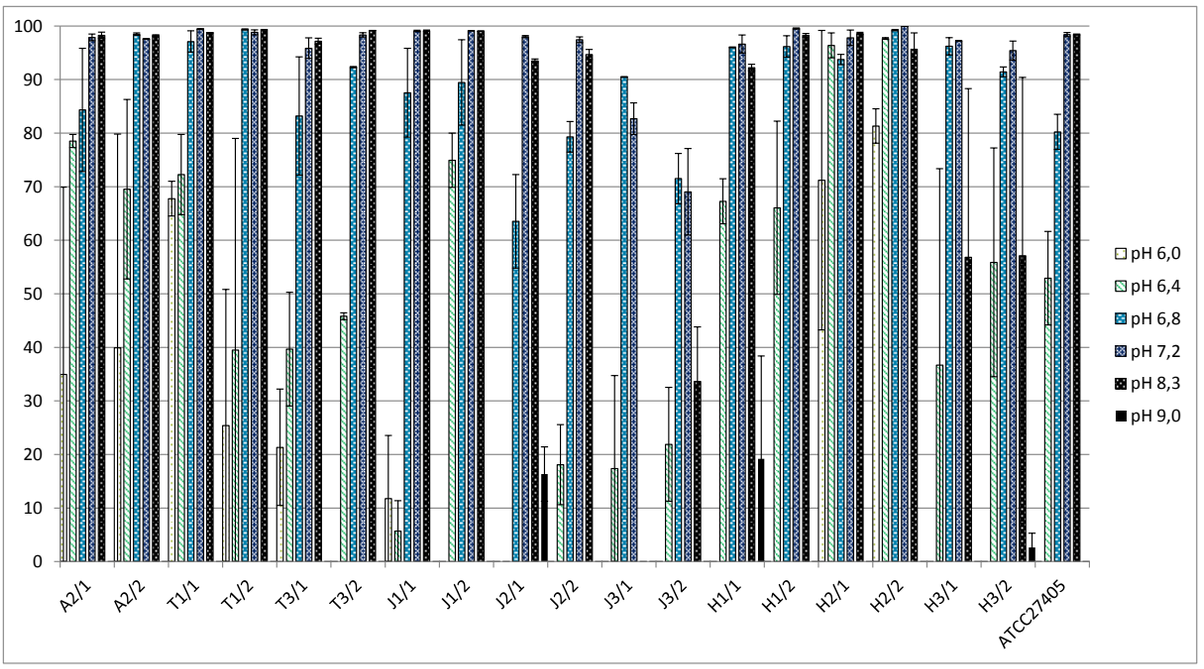


Abbildung 26: Cellulose Abbau (%) (y-Achse) aller Isolate und des C. thermocellum Typstammes ATCC 27405 bei verschiedenen pH Werten. Fehlerbalken repräsentieren die Standardabweichung

Auch der Celluloseabbau der Isolate der Biogasanlage H2 war bei pH 6,0 signifikant ($p = 0.002$) höher als bei allen anderen getesteten Isolaten (Abbildung 26). Die beste Abbauleistung lieferten alle Isolate bei pH 7,2 und 8,3. Nur die Isolate J2/1, H1/1 und H3/2 konnten bei pH 9,2 noch Cellulose abbauen. Die Ergebnisse des Celluloseabbaus zeigen eine positive Korrelation mit der Zellproteinkonzentration bei den entsprechenden Temperaturen/-pH Werten. In den meisten Fällen ergab sich ebenfalls eine positive Korrelation zwischen den beiden Isolaten einer Biogasanlage bei der Bestimmung des Celluloseabbaus und der Zellproteinkonzentration, besonders bei niedrigen pH Werten. Verglichen mit den anderen Isolaten wuchs der Typstamm am schlechtesten bei pH Werten und Temperaturen außerhalb des für die neuen *C. thermocellum* Stämme üblichen Bereichs.

Tabelle 44: Verwertung verschiedener Kohlenhydrate durch C. thermocellum Isolate in GS2 Medium. Bestimmung der Zellproteinkonzentration, um Wachstum nachzuweisen. Keines der Isolate konnte auf Xylan, Xylose, Stärke, Galactose, Mannose, Inulin, Sucrose, und Arabinose wachsen.

Substrat:	Cellulose	Glucose	Fructose	Sorbitol
A2/1	0,139	0,052	0,039	0,049
A2/2	0,113	0,049	0,042	0,031
T1/1	0,195	0,064	0,054	0,042
T1/2	0,177	0,084	0,051	0,038
T3/1	0,057	0,064	0,045	0,041
T3/2	0,125	0,042	-	0,06
J1/1	0,182	-	-	-
J1/2	0,169	-	-	-
J2/1	0,029	-	-	-
J2/2	0,091	-	-	-
J3/1	0,109	0,05	0,028	0,032
J3/2	0,128	0,067	0,027	-
H1/1	0,112	0,025	0,014	0,018
H1/2	0,184	0,054	0,026	0,02
H2/1	0,058	0,038	0,026	0,037
H2/2	0,035	-	0,022	0,038
H3/1	0,119	-	-	-
H3/2	0,167	-	-	-
ATCC 27405	0,108	-	0,077	0,128

Alle Isolate wuchsen, wie für bisher charakterisierte *C. thermocellum* Stämme beschrieben, auf kristalliner Cellulose (Filterpapier und Avicel) und Cellobiose (Freier et al. 1988, Mc Bee 1954). Nach längerer Inkubation für mindestens sieben Tage konnten einige Isolate auch Glucose, Fructose oder Sorbitol als Kohlenstoffquelle verwerten (Tabelle 44). Keines der Isolate konnte Xylan, Xylose, Stärke, Galactose, Mannose, Inulin, Sucrose, und Arabinose verwerten. In den meisten Fällen konnten die beiden Isolate einer Biogasanlage auf denselben Substraten wachsen, zeigten also auch auf der physiologischen Ebene einen hohen Grad an Übereinstimmung. Nach einer zweiten Überimpfung auf Glucose, Fructose und Sorbitol konnte eine Trübung des Mediums bereits nach zwei Tagen Inkubation beobachtet werden. Eine 16S rRNA Sequenz Analyse bestätigte, dass es sich nach der zweiten Überimpfung immer noch um *C. thermocellum* Reinkulturen handelte. Die Adaption ging nach einer darauf folgenden Kultivierung auf Cellulose wieder verloren. Der Typstamm ATCC 27405 konnte ebenfalls auf Fructose und Sorbitol angezogen werden.

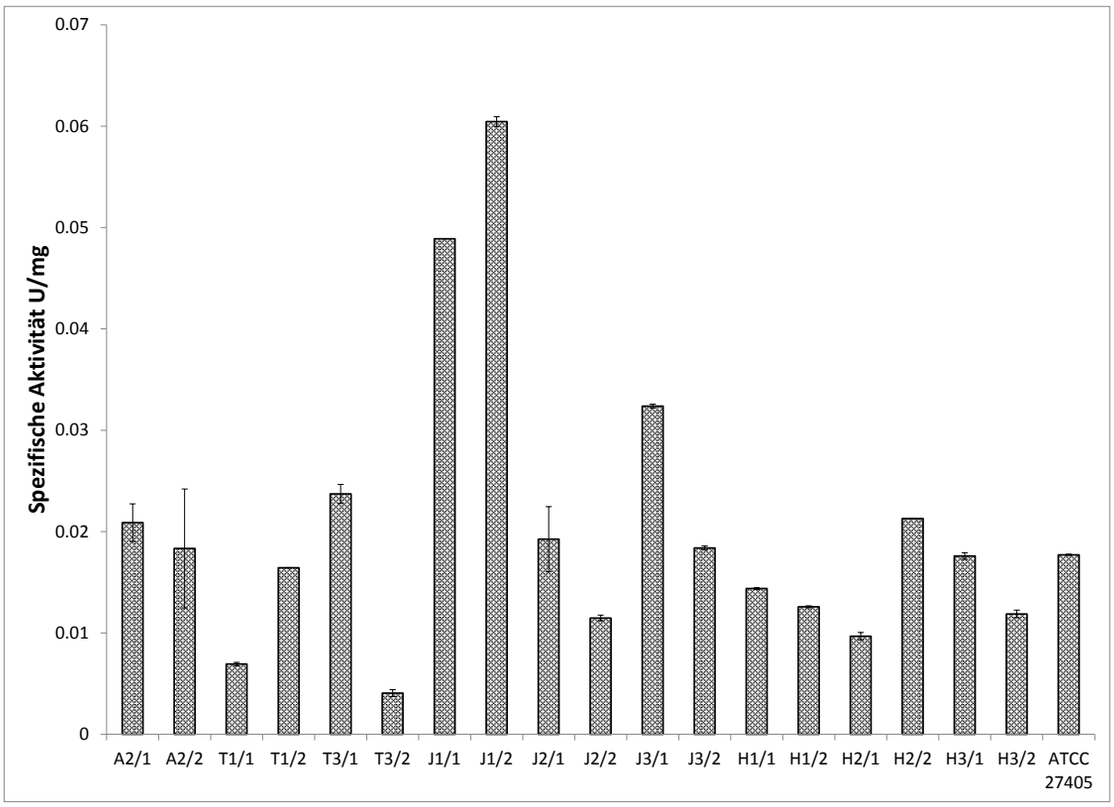


Abbildung 27: DNSA Assay zur Ermittlung der spezifischen Cellulase Aktivität des Kulturüberstandes aller Isolate und C. thermocellum ATCC 27405 auf Avicel unter Standardbedingungen. Ausgedrückt als µmol Glucose Äquivalent pro mg Protein aus Zellüberstand und Minute (U/mg). Von jeder Proteinprobe wurden drei Replikate gemessen und die numerische Standardabweichung mit Fehlerbalken dargestellt

Zur Ermittlung der spezifischen Cellulase Aktivität von allen Isolaten und *C. thermocellum* ATCC 27405, wurden diese in GS2 Medium mit Filter Papier (0,5 %) als C-Quelle bis zum vollständigen Abbau des Filterpapiers bei 60 °C angezogen. Die spezifische Cellulase-Aktivität des Kulturüberstandes wurde mittels DNSA Assay bestimmt. Diese war bei den Isolaten J1/1 und J1/2 signifikant höher ($p > 0.0001$) als bei allen anderen Isolaten. Die Stämme T1/1 und T3/2 zeigten die niedrigsten Aktivitäten.

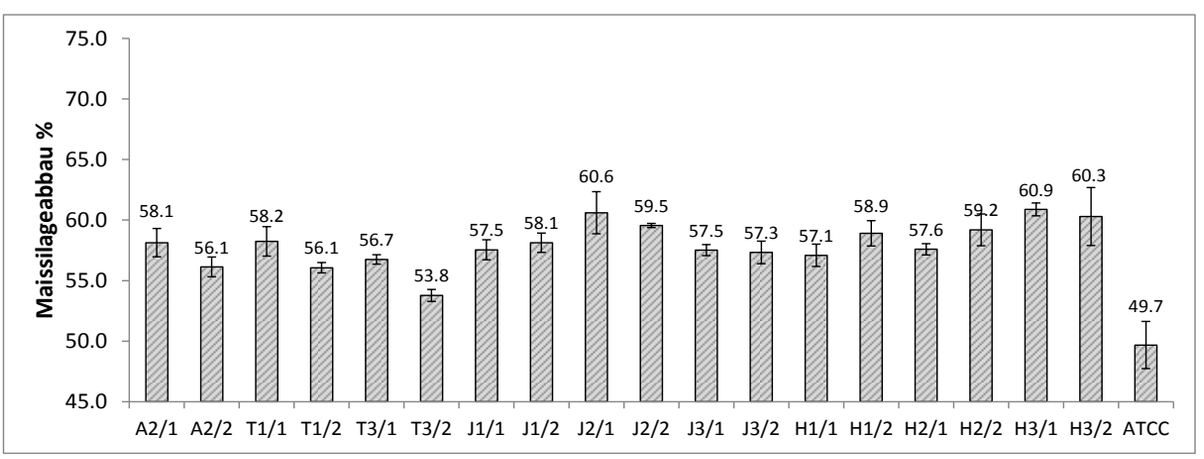


Abbildung 28: Abbau von Maissilage aller Isolate und des C. thermocellum Typstamms ATCC 27405 nach Inkubation in GS2 Medium mit 0.5 % Maissilage (wt/vol) für sechs Tage

Auf Maissilage, einem natürlichen Substrat, zeigten alle Isolate ähnliche Abbaueffizienzen (61-54 %). Der Typstamm dagegen zeigte einen signifikant (p = 0.0006) niedrigeren Abbau von Maissilage (49.7 %) (Abbildung 28).

3.1.7 Vergleich des Isolats Clostridium thermocellum BC1 mit Clostridium thermocellum ATCC 27405 und DSM 1313 auf enzymatischer und genomischer Ebene

In vorangegangenen Arbeiten wurde ein weiterer *C. thermocellum* Stamm BC1 aus einer thermophilen Kompostierungsanlage bei München (Deutschland) isoliert (Köllmeier 2013). Der Stamm zeigte, ähnlich wie einige der hier beschriebenen *C. thermocellum* Isolate aus den Biogasanlagen eine bessere Effizienz des Celluloseabbaus bei niedrigeren pH Werten (Abbau bis pH 6,2) und höheren Temperaturen (Abbau bis 67 °C) als der Typstamm ATCC 27405 und der Stamm DSM 1313. Die biotechnologische Anwendung des Stammes als Additiv für Biogasanlage zeigte eine positive Wirkung auf die Biogasproduktion aus Biomasse (Köllmeier 2013). Um zu überprüfen, ob diese natürliche Anpassung an niedrigere pH Werte an einer höheren Säuretoleranz des cellulolytischen Systems liegt, wurden mit Cellulosompräparationen der Stämme BC1, ATTC 27405 und DSM 1313 bei unterschiedlichen pH Werten in verschiedenen Puffersystemen Aktivitätstests auf Avicel, CMC und Birkenxylan durchgeführt. Die Aktivität von BC1 auf Birkenxylan war generell schlechter als die der beiden anderen Cellulosompräparationen. Das BC1 Cellulosom zeigte aber eine signifikant bessere Aktivität auf kristalliner Cellulose (Avicel) bei leicht sauren pH Werten (5,5 – 6,5) als die Cellulosomen der beiden anderen Stämme. Die Aktivität auf CMC war bei den Stämmen DSM 1313 und BC1 sehr ähnlich und bei ATCC 27405 signifikant geringer (Abbildung 29). Somit scheinen vor allen die Cellulasen von BC1 eine höhere Säuretoleranz zu besitzen.

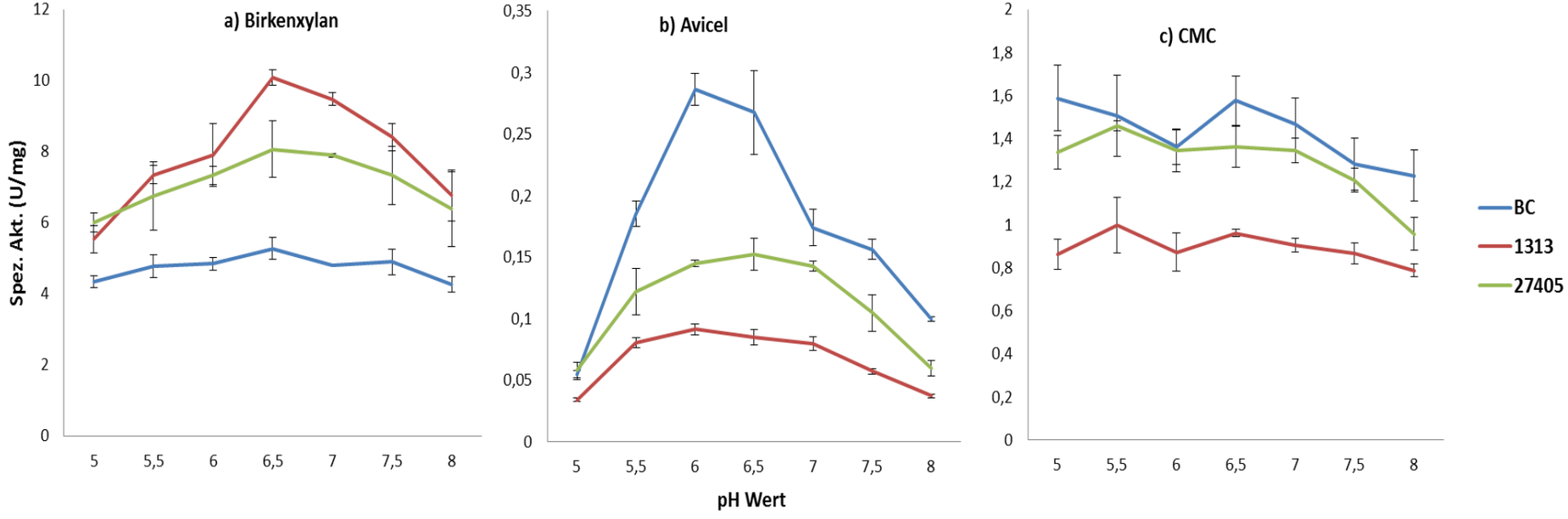


Abbildung 29: Spezifische Aktivität in Units pro Milligramm Enzym der Cellulosompräparationen von C. thermocellum BC1 (blau), DSM 1313 (rot), ATCC 27405 (grün) bei verschiedenen pH Werten auf Birkenxylan (a), Avicel (b) und Carboxymethylcellulose (c)

Ein weiterer Unterschied zwischen den Stämmen ist die Fähigkeit, Glucose als Kohlenstoffquelle direkt aus dem Medium zu verwerten. Dies wurde nur bei DSM 1313 und BC1 beobachtet, jedoch nicht bei ATCC 27405 (Ng and Zeikus 1981).

Eine Sequenzierung des BC1 Genoms sollte darüber Aufschluss geben inwieweit sich die Unterschiede in der enzymatischen Aktivität und Säuretoleranz auf Variationen in der Genomsequenz und Genausstattung zurückführen lassen. Hierzu wurde das annotierte Genom von BC1 mit den bereits in der Datenbank verfügbaren Genomen von DSM 1313 und ATCC 27405 verglichen. Die „Draft"-Genomsequenz von *C. thermocellum* BC1 mit 3.150.495 Einzelsequenzen (787,62 Mb) wurde mit Hilfe des Illumina MiSeq Systems am Zentrum für Biotechnologie in Bielefeld generiert. Die Assemblierung der Einzelsequenzen mit der GS *De Novo* Assembler Software (Version 2.6) lieferte ein nicht geschlossenes Genom bestehend aus 72 Einzelstücken (Koeck et al. 2013). Die Annotation wurde mit Hilfe der GenDB Plattform durchgeführt (Meyer et al. 2003).

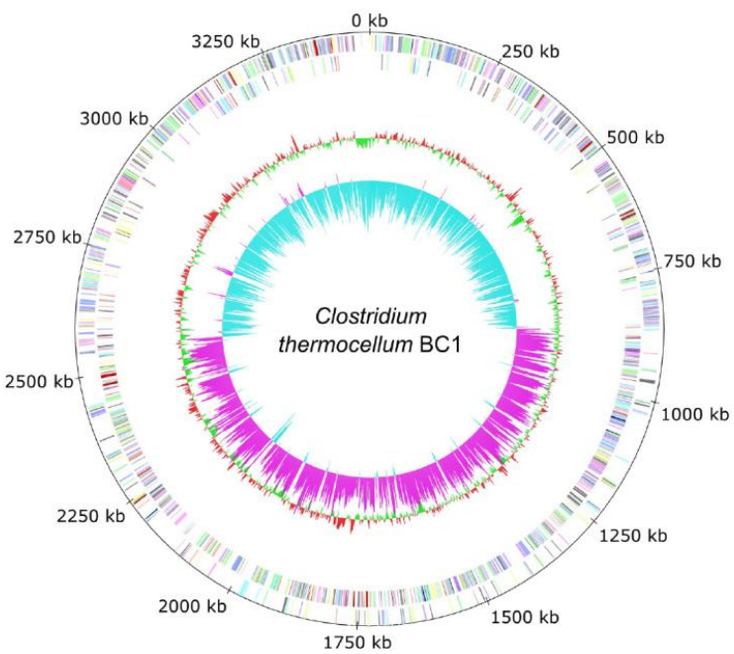


Abbildung 30 Zirkulärer Plot des C. thermocellum B1 Genom, von außen nach innen: annotierte Gene des Vorwärts-Strangs, Gene des Rückwärts-Strangs, GC Gehalt, GC Verhältnis (G - C)/(G + C)

Die wichtigsten Eigenschaften des Genoms sind im direkten Vergleich mit den Genomen von DSM 1313 und ATCC 27405 in Tabelle 45 zusammengefasst.

Tabelle 45: Wichtige Merkmale der Genome von Clostridium thermocellum BC1, DSM 1313 und ATCC 27405 im direkten Vergleich.

Merkmal	BC1	DSM 1313	ATCC 27405
Größe	3,45 Mb	3,56 Mb	3,84 Mb
GC Gehalt	39,1 %	39,1 %	39,0 %
kodierende Regionen	3.094	2.911	3.146
tRNAs	52	59	59
rrn-Operone	4	4	4

Durch eine Analyse und Interpretation der Genomdaten mit GenDB und EDGAR konnten alle bekannten *Clostridium thermocellum* Gene, die für den Abbau von kristalliner Cellulose notwendig sind, identifiziert werden. Eine dbCAN Analyse zur automatischen Annotation von den auf Kohlenhydraten aktiven Enzymen (CAZY), identifizierte 358 orthologe Gene im Genom von BC1 (Yin et al. 2012). Insgesamt sind die Genome von *C. thermocellum* BC1, *C. thermocellum* ATCC 27405 und *C. thermocellum* DSM 1313 untereinander hoch konserviert. Die komparative Analyse einzelner Cellulasen Gene (*celS, celR, celK, celJ*) zeigt innerhalb der drei verglichenen Genomen 99 % Indetität. Zum Vergleich der Reihenfolge von Genen oder Gensegmenten zwischen den Genomen, wurde ein sogenannter Syntenie Plot erstellt. Als Syntenie werden in der Genomik Gemeinsamkeiten in der Reihenfolge von Genen zwischen verschiedenen Genomen bezeichnet. Der Syntenie-Plot zeigt größere evolutionäre Ereignisse, wie z. B. Genom Neuordnungen (Inversion, Deletion, Insertion oder Translokation) an. Der Grad der Syntenie ist ein Maß für die genetische Verwandtschaft zwischen zwei Spezies. Je größer der Verwandtschaftsgrad, desto mehr ist auch die Anordnung der Gene konserviert. Der Syntenie-Plot zeigt auf der x-Achse die Position der Gene im Referenzgenom und auf der y-Achse die relative Position des entsprechenden orthologen Gens im Vergleichsgenom. Im Vergleich zum Referenzgenom von Stamm ATCC 27405 sind im Genom DSM 1313 ca. 50 % des Genoms invertiert und einige Einzelgene transloziert (Abbildung 31). Vergleicht man dagegen die Genomstruktur von BC1 mit ATCC 27405 haben keine Umstrukturierungen in der Anordnung der orthologen Gene stattgefunden.

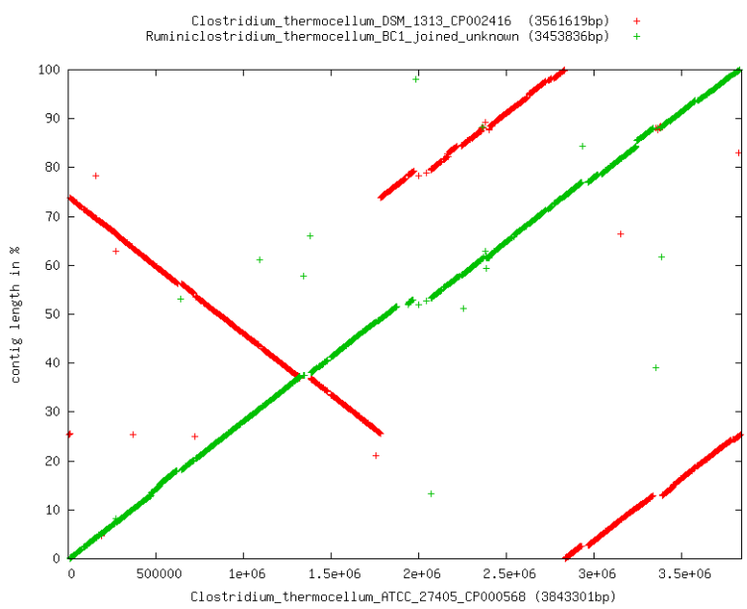


Abbildung 31: Syntenie Plot zur Darstellung der relativen Position der Gene von C. thermocellum DSM 1313 (rot) und BC1 (grün) im Vergleich zu Referenzgenom von ATCC 27405.

Das Venn Diagramm zeigt für jede Kombination der drei Genome gemeinsame und individuelle Gene („Singeleton“ Gene). Jedes Genom ist mit einer eigenen Farbe versehen, überlappende

Bereich sind in der Mischfarbe dargestellt. Das Core Genom, der Anteil gemeinsamer Leserahmen (kodierende Sequenzen) an der Gesamtzahl von Leserahmen in den Genomen, liegt bei über 88 %. Dies entspricht 2580 Genen (Abbildung 32).

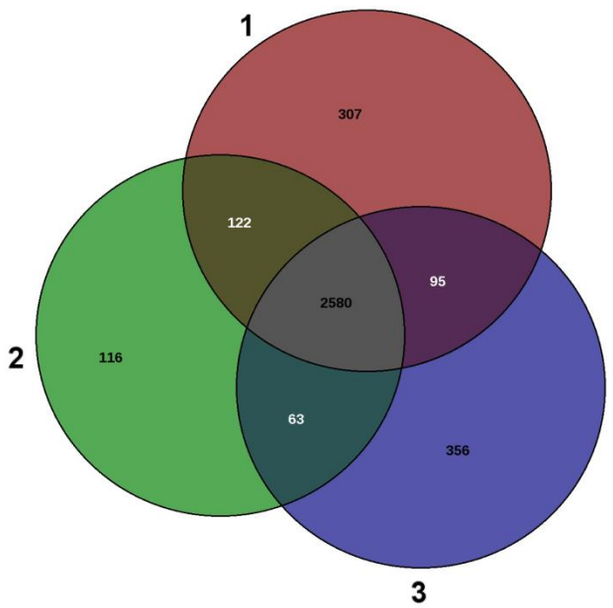


Abbildung 32: Venn Diagramm zur Darstellung des Core Genoms und der „Singelton" Gene zwischen den drei untersuchten C. thermocellum Stämmen (ATCC 27405 (1), DSM 1313 (2), BC1 (3)).

Individuelle Gene, sogenannte Singleton Gene, von *C. thermocellum* BC1 die bei den anderen beiden Stämmen fehlen, kodieren für Restriktionsenzyme und verschiedene Antibiotika Resistenz Faktoren. Auf der anderen Seite wurden im Genom von *C. thermocellum* BC1 keine Nitrogenase Gene gefunden, ein Hinweis, dass BC1 nicht fähig ist, atmosphärischen Stickstoff zu fixieren.

Basierend auf dem Core Genom wurde zusätzlich noch eine phylogenetische Analyse der Genome aller bisher sequenzierten *C. thermocellum* Stämme durchgeführt, welche in Abbildung 33 dargestellt ist.

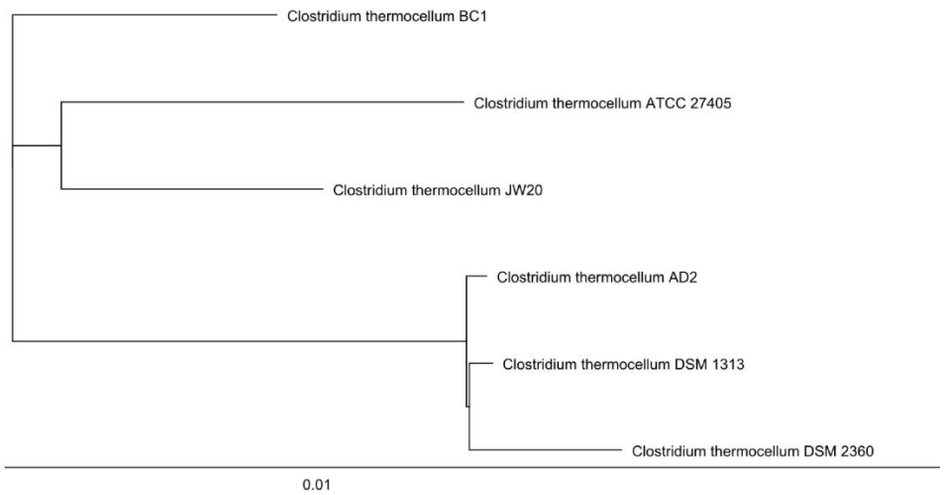


Abbildung 33: Phylogenetischer Baum basierend auf dem „Core" Genom aller bisher Sequenzierten C. thermocellum Stämme berechnet mit dem *Neigbour joining Algorithmus*

Insgesamt sind die Sequenzen der einzelnen Gene zwischen den drei untersuchten *C. thermocellum* Stämmen hoch konserviert und die Enzymausstattung sowie die zentralen Stoffwechselwege weitgehend identisch. Die Aminosäuresequenz homologer Glycosidasen (CelS, CelR, CelK) unterscheiden sich nicht oder maximal durch einen Aminosäureaustausch. Nur die Anordnung der Gene zueinander weicht teilweise ab. Letztendlich lassen sich aus der komparativen Analyse der Genome keine Rückschlüsse auf die Unterschiede in der Abbaueffizienz von Cellulose bei niedrigen pH Werten ziehen.

3.2 ISOLIERUNG UND CHARAKTERISIERUNG NEUER BAKTERIEN

Für die Anreicherung und Isolierung von Mikroorganismen aus Umweltproben gibt es diverse Ansätze. In dieser Arbeit lag der Focus auf der Isolierung von hydrolytischen, insbesondere cellulolytischen Bakterien aus verschiedenen Biogasanlagen, die mit pflanzlichen Substraten betrieben werden. Es wurden zahlreiche neuartige Isolierungs- und Reinigungsstrategien zur Gewinnung neuer Isolaten aus dem hoch komplexen mikrobiellen Konsortium in Biogasanlagen getestet und angewandt. Bei der Wahl der Isolierungsstrategien wurde angestrebt eine möglichst große Vielfalt hydrolytisch aktiven Bakterien in Reinkultur zu erhalten. Für eine maximale Diversität an Isolaten wurden zudem verschiedenste Medien, Additive und Kultivierungsbedingungen getestet. Im Wesentlichen wurden nach gründlicher Literaturrecherche vier Isolierungsstrategien ausgewählt und für die speziellen Anforderungen weiter entwickelt. Die einzelnen Strategien sind detailliert im Kapitel 2.6 beschrieben. In diesem Teil der Arbeit wird nur bei Isolaten, die von besonderem Interesse waren auf die individuelle Isolierungsstrategie eingegangen. Mit welcher Methode die restlichen Stämme gewonnen wurden, wird zugunsten der besseren Übersicht und aufgrund der geringen Relevanz nicht im Detail beschrieben.

3.2.1 ISOLATE AUS DER THERMOPHILEN BIOGASANLAGE 4BIOGAS NOVA FERMENTER (H1)

Als Modellbiogasanlage zur Isolierung möglichst vieler Arten/Stämme im thermophilen Bereich wurden der 4Biogas Nova Fermenter in Viersen (Deutschland) ausgewählt (vgl. 2.3). In der folgenden Tabelle sind alle mit den verschiedenen Methoden isolierten thermophilen Bakterien aus dieser Anlage aufgelistet (Tabelle 46).

Tabelle 46: Liste aller bisherigen Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten 16S rRNA Gen Sequenzähnlichkeit (laut NCBI Datenbankvergleich). Isolate wurden der gleichen Zeile zugeordnet, wenn ihre 16S rRNA Gensequenzähnlichkeit mindestens 98 % betrug. [1]Inc. sed. = incertae sedis

Nächster verwandter NCBI Datenbank Treffer			Isolate	Max. Indent. (%)	Indent. Isolate (min. 98 %)	Cellulose-abbau	Strategie
Klasse	**Familie**	**Spezies**					
Bacilli	*Bacillaceae*	*Bacillus thermoamylovorans*	K GR steril 6c, Neu19	99 %	2	-	1, 2
Clostridia	*Clostridiaceae*	*Clostridium isatidis*	RX1, RX3, CV1	99 %	3	-	1
		Clostridium putrefaciens	Neu23, Neu30	95 %	2	-	2
		Clostridium tertium	diverse	98 %	12	-	1, 2
		Clostridium thermopalmarium	Neu4	99 %	1	-	2
	Gracilibacteraceae	*Lutispora thermophila*	XV1	99 %	1	-	1
	Defluviitaleaceae	*Defluviitalea* sp.	GRX3	99 %	1	-	1
	Lachnospiraceae	*Mobilitalea sp.*	**T3/55**	94 %	1	**JA**	2
		Mobilitalea sp.	A, B, **SD1D,** SD1G, SD1I; SD1J, RK1PASC; GS2neu GRsteril direct	95 %	8	**JA**	2, 3
	Ruminococcaceae	*Acetivibrio cellulolyticus*	Neu18, Iso8/2b, GS2neu/4/2	98 %	3	-	2
		Clostridium caenicola	Neu21, Neu24, 5VG 1 PASC	98 %	3	-	1, 2
		Clostridium cellulosi	**diverse**	99 %	14	**JA**	1, 2
		Clostridium clariflavum	diverse	99 %	34	**JA**	1, 2
		Clostridium stercorarium	diverse	98-99 %	13	**JA**	1
		Clostridium thermocellum	diverse	99 %	27	**JA**	1, 2, 3
		Clostridium thermosuccinogenes	diverse	98 %	14	-	1, 2
	Inc. sed.[1]	*Thermoanaerobacterium thermosaccharolyticum*	diverse	99 %	24	-	1
		Tissierella creatinini	DG3	96 %	1	-	1
		Tepidimicrobium ferriphilum	GRX2; GRM1	97 %	2	-	1
		Tepidimicrobium ferriphilum	GRC5	99 %	1	-	1
		Tepidimicrobium xylanilyticum	GRC3, GRC4	99 %	2	-	1
		Tepidimicrobium xylanilyticum	GRC1, 5VG 1b MN	97 %	2	-	1
		Tepidimicrobium sp.	XV2	96 %	1	-	1
Summe					**172**		

Insgesamt wurden 172 thermophile Reinisolate aus der Biogasanlage Viersen mit den verschiedensten Isolierungsmethoden gewonnen. Davon sind 91 (56 %) in der Lage, in Reinkultur kristalline Cellulose abzubauen. Diese teilen sich auf fünf taxonomische Einheiten auf. Für die genaue Zuordnung zu einzelnen Taxa muss beachtet werden, dass zwei Stämme, die weniger als 97 % 16S rRNA Gensequenzähnlichkeit aufweisen, nicht als derselben Art zugehörig betrachtet werden können. Sequenzähnlichkeiten unter 95 % weisen auf unterschiedliche Gattungen hin. Dies muss aber durch andere Methoden wie z.B. DNA-DNA Hybridisierung oder Fettsäureanalyse bestätigt werden (Tindall et al. 2010). Die meisten Isolate konnten aufgrund ihrer 16S rRNA-Sequenzanalysen bekannten Spezies zugeordnet werden (91 %), wobei von den 15 nicht zuzuordnenden Isolaten neun in der Lage sind, Cellulose zu hydrolysieren. Diese können jedoch zwei nahe verwandten taxonomischen Einheiten zugeordnet werden. Jeweils ein Vertreter wurde für eine weitere, ausführliche Charakterisierung, sowie für die Genomsequenzierung ausgewählt (in fett markiert). Auf diese wird im Kapitel 3.2.3 näher eingegangen. Tabelle 46 zeigt die taxonomische Zuordnung der einzelnen Isolate. Alle Isolate stammen aus dem Phylum Firmicutes und lassen sich wiederum zwei Klassen (Clostridia, Bacilli) zuordnen. Von den 8 isolierten Hauptgattungen gehört jedoch nur *Bacillus* in die Klasse der Bacilli. Der überwiegende Anteil der cellulolytischen Isolate kann der Gattung *Clostridium* zugeordnet werden.

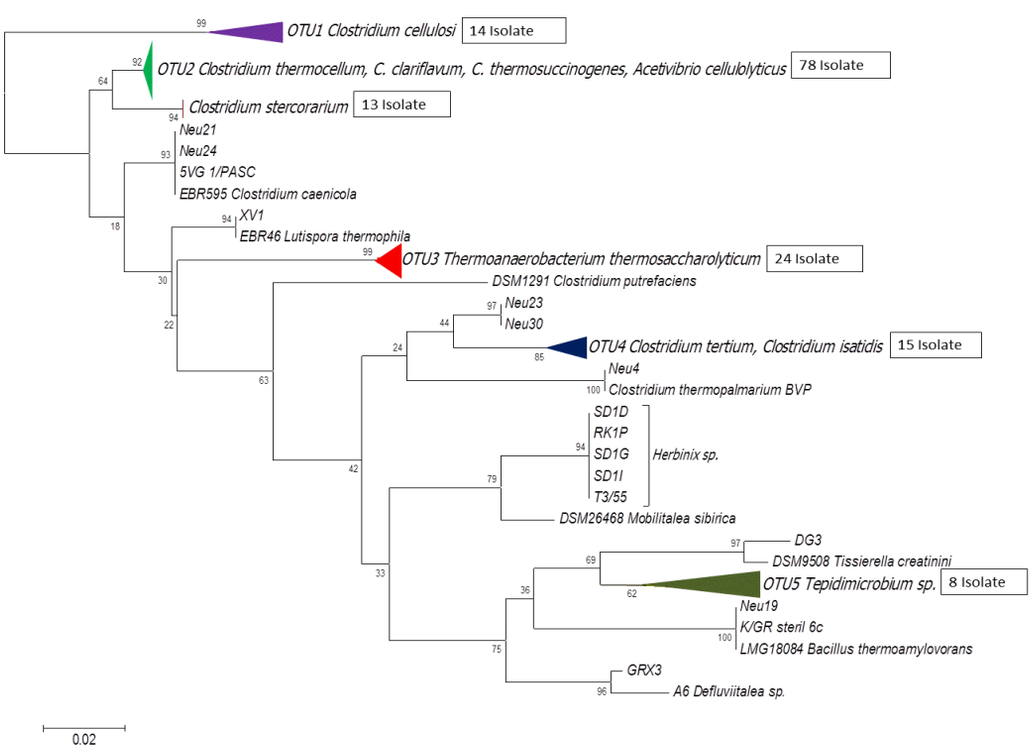


Abbildung 34: Ungewurzelter phylogenetischer Baum der 16S rRNA Gen Sequenzen aller thermophilen Isolaten und ausgewählter Referenzstämme unter Verwendung eines Maximum likelihood-basierten Algorithmus. Auf den Ästen sind die einzelnen Bootstrap Werte angegeben. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in „Operational taxonomic units" (OTU`s) zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Isolate, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede.

Vergleicht man die Isolierungsstrategien wird deutlich, dass der überwiegende Anteil der Isolate mit Strategie 1 gewonnen wurde (einfache Anreicherung in Flüssigmedium). Vor allem aufgrund der Anreicherung auf verschiedenen Kohlenstoffquellen (Xylan, Cellulose, Maissilage etc.) und der größeren Variation der Medien wurden auch bei Strategie 1 die meisten unterschiedlichen Isolate generiert (Tabelle 47). Jedoch waren dies meist Vertreter bereits charakterisierter Spezies. Neue, bisher unbekannte cellulolytische Spezies wurden nur mit Strategie 2 (Verdünnungsreihen in Flüssigmedium) und Strategie 3 (Verdünnungsreihen auf Festmedium) gewonnen. Diese bilden im phylogenetischen Baum aller Isolate ein eigenes Subcluster mit dem erst kürzlich beschriebenen Bakterium *Mobilitalea sibirica* als nächsten Verwandten (Podosokorskaya et al. 2014) (Abbildung 34).

Tabelle 47: Zusammenfassung der bisherigen Isolate mit drei Isolierungsstrategien

Isolierungsstrategie	Gesamtzahl Isolate	Anhand der 16S rRNA Sequenz unterschiedliche Isolate (min 5% Sequenzunterschied)
1	104 (60,5 %)	18
2	54 (31,4 %)	12
3	14 (8,1 %)	2

Aufgrund fehlender Referenzspektren konnte bis auf *C. thermocellum* keines der Isolate zuverlässig mittels Zellproteinprofilanalyse durch MALDI TOF MS identifiziert werden.

3.2.2 *Metagenomanalyse der thermophilen Biogasanlage 4Biogas Nova Fermenter (Viersen)*

Für die taxonomische Evaluierung des 4Biogas Nova Fermenters (Viersen) wurde eine umfangreiche Metagenomanalyse durchgeführt. Hierzu wurde von verschiedenen Institutionen Zentrum für Biotechnologie Universität Bielefeld (CeBiTec), Hochschule für Angewandte Wissenschaften Hamburg (HAW), Johannes Gutenberg-Universität Mainz (JGU), TU München (TUM), Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)) genomische DNA mit verschiedenen Methoden isoliert. Die Proben wurden alle zum gleichen Zeitpunkt aus dem Fermenter entnommen. Die Sequenzierung wurde vom Zentrum für Biotechnologie in Bielefeld mit dem Illumina MiSeq System durchgeführt. Die Metagenomanalyse sollte Aufschluss über die exakte kultivierungsunabhängige Zusammensetzung der mikrobiellen Gemeinschaft geben und aufdecken ob die Methode der DNA-Isolierung die Zusammensetzung des taxonomischen Profils beeinflussen kann. Die generierten Sequenzen wurden anschließend mit der Analyseplattform MGX analysiert.

Tabelle 48: DNA-Isolierungsmethoden für die Metagenomanalyse der Biogasanlage und Anzahl der generierten und bereinigten Einzelsequenzen

Probe	Methode	Bereinigte Einzelsequenzen
A (CebiTec)	Chloroform-Isoamyl Alkohol DNA Extraktion	520.357
B (ATB)	Phenol-Chloroform Extraktion mit „*Bead Beating*“ Schritt	499.640
C (HAW)	Chloroform-Isoamyl Alkohol DNA Extraktion	632.251
D (JGU)	GeneMATRIX Stool DNA Purification Kit	685.903
E (TUM)	Chloroform Extraktion mit Isopropanol Fällung	583.417

Die Gesamtzahl der bereinigten Einzelsequenzen die von den verschiedenen DNA-Proben generiert wurden, liegt in derselben Größenordnung. Die klassische Chloroform-Isoamyl-Alkohol Methode (A und C) lieferte die höchste durchschnittliche Anzahl klassifizierbarer bakterieller Sequenzen. Zwischen den einzelnen Proben war die Verteilung der Sequenzen auf die Familien sehr ähnlich. Insgesamt wurde in allen Proben die Gemeinschaft der Bakterien von der Familie Clostridia dominiert, mit größerem Abstand gefolgt von den Thermotogae und Bacilli (Abbildung 35). In Probe B waren Sequenzen, die der Familie Thermotogae zugeordnet werden konnten, häufiger vertreten als in den restlichen Proben. Generell zeige die Wahl der DNA-Isolierungsmethode wenig Einfluss auf das taxonomische Profil der Metagenomanalyse.

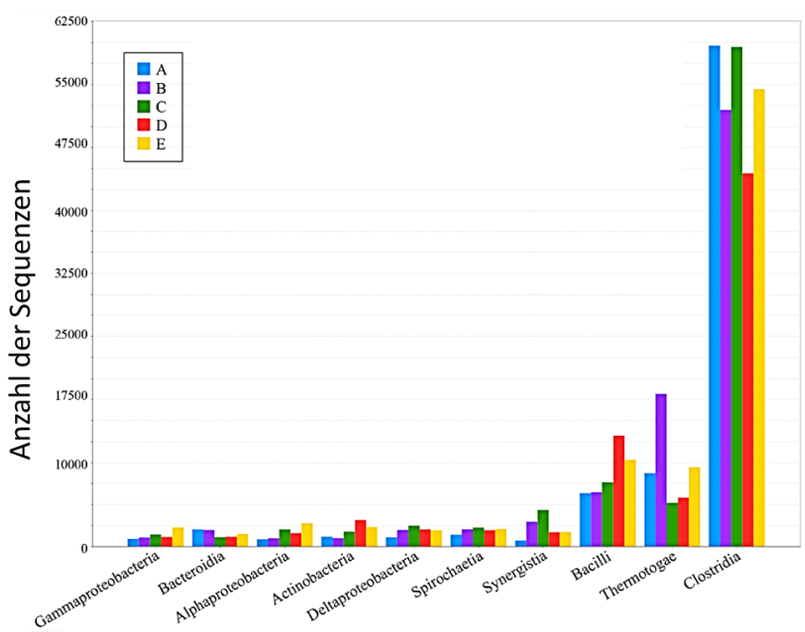


Abbildung 35: Verteilung der Metagenomsequenzen aus verschiedenen DNA Proben auf die bakteriellen Taxa (Die bezeichnungen sind durch das Programm MGX vorgegeben).

Bei der Darstellung aller Sequenzen in einem einzigen Kronaplot sieht man die Verteilung der Sequenzstücke auf die verschiedenen Taxa und erhält daher ein ähnliches Ergebnis wie für die Einzelproben (Abbildung 35). Das Reich der Bakterien wird mit über 60 % der Sequenzen vom

Phylum Firmicutes dominiert, gefolgt von den Thermotogae. Etwa 10 % der Sequenzen konnten keinem Phylum zugeordnet werden. Innerhalb der Firmicutes dominierten wiederum die Klassen Clostridia und Bacilli mit über 90 % der Sequenzen. Daraus resultieren die Clostridiales bzw. Clostridiaceae als die dominierende Ordnung bzw. Familie. 25 % aller Sequenzen konnten der einzelnen Gattung *Clostridium* zugeordnet werden (Abbildung 36).

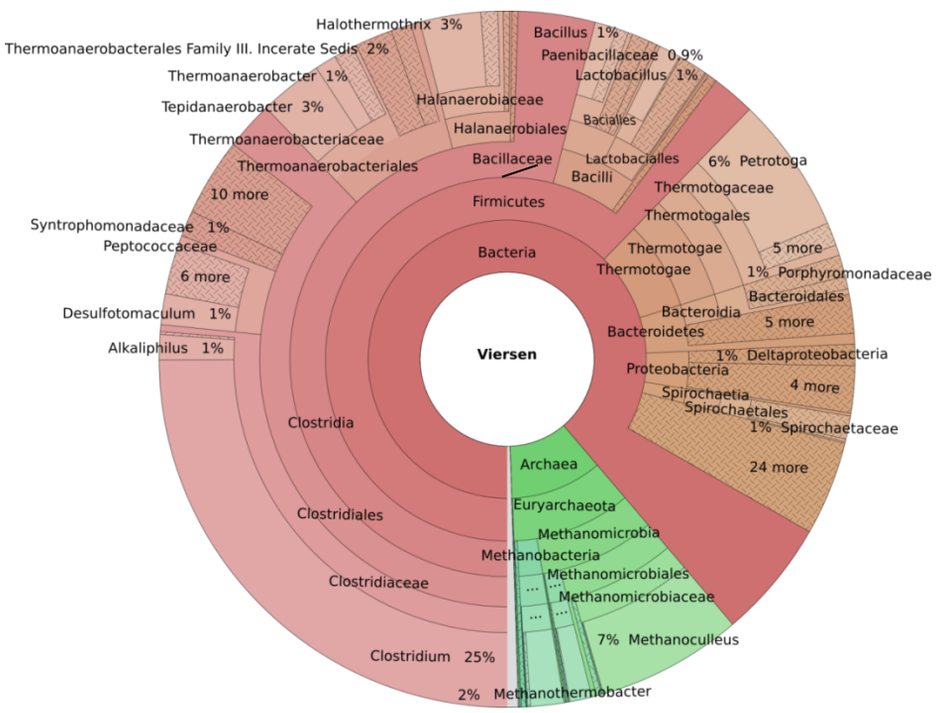


Abbildung 36: Kronaplot aller Sequenzen der Metagenomanalyse der Biogasanlage Viersen

Innerhalb der Gattung überwiegen wiederum die Spezies *Clostridium thermocellum* und *Clostridium clariflavum* (Abbildung 37). Beide Arten sind in der Lage kristalline Cellulose zu verwerten.

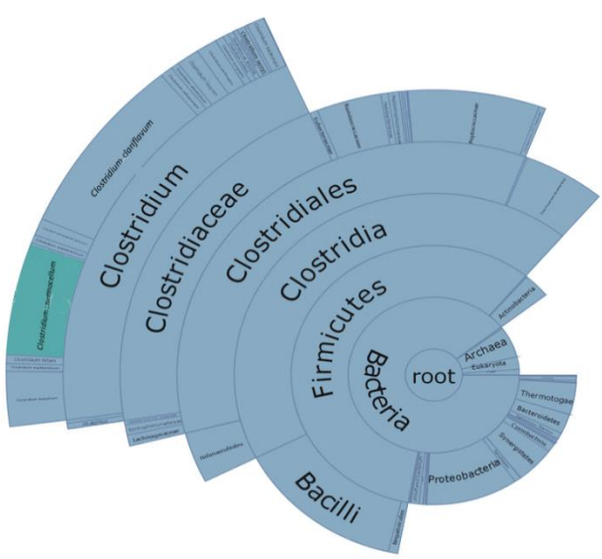


Abbildung 37: Radiale Darstellung der taxonomischen Hauptgruppen aus dem Metagenom der Biogasanlage Viersen

3.2.3 *CHARAKTERISIERUNG AUSGEWÄHLTER THERMOPHILER ISOLATE*

Insgesamt wurden vier interessant erscheinende thermophile Isolate zur Genomsequenzierung und weiteren Charakterisierung ausgewählt, die in Tabelle 49 zusammengefasst sind. Die taxonomische Einordnung kann aus Abbildung 34 und Tabelle 46 entnommen werden. Auswahlkriterien waren vor allem besondere Eigenschaften (z.B. sehr gute Cellulose-Hydrolyse) und Häufigkeit des Isolates. Bis auf Isolat 1A1 stammen alle Isolate aus der thermophilen Biogasanlage H1 (Viersen). 1A1 wurde aus einer sauren Anreicherungskultur (pH 6,2) von Proben einer thermophilen Biogasanlage bei München in Kokultur mit *C. thermocellum* isoliert: Bei 52 °C aus einer Anreicherung in GS2 Medium mit Filterpapier (Isolierungsstrategie 1). In einen zweiten Reinigungsschritt mit Cellobiose statt Cellulose als C-Quelle wurden die Kulturen aufgetrennt. Isolat DG5 wurde ebenfalls mit Strategie 1 aus einer Anreicherung in DNB (Dilute nutrient broth) Medium mit 10 % Gärrest (v/v) und Filterpapier isoliert. Die Isolate T3/55 und SD1D wurden mit Strategie 2 aus einer Verdünnungsreihe in flüssigem GS2 Medium mit 0,5 % steril filtriertem Rumensaft (v/v) gewonnen und mit Agarplatten des selben Mediums mit PASC-Topagar gereinigt.

Tabelle 49: Liste der zur genaueren Charakterisierung ausgewählten thermophilen Isolate. BGA = Biogasanlage

Name Isolat	Max. Sequenzident. (NCBI)	Ursprung	EMBL/NCBI Zugangsnummern	DSMZ Nummer
1A1	*Bacillus thermoamylovorans (99 %)*	Saure (pH 5,6) Anreicherungskultur	Genom: PRJEB6526	29353
T3/55	*Herbinix hemicellulosilytica (100 %)*	BGA H1 (Viersen)	Genom: CVTD01000001 - CVTD01000075 16S rRNA: LN626355	29228
SD1D	*Herbinix hemicellulosilytica (96 %)*	BGA H1 (Viersen)	16S rRNA: LN626359	100831
DG5	*Clostridium cellulosi (99 %)*	BGA H1 (Viersen)	Genom: LM995447	29189

3.2.3.1 Physiologische Charakterisierung ausgewählter thermophiler Isolate

Zur physiologischen Charakterisierung der Isolate wurden zunächst Substratspektrum, Produktspektrum, pH-, Temperaturoptimum und weitere wichtige morphologische und physiologische Eigenschaften bestimmt. Die Ergebnisse sind in Tabelle 50 zusammengefasst.

Tabelle 50: Wichtige morphologische und physiologische Eigenschaften von den Isolaten T3/55, SD1D, 1A1, DG5.

Characteristika	1A1	DG5	T3/55	SD1D
Temperatur (°C)	40 -58	40 - 65	45 - 62	40 - 62
pH	5,4 – 8,5	6,2 – 8,5	6,5 – 9,0	6,5 – 8,5
Katalase	+	–	–	–
Gram Reaktion	+	–	+	+
Flagella	+	+	+	n.b.
Zellmorphologie	Stäbchen	Stäbchen	Stäbchen	Stäbchen
Endosporen	+	+	–	–
Beweglichkeit	+	+	–	–
Abbau von:				
Galactose	+	+	–	+
Arabinose	+	–	+	+
Xylan	+	n.b.	+	–
Xylose	n.b.	+	+	–
Cellulose*	–	+	+	+
Mannose	–	+	+	+
Fructose	+	+	+	–
Sucrose	+	+	+	–
Glucose	+	+	+	+
Cellobiose	+	+	+	+
Sorbitol	+	–	–	–
Stärke	+	+	+	+
Haupt-Fermentationsprodukte	Acetat, Ethanol, und Formiat	H_2, Ethanol, Acetat	H_2, Ethanol, Acetat, Butyrat, Propionat	H_2, Ethanol, Acetat, Butyrat

**Filterpapier (Whatman No. 1); n.b. = nicht bestimmt*

Eine Sequenzierung des 16S rRNA Gens zeigte, dass Isolat 1A1 zu 99% mit der 16S rRNA Gen Sequenz des *Bacillus thermoamylovorans* Typstammes DKP^T identisch ist. Dies rechtfertigt die Zuordnung von Isolat 1A1 zur Spezies *B. thermoamylovorans*, wobei 1A1 das erste *B. thermoamylovorans* Isolat aus einer industriellen Biogasanlage ist. Bereits bekannte *B. thermoamylovorans* Stämme aus dem Phylum *Firmicutes* und der Familie *Bacillaceae* wurden ursprünglich aus mesophilen Habitaten wie senegalesischem Palmwein, Melkständen, Stroh und Reinräumen der Luft- und Raumfahrt isoliert (Coorevits et al. 2011, Stieglmeier et al. 2009). Obwohl *B. thermoamylovorans* 1A1 sein Wachstumsoptimum bei einer Temperatur von 50°C hat und unter 40°C nicht wächst, sind andere Isolate dieser Spezies in zahlreichen mesophilen Lebensräumen zu finden. Die Spezies scheint eine wichtige Rolle innerhalb bakterieller Gemeinschaften, die am anaeroben Abbau pflanzlicher Biomasse beteiligt sind, einzunehmen (Lakaniemi et al. 2011, Wang et al. 2003). Alle Isolate sind fakultativ anaerob, Gram-positiv, amylolytisch und sporulierend. Der Stamm 1A1 wächst anaerob auf diversen Poly- und Monosacchariden wie Xylan, Pectin und Stärke. Dies lässt vermuten, dass er besonders bei leicht sauren pH Werten eine wichtige Rolle in der hydrolytischen Phase des Biomasseabbaus in Biogasanlagen spielt. Der Typstamm kann dagegen weder auf Xylanen noch auf Pectin wachsen (Combet-Blanc et al. 1995). Da der Stamm *B. thermoamylovorans* 1A1 ursprünglich in Kokultur mit einem *C. thermocellum* Stamm isoliert wurde, wurde vermutet, dass bei dem Abbau von natürlichem Substrat (Maissilage) ein Synergismus besteht. Dies konnte jedoch experimentell nicht bestätigt werden (Abbildung 38).

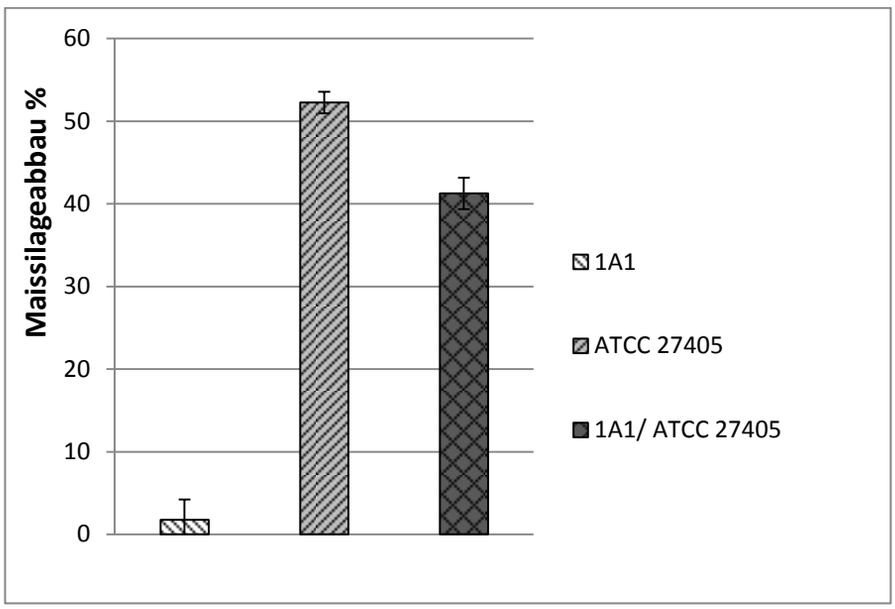


Abbildung 38: Abbau von Maissilage durch Clostridium thermocellum ATCC 27405, Bacillus thermoamylovorans 1A1 und einer Kokultur aus beiden. Fehlerbalken repräsentieren die Standardabweichung von drei Messwerten

Das Isolat DG5 konnte aufgrund einer 99 %igen Sequenzübereinstimmung mit dem Typstamm AS 1.1777 der Spezies *Clostridium cellulosi* zugeordnet werden. *Clostridium cellulosi* gehört innerhalb des Phylums Firmicutes zur Familie der *Ruminococcaceae* und wurde bisher nur aus Kuhmist und Kompost isoliert (He et al. 1991, Yutin and Galperin 2013). Es handelt sich um ein thermophiles, cellulolytisches Bakterium, welches Wasserstoff, Ethanol und Acetat als Fermentationsprodukte produziert. *C. cellulosi* DG5 ist ein beweglicher Sporulierer, der optimal zwischen 55 und 60°C und bei pH 7,3 – 7,5 wächst (Tabelle 50). Genauso wie der Typstamm AS 1.1777 konnte das neue Isolat anaerob eine große Bandbreite an Poly- und Monosacchariden verstoffwechseln (He et al. 1991). DG5 ist das erste bekannte *C. cellulosi* Isolat welches aus einer großtechnischen Biogasanlage stammt. Bisher konnte *C. cellulosi* in Metagenomanalysen von Biogasfermentern nicht als Bestandteil des für die Hydrolyse verantwortlichen mikrobiellen Konsortiums identifiziert werden (Krause et al. 2008, Schlüter et al. 2008, Wirth et al. 2012). Allerdings weisen aktuelle Ergebnisse von anderen Studien darauf hin, dass diese Spezies eine wichtige Rolle bei der Substrathydrolyse in Biogasanlagen spielen könnte (Nissila et al., 2011; Ueno et al., 2001; Zhang et al., 2014). DGGE (diskontinuierliche Gradientengelelektrophorese) Analysen haben gezeigt, dass in Kompostproben vor allem mit *C. cellulosi, C. stercorarium* und *C. thermocellum* eng verwandte Bakterien für den Celluloseabbau verantwortlich sind (Nissila et al. 2011, Ueno et al. 2001). Dabei sind vor allem *C. cellulosi* und *C. stercorarium* verantwortlich für den Hemicelluloseabbau (Nissila et al. 2011). Zhang et al. (2014) haben einen neuen Stamm von *C. stercorarium* (CS-3-2) und *C. cellulosi* (CS-4-4) aus verrottenden Strohhalmen isoliert (Zhang et al. 2014). Die Endo-Glucanase und Exo-Glucanase Aktivitäten im Kulturüberstand von Stamm CS-4-4 waren etwa 3 bis 5 mal höher als die von *C. stercorarium* CS-3-2, was wiederum ein Hinweis auf das hohe Lignocellulose Hydrolysepotential von *C. cellulosi* ist (Zhang et al., 2014). Außerdem

konnte *C. cellulosi* eine größere Bandbreite verschiedener Kohlenhydrate abbauen als *C. stercorarium* und *C. thermocellum* (He et al. 1991).

Die Isolate SD1D und T3/55 bilden, basierend auf den Ergebnissen der phylogenetischen Analyse der 16S rRNA Gensequenzen, einen neuen Genus innerhalb der Familie *Lachnospiraceae*. *Lachnospiraceae* ist eine Familie innerhalb der Clostridia, welche bisher hauptsächlich Mikroorganismen des Gastrointestinal-Traktes von Säugetieren beinhaltet (Meehan and Beiko 2014), speziell aus Wiederkäuern (Kittelmann et al. 2013) und Menschen (Gosalbes et al. 2011). Vertreter der Familie sind dabei hauptsächlich verantwortlich für den Abbau von komplexem Pflanzenmaterial. Die Familie *Lachnospiraceae* wurde erstmals von Rainey (2009) auf Basis von phylogenetischen Analysen der 16S rRNA Gensequenzen beschrieben (Euzéby 2010, Ludwig et al. 2009). Die Familie umfasst derzeit 25 Genera (Parte 2014). Alle Spezies sind strikt anaerob und bilden keine Sporen. Sequenzähnlichkeitsberechnungen mit der kompletten 16S rRNA Sequenz von Isolat T3/55 ergaben *Mobilitalea sibirica* als nächsten Verwandten unter den gültig beschriebenen Taxa, mit nur 93,9 % 16S rRNA Sequenzähnlichkeit. Daher repräsentiert das Isolat T3/55 den Typstamm der kürzlich im Rahmen dieser Arbeit beschriebenen neuen Gattung und Spezies *Herbinix hemicellulosilytica* (Koeck et al. 2015b). Das Isolat SD1D gehört mit 97 % Sequenzähnlichkeit zu T3/55 auch zur Gattung *Herbinix,* zählt allerdings gemäß den Kriterien von Tindall et al. (2010) zu einer bisher nicht beschriebenen neuen Spezies. Beide Isolate unterscheiden sich vor allem in ihrem Substratspektrum und in ihrer Salztoleranz. T3/55 war in der Lage, eine größere Bandbreite an Kohlenhydraten zu verstoffwechseln als SD1D, darunter Xylan, Xylose, Fructose und Sucrose. SD1D konnte dafür im Gegensatz zu T3/55 auf Galactose wachsen (Tabelle 50). Während Isolat SD1D bei 4 % NaCl (w/v) im Medium noch problemlos wächst, toleriert Isolat T3/55 maximal 1 % NaCl. Morphologie und Produktspektrum waren weitgehend identisch. Ein Vergleich des Zellwachstums von T3/55, basierend auf der Zellproteinkonzentration, auf verschiedenen Substraten mit der Konzentration der Abbauprodukte zeigte eine direkte Korrelation. Von den Substraten, auf denen die Zellen am besten wuchsen, konnten auch die höchsten Konzentrationen der Abbauprodukte gemessen werden (Abbildung 39).

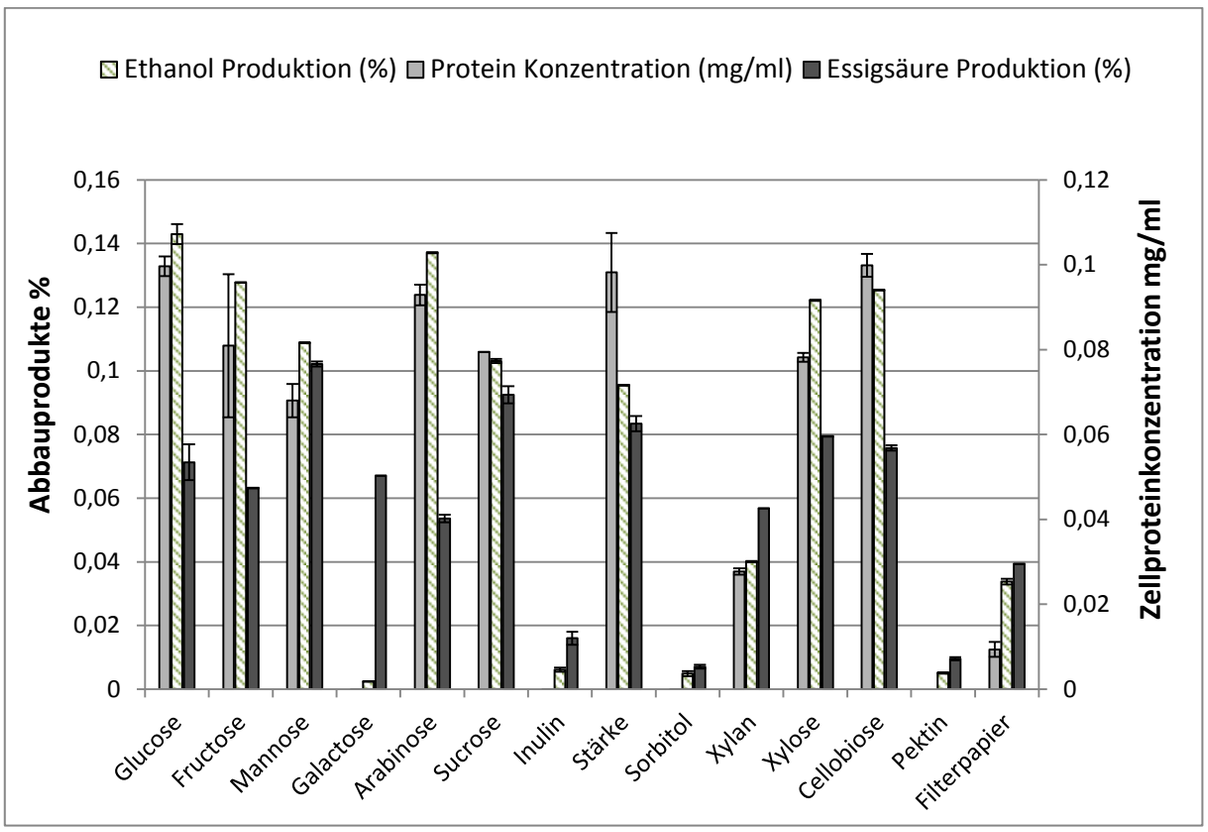


Abbildung 39: Vergleich der Zellproteinkonzentration von T3/55 auf verschiedenen Substraten mit der Konzentration der Abbauprodukte nach sieben Tagen Inkubation bei 55 °C in GS2 Medium

Die Aktivität der Enzyme von SD1D und T3/55 im Zellüberstand wurde zunächst auf Xylan, ß-Glucan, CMC, PASC und Avicel getestet (Tabelle 51). Die Aktivität der beiden Isolate auf CMC ist gleich, von den anderen Substraten setzen die Enzyme von SD1D mehr reduzierende Zuckerenden frei. Selbst auf Xylan ist die Aktivität von SD1D höher, obwohl das Isolat nicht auf Xylan wachsen konnte. Beide Isolate konnten Cellulose (Filterpapier) abbauen und zeigten auch Enzymaktivität auf kristalliner Cellulose (Avicel).

Tabelle 51: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat T3/55 und SD1D auf verschiedenen Substraten mit Standardabweichung von drei Messwerten

Isolate	Xylan	ß-Glucan	CMC	PASC	Avicel
T3/55	2,27 +/- 0,13	0,23 +/- 0,001	0,09 +/- 0,002	0,02 +/- 0,0007	0,008 +/- 8,66 * 10^{-6}
SD1D	3,58 +/- 0,0003	1,28 +/- 0,05	0,09 +/- 0,003	0,06 +/- 0,0012	0,02 +/- 0,0003

Exemplarisch für beide Vertreter der neuen Gattung *Herbinix* wurden mit dem Isolat T3/55 weitere Enzym- und Physiologietests durchgeführt.

Da T3/55 das breitere Substratspektrum hat, wurde die Enzymaktivität des Zellüberstands zusätzlich auf diversen Hemicellulosen getestet. Dabei wurde *Clostridium stercorarium* DSM 8532 Zellüberstand, welcher für sein gutes hemicellulolytisches Potential bekannt ist, zum Vergleich heran gezogen (Zhang et al. 2014). Besonders auf den Xylanen ist die spezifische Enzymaktivität von *Herbinix hemicellulosilytica* (T3/55) höher als in den Vergleichsproben von

Clostridium stercorarium (Abbildung 40). *Eine* zweite Messung auf unlöslichem und viskosem Arabinoxylan zeigte, dass T3/55 auch auf diesen Substraten eine höhere Aktivität aufweist.

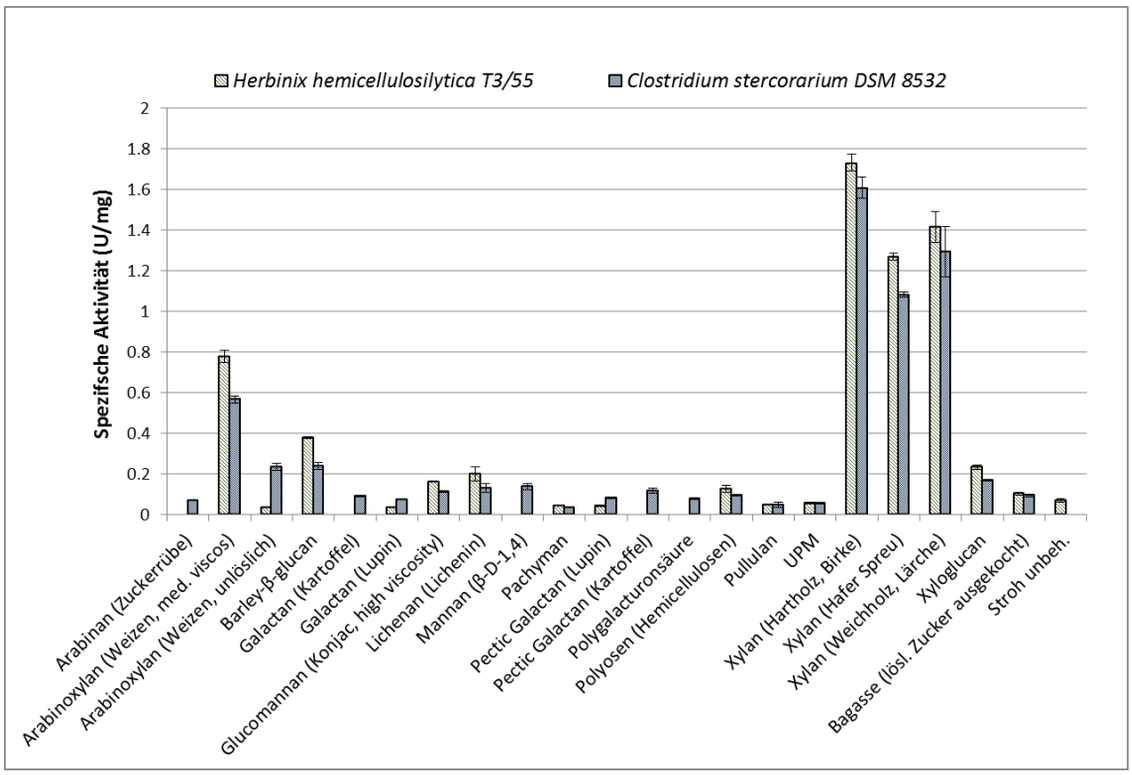


Abbildung 40: Spezifische Aktivität in Units pro Milligramm Protein im Kulturüberstand von C. stercorarium (DSM 8532) und Herbinix hemicellulosilytica (T3/55) Zellüberstand auf diversen hemicellulosehaltigen Substraten

Von den Substraten, auf denen mit DNSA-Tests Enzymaktivität des T3/55 Zellüberstands messbar war, wurden zusätzlich Dünnschichtchromatographien zur Analyse des Produktspektrums angefertigt. Auch hier war der höchste Abbaugrad bei den Xylanen zu sehen (Abbildung 41). Alle Xylane konnten von den Enzymen des Zellüberstandes bis zu Xylose abgebaut werden (X1).

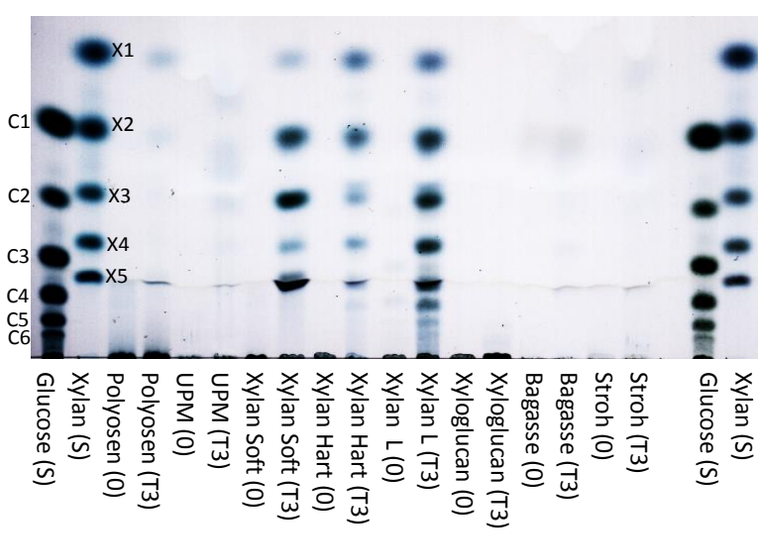


Abbildung 41: Dünnschichtchromatographie von T3/55 Zellüberstand auf ausgewählten Hemicellulosen. T3 = T3/55, 0 = negativ Kontrolle, S = Standard

Bei Xyloglucan, Stroh und Bagasse waren im Gegensatz zu den DNSA Tests keine Abbauprodukte erkennbar. Zusätzlich wurde der Abbau von natürlichem Substrat (Maissilage) getestet. Hierbei wurde der prozentuale Abbau von *C. stercorarium* DSM 8532 (Cst), *C. thermocellum* ATCC 27405 (Cth), den jeweiligen Mischkulturen T3/55-Cst, T3/55-Cth und *Herbinix hemicellulosilytica* T3/55 einzeln getestet. Den höchsten Abbaugrad erreichte mit fast 70 % *C. thermocellum*, gefolgt von T3/55 mit 60 % und *C. stercorarium* mit nur 35 %. Es konnte kein synergistischer Abbau der beiden Mischkulturen festgestellt werden (Abbildung 42).

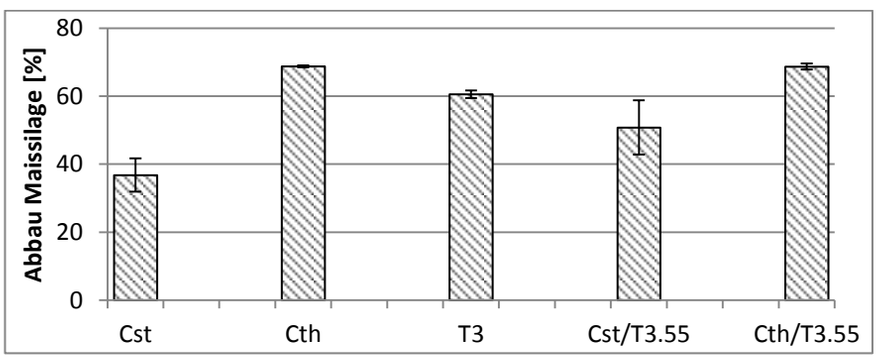


Abbildung 42: Messung des Abbaugrads von gewaschener Maissilage (~0,5 g in GS2 Medium); Cst = Clostridium stercorarium; Cth = Clostridium thermocellum; T3.55 = Herbinix hemicellulosilytica

Des Weiteren wurde als wichtiges Unterscheidungsmerkmal zwischen eng verwandten Spezies das zelluläre Fettsäuremuster bestimmt. Die häufigsten Fettsäuren in T3/55 Zellen waren 16:0 DMA (Dimethylacetal), 16:0 FAME (Fettsäuremethylester) und 16:0 ALDE (Aldehyd) (Tabelle 52). Mit dem Enterotube II lieferten folgende Tests ein negatives Ergebniss für T3/55: Sulfat Reduktion, Indolbildung, Lactosefermentation, Dulcitolfermentation, Desaminierung von Phenylalanin und Citratabbau. Harnstoff konnte von T3/55 verstoffwechselt werden.

*Tabelle 52: Zelluläres Fettsäuremuster (> 1 %) von Herbinix hemicellulosilytica T3/55. *Gruppe 1 enthielt 13:1 CIS 12 FAME und 14:0 ALDE; Gruppe 4 enthielt eine unbekannte Fettsäure und 15:2 FAME; Gruppe 6 enthielt 15:0 anteiso 3OH FAME und 16:1 CIS 7 DMA; Gruppe 8 enthielt 17:1 CIS 9 FAME und 17:2 FAME; Gruppe 10 enthielt 18:1c11/t9/t6 FAME eine unbekannte Fettsäure.*

Fettsäure	[%]
14:0 FAME	1.79
14:0 DMA	2.82
16:0 ALDE	11.48
16:0 FAME	37.51
16.1 CIS 9 DMA	1.27
16:0 DMA	21.56
18:0 FAME	1.73
18.1 CIS 9 DMA	1.27
18.1 CIS 11 DMA	5.77
Gruppe 1*	1.67
Gruppe 4*	1.67
Gruppe 6*	1.76
Gruppe 8*	2.91
Gruppe 10*	2.50

Außerdem wurden elektronenmikroskopische Aufnahmen von Zellen des neuen Isolats angefertigt. T3/55 bildet regelmäßige Stäbchen mit abgerundeten Enden, die von feinen geißelartigen Strukturen umgeben sind (Abbildung 43a). Bemerkenswert ist, dass die Zellstruktur sich nach drei Kultivierungstagen anfängt aufzulösen und die Zellen nach fünf Tagen vollständig kollabiert sind (Abbildung 43b, c). Dieser Umstand und die Tatsache, dass die Kultur keine Sporen bildet, erschwert die Kultivierung von T3/55. Bei Isolat SD1D konnte unter dem Lichtmikroskop ähnliches beobachtet werden.

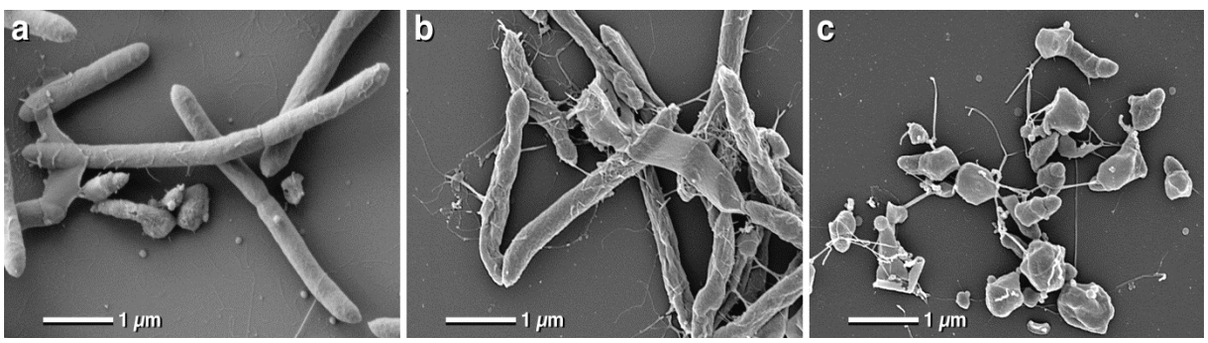


Abbildung 43: Elektronenmikroskopische Aufnahmen von T3/55 Zellen nach einem (a), drei (b) und fünf (c) Tagen Kultivierung in GS2 Medium bei 55 °C (Koeck et al. 2015b)

3.2.3.2 Genomsequenzierung ausgewählter thermophiler Isolate

Vor allem um Erkenntnisse zur Ausstattung mit Glykosidhydrolasen für den Abbau von Biomasse und über die Beschaffenheit und Art des cellulolytischen Systems (falls vorhanden) zu erhalten wurden von ausgewählten thermophilen Isolaten die Genome sequenziert. Die Genominformationen liefern in Kombination mit Metagenomdaten zudem wichtige Einblicke in die Rolle und Abundanz eines Isolats innerhalb der bakteriellen Gemeinschaft in der Biogasanlage. In Tabelle 53 sind zunächst die grundlegenden Eigenschaften der sequenzierten Genome zusammengefasst.

Tabelle 53: Eigenschaften der von den Isolaten 1A1, T3/55, SD1D und DG5 sequenzierten Genome

Eigenschaft	1A1 (Koeck et al. 2014c)	T3/55 (Koeck et al. 2015c)	SD1D	DG5 (Koeck et al. 2014b)
Größe [bp]	3.708.331	3.037.031	2.567.185	2.229.578
GC Gehalt [%]	37,28	36,69	35,13	44,15
Gesamtzahl Gene	3536	2681	2340	2088
rRNA Operons	10	4	4	4
tRNAs	59	35	49	59
Protein kodierende Sequenzen (cds)	3472	-	-	2017
Gene mit einer bekannten Funktion	2112	1187	-	1069

Die Genomsequenz von *R. cellulosi* DG5 wurde auf dem Illumina MiSeq System angefertigt. Nach einer automatischen Assemblierung der Sequenzen mit der GS *De Novo* Assembler Software

(Roche, Version 2.8) wurden verbleibende Lücken im Genom am CeBiTec (Bielefeld) *in silico* geschlossen (Kuberl et al. 2011, Maus et al. 2013, Wibberg et al. 2011). Dies resultierte in einem lückenlos geschlossenen *C. cellulosi DG5* Chromosom mit 2.229.578 bp und mit einem G+C Gehalt von 44,15 %. Die Genomannotation erfolgte mit der Annotationsplattform GenDB (Meyer et al. 2003). Wie man dem zirkulären Plot entnehmen kann, konnte das Genom zwar geschlossen werden, jedoch sind noch nicht alle Gene vollständig annotiert (Abbildung 44).

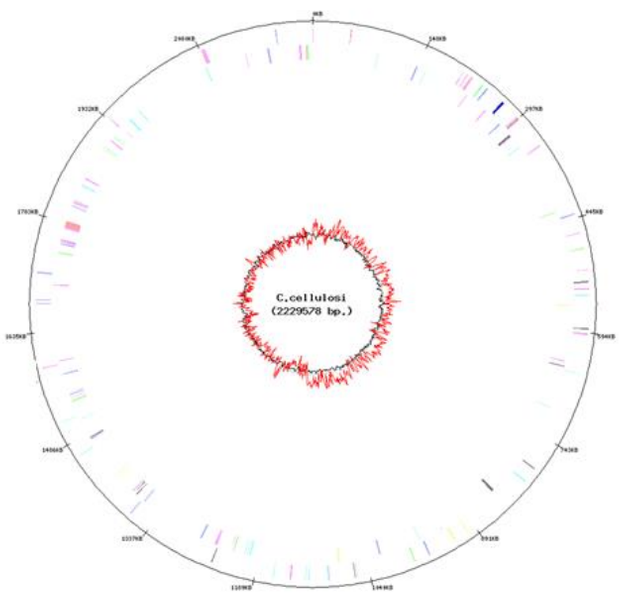


Abbildung 44 Zirkulärer Plot des C. cellulosi DG5 Genoms, von außen nach innen: annotierte Gene des Leitstrangs, Gene des Gegenstrangs, GC Gehalt (schwarz), GC Verhältnis (G - C)/(G + C) (rot)

Die drei anderen DNA Proben wurden auf gleiche Weise sequenziert und die Genome assembliert. Jedoch konnten nicht alle Lücken geschlossen werden, so dass die Genomsequenzen nur als „Draft"-Sequenz zu Verfügung standen. Das sogenannte „Fragment recruitment" wurde in Kooperation mit dem CeBiTec angewandt, um die Einzelsequenzen der Metagenomanalyse der Biogasanlage H2 (Viersen) mit den Referenzgenomen der Isolate zu vergleichen (= „Readmapping") (Rusch et al. 2007). Tabelle 54 zeigt den Prozentsatz der 520.000 Metagenomeinzelsequenzen der Biogasanlage Viersen mit einem der sequenzierten Genome übereinstimmen (Einstellungen: 97 % Übereinstimmung, mindestens 40 bp Überlappung).

Tabelle 54: „Readmapping“ der sequenzierten Genome auf den Metagenomdatensatz der Biogasanlage (BGA) Viersen (97% Sequenzübereinstimmung bei mindestens 40 Basenpaaren Überlappung)

Isolat	BGA Viersen
1A1 *(Bacillus thermoamylovorans)*	0,29 %
BC1 *(Clostridium thermocellum)*	0,80 %
T3/55 *(Herbinix hemicellulosilytica)*	0,52 %
DG5 *(Costridium cellulosi)*	0,14 %
SD1D (Herbinix sp.)	1,0 %

Auf die Genome von *Herbinix sp.* SD1D und *C. thermocellum* BC1 konnten mit 1 bzw. 0,8 % die meisten der Metagenomsequenzen im Vergleich zu den andern Isolaten aligniert werden. Sequenzen die auf das Genom von *C. cellulosi* „gemappt“ werden können, machen dagegen nur 0,14 % des Datensatzes aus. Dies ist ein Indiz dafür, dass *C. cellulosi* in der Biogasanlage H1 eher in geringeren Abundanzen vorkommt und im Gegensatz zu *C. thermocellum* kein Schlüsselorganismus für den Abbau von lignocellulosehaltiger Biomasse ist.

Die weitere Analyse und Interpretation der Genomdaten erfolgte mittels GenDB und mit Hilfe der CAZy Datenbank (Lombard et al. 2014) und dem „Carbohydrate-active-enzyme Annotation“ Web-server dbCAN (Yin et al. 2012). Die Analyse zeigte, dass das *C. cellulosi DG5* Genom insgesamt für 136 auf Kohlenhydraten aktive Enzyme codiert (Tabelle 55). Diese gehören in erster Linie zu verschiedenen Familien von Glycosidhydrolasen, welche an der Hydrolyse und/oder Umordnung von glycosidischen Bindungen beteiligt sind. Darunter wurden alle Gene die für den Abbau von kristalliner Cellulose nötig sind, identifiziert: eine prozessive Cellobiohydrolase der Glycosidhydrolase Familie 48, zwei Cellobiohydrolasen der Familie 5 und zwei Endo-Glucanasen der Familie 9. Unter den 107 auf Kohlenhydraten aktiven Enzymen im Genom von *Bacillus thermoamylovorans* 1A1 konnte dagegen nur eine Endo-Glucanase der Familie 9 annotiert werden. Die für den Abbau von kristalliner Cellulose notwendige Exo-Glucanase fehlt. Dafür wurden zwei α-Glucosidasen der Familie 4 identifiziert, die in keinem der anderen drei Genome zu finden sind und vermutlich zum Abbau von Stärke beitragen. Beide Genome enthalten keine cellulosomalen Komponenten, d.h. den Cohesinen oder Dockerinen homologe Sequenzen.

Das Genom von Isolat T3/55 enthält 155 auf Kohlenhydraten aktive Enzyme. Darunter sind 26 CBMs, 33 Esterasen, 50 Hydrolasen und 39 Transferasen aber ebenfalls keine cellulosomalen Komponenten. Wiederum wurden alle Gene für den Abbau von kristalliner Cellulose identifiziert: Eine prozessive Cellobiohydrolase der Glycosidhydrolase Familie 48, und eine Endo-Glucanasen der Familie 9.

Das Genom von Isolat SD1D codiert mit 154 Komponenten für eine ähnliche Anzahl an „CAZymes“, wie das Genom von Isolat T3/55. Darunter wurden auch alle Gene die für den Abbau von kristalliner Cellulose nötig sind, identifiziert. Obwohl beide sehr nah verwandt sind, zeigte ein genauer Vergleich der cellulolytischen Systeme deutliche Unterschiede auf. Im Gegensatz zu Isolat T3/55, welches jeweils nur eine Glycosidhydrolase der Familien 9 und 5 besitzt, konnten im Genom von SD1D fünf potentielle Endo-Glucanasen der Familie 9 und fünf potentielle Cellulasen der Familie 5 identifiziert werden. Zusätzlich wurden noch zwei Exo-Glucanasen der Familie 48 gefunden. Außerdem wurden zu Cohesin Typ 1 und Dockerin Typ 1 homologen Sequenzen identifiziert. Dies spricht für cellulosomale Komponenten. Eine genauere Analyse der entsprechenden Leserahmen offenbarte fünf potentielle Cellulosomale Gene (Abbildung 45).

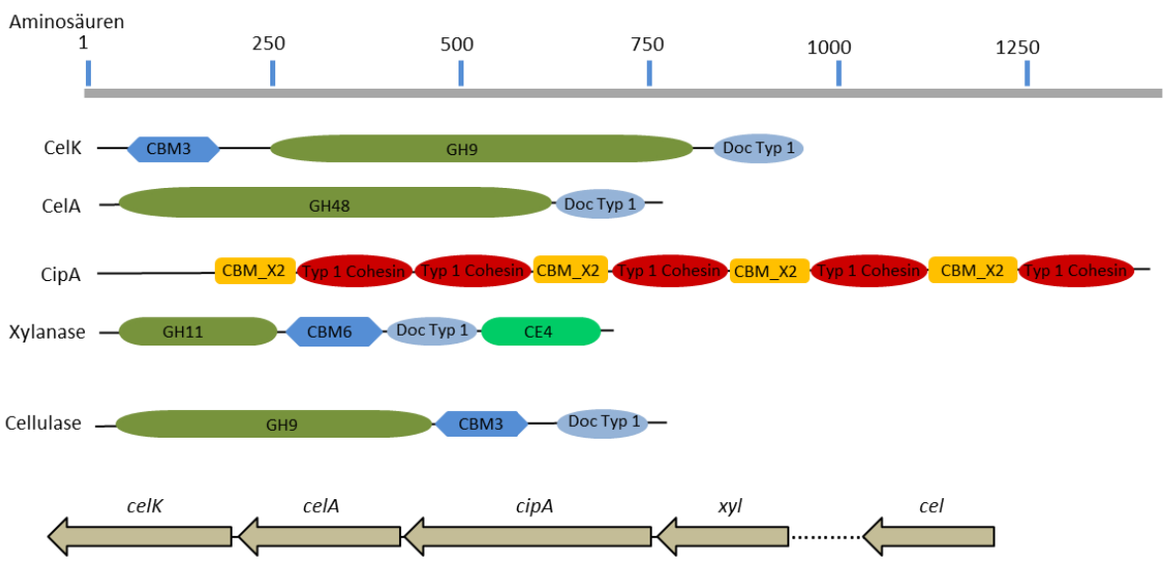


Abbildung 45: Vermeintliche Cellulosomale Komponenten von Isolat SD1D und deren modularer Aufbau, abgeleitet aus der Genomsequenz von SD1D. Doc = Dockerin, cel = Gene die für Cellulasen codieren, xyl = Gene die für Xylanasen codieren.

Darunter befindet sich ein makromolekulares Scaffoldinprotein mit fünf Typ 1 Cohesinen, zwei Endo-Glucanasen der Familie 9, eine Exo-Glucanasen der Familie 48 und eine Xylanase der Familie 11. Bis auf eine Endo-Glucanase sind alle Gene in einem Cluster angeordnet. Eine Analyse des Kulturüberstands mittels SDS-PAGE und Western Blot mit Anti-Cohesin Antikörper, bestätigte die Expression des cellulosomalen Komplexes (Daten nicht gezeigt). Die restlichen Cellulasen von Isolat SD1D besitzen kein Dockerin Modul und sind somit nicht in den cellulosomalen Multienzymkomplex eigebunden. Vergleicht man nochmal die spezifischen Aktivitäten von T3/55 und SD1D Kulturüberstand in Tabelle 51, wird ersichtlich, dass sowohl die Enzymkomplexierung in einem Cellulosom als auch höhere Anzahl an entsprechenden Enzymen zu einer signifikant besseren Abbaueffizienz von Xylan und Avicel bei trägt.

Tabelle 55: Kohlenhydrat aktive Enzyme die in den Genomen von Clostridium cellulosi DG5, Bacillus thermoamlylovorans 1A, Herbinix hemicellulosilytica T3/55 und Herbinix sp. SD1D codiert sind; CBM: Kohlenhydratbindemodule, CE: Carbohydratesterasen, GH: Glycosidhydrolasen, GT: Glycosyltransferasen. Potentielle Cellulasen sind grün hinterlegt

		DG5	1A1	T3/55	SD1D
Alle CAZy Treffer		**136**	**107**	**155**	**154**
CBM	**Summe**	**22**	**24**	**26**	**59**
	CBM2	2	0	0	5
	CBM3	7	0	3	8
	CBM4	0	0	0	2
	CBM5	0	0	0	2
	CBM6	1	0	2	8
	CBM9	0	0	0	2
	CBM17	1	0	0	0
	CBM20	1	0	0	0
	CBM22	1	0	1	2
	CBM28	1	0	0	0
	CBM29	0	0	1	0
	CBM30	0	0	0	1
	CBM32	0	0	0	2
	CBM34	1	1	0	0
	CBM35	0	0	0	1
	CBM36	0	0	1	2
	CBM41	0	1	0	0
	CBM46	0	0	3	7
	CBM48	0	2	1	0
	CBM50	7	18	14	15
	CBM61	0	1	0	0
	CBM65	0	1	0	2
CE	**Summe**	**6**	**18**	**33**	**20**
	CE1	1	6	19	6
	CE3	0	0	3	2
	CE4	3	4	5	8
	CE6	0	0	1	1
	CE7	1	2	1	0
	CE9	1	2	1	1
	CE10	0	2	2	1
	CH12	0	1	1	0
	CH14	0	1	0	1
GH	**Summe**	**46**	**32**	**50**	**52**
	GH1	0	4	2	0
	GH2	1	1	1	1
	GH3	3	2	4	3
	GH4	0	2	0	1
	GH5	2	0	1	5
	GH8	1	1	1	1
	GH9	2	1	1	5
	GH10	2	0	4	5
	GH11	0	0	2	3
	GH13	3	6	5	2
	GH15	1	0	0	0
	GH16	0	0	1	0
	GH18	1	2	5	5
	GH23	1	1	1	1
	GH24	1	0	0	0
	GH25	2	0	1	1
	GH28	0	0	0	1
	GH30	0	0	0	0
	GH31	1	0	1	0
	GH36	0	0	1	0
	GH39	2	0	1	0
	GH42	1	1	0	0
	GH43	2	1	3	4

	GH48	1	0	1	2
	GH51	4	0	2	1
	GH53	0	1	0	0
	GH62	0	0	0	1
	GH63	1	0	0	0
	GH65	0	2	0	0
	GH67	1	0	1	1
	GH73	2	1	0	0
	GH74	0	1	0	0
	GH76	0	1	0	0
	GH77	1	0	1	1
	GH94	4	0	5	3
	GH109	0	5	5	3
	GH113	2	0	0	0
	GH115	1	0	0	0
	GH120	1	0	1	1
	GH127	0	0	1	0
	GH129	0	0	0	1
	GH130	3	0	0	0
GT	**Summe**	**27**	**21**	**39**	**23**
	GT1	0	0	0	0
	GT2	9	10	21	11
	GT4	8	4	9	6
	GT8	0	1	1	1
	GT15	0	0	0	1
	GT19	0	2	0	0
	GT23	2	0	0	0
	GT26	1	0	0	0
	GT28	3	1	2	1
	GT32	0	0	1	0
	GT35	1	0	1	1
	GT44	0	0	1	0
	GT51	2	3	2	2
	GT84	1	0	1	0

3.2.3.3 Charakterisierung des cellulolytischen Systems von Isolat T3/55

Da die Isolate T3/55 und SD1D eine neue Gattung bisher nicht bekannter cellulolytischer Bakterien darstellen, wurde exemplarisch für diese Gruppe das cellulolytische System von *Herbinix hemicellulosilytica* T3/55 genauer charakterisiert. Wie im vorangehenden Kapitel beschrieben konnten drei potentielle Cellulasen identifiziert werden, die der GH Familie 5, 9 und 48 zuzuordnen sind.

GH Familie	Max. Sequenzidentität (NCBI)	Leserahmen
GH9	Endo-1,4-beta-glucanase (EC 3.2.1.4)	TUM355_rep1_2138
GH48	1,4-beta-cellobiohydrolase (EC 3.2.1.91)	TUM355_rep1_2139
GH5	Endo-1,4-beta-glucanase (EC 3.2.1.4)	TUM355_rep1_438

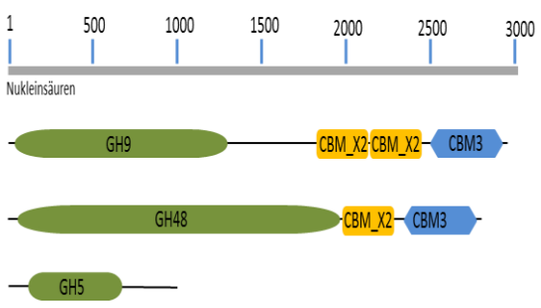


Abbildung 46: Potentielle Cellulasen im T3/55 Genom. Links: Protein-BLAST der Aminosäuresequenz und Lokalisation des Gens im Genom von T3/55. Rechts: Vorhergesagte Domänenstruktur der potentiellen Cellulasen.

In Abbildung 46 ist die Modulstruktur der T3/55-Cellulasen dargestellt. Die GH9- und GH48-Module besitzen neben dem katalytischen Glycosidhydrolase-Modul noch Cellulosebindemodule der Gruppe X2 und 3, welche dafür verantwortlich sind das Substrat zu binden und in die Nähe des Enzyms zu bringen. Die GH5 wurden ebenso wie die GH9 als Endo-Glucanase klassifiziert, besitzt jedoch kein zusätzliches Bindemodul. Die potentiellen Cellulasen wurden in einen Expressionsvektor (pET24) kloniert und in den Expressionsstamm BL21Star transformiert. Nach erfolgreicher Expression und Aufreinigung der Proteine wurde die Aktivität mit Aktivitätsassays auf diversen Substraten getestet und mittels kolorimetrischer Reaktion von DNSA mit den entstandenen reduzierenden Enden ausgewertet. Nach einer Stunde Inkubation auf den Substraten zeigten die GH5 und die GH9 die höchste Aktivität auf ß-Glucan, GH48 dagegen auf PASC. Die glycosidische Bindung zwischen den ß-D-Glucopyranose Einheiten von ß-Glucan, PASC und CMC konnte von allen drei Enzymen hydrolysiert werden. Nur das Enzym mit dem GH9 Modul war in der Lage, die kristalline Struktur von Avicel messbar aufzubrechen und geringe Mengen reduzierender Zucker freizusetzen. Jedoch konnte Avicel nach einer längeren Inkubationszeit (24 Stunden) von allen Enzymen angegriffen werden. Auf Xylan und Arabinoxylan war auch nach 24 Stunden nur die GH5 in der Lage, die β-1,4-glycosidischen Bindungen zwischen den β-D-Xylopyranose-Einheiten zu hydrolysieren (Abbildung 47).

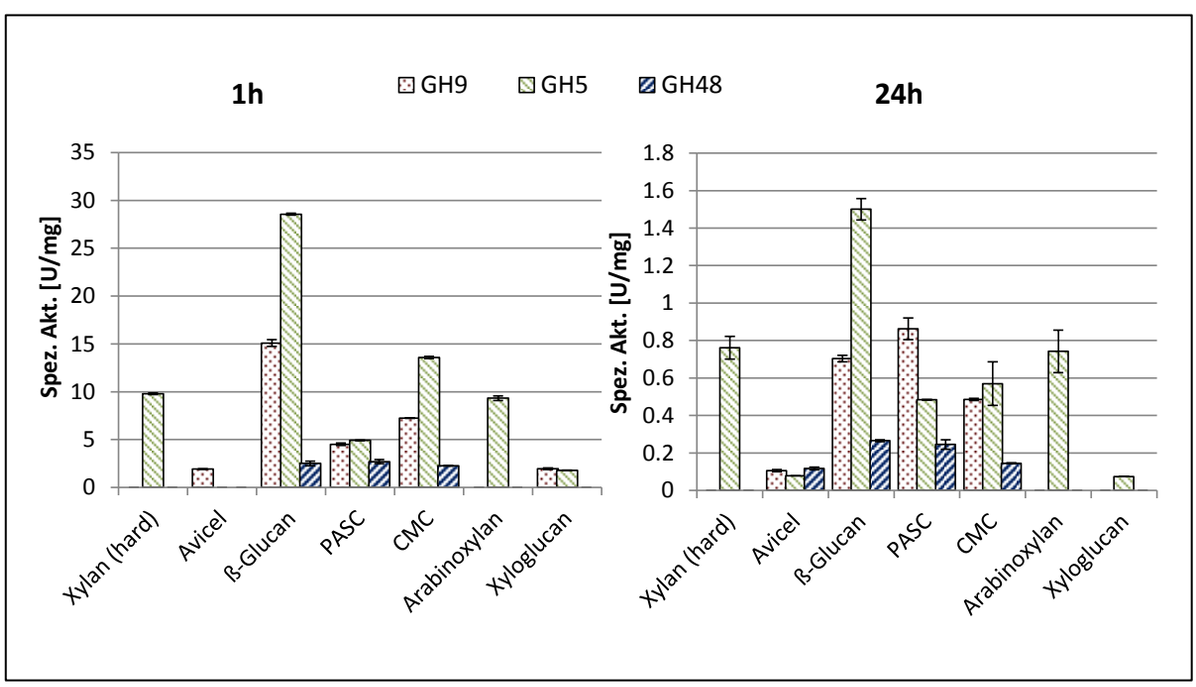


Abbildung 47: Spezifische Aktivität in Units pro Milligramm Enzym der T3/55 Cellulasen. Links nach einer Stunde Inkubation bei 55 °C (1h). Rechts nach 24 Stunden Inkubation bei 55 °C (24h).

Als nächstes wurden pH- und Temperaturoptimum der Cellulasen anhand der Aktivität nach 2h Inkubation auf PASC (GH48) und ß-Glucan (GH9, GH5) bei verschiedenen pH-Werten und Temperaturen bestimmt.

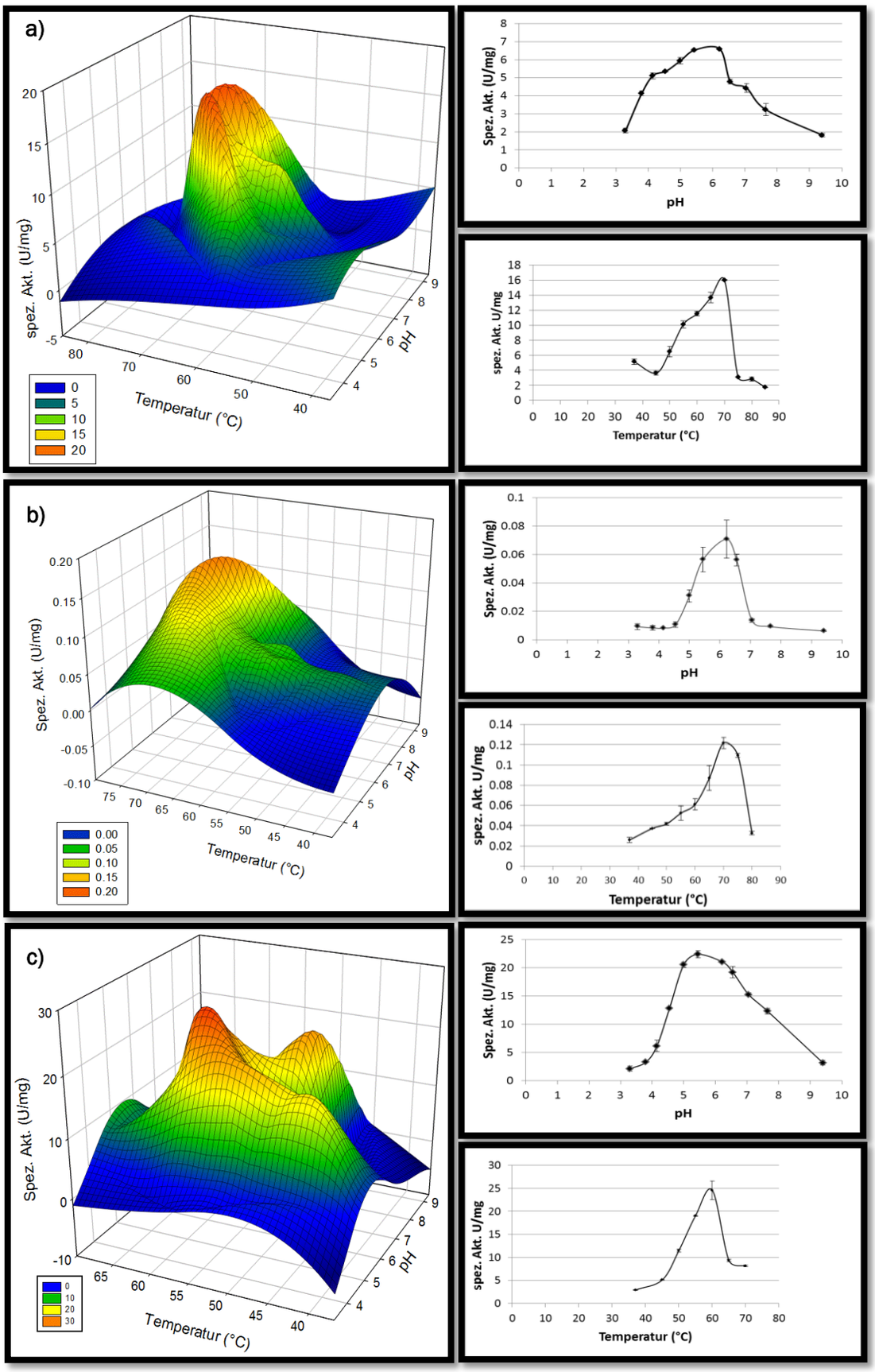


Abbildung 48: pH und Temperaturoptimum des GH9 (a), GH48 (b) und GH5 (c) Enzyms von Isolat T3/55 dargestellt als 3D Plot. Die Farbskale zeigt die spezifische Aktivität in U/mg. Die links dargestellten pH bzw. Temperaturkurven wurden bei der jeweils optimalen Temperatur bzw. dem optimalen pH-Wert ermittelt.

Das GH9 Enzym zeigte ein Optimum der Aktivität bei pH 6,2 und 70 °C; das GH48 Enzym bei pH 6,2 und 70 °C und schließlich das GH5 Enzym bei pH 5,5 und 60 °C. Das pH Optimum der Enzymaktivität liegt damit deutlich unter dem pH-Wert des Wachstumsoptimums (Tabelle 50). Dies gewährleistet eine stabile Hydrolyse auch in leicht saurem Milieu. Bemerkenswert ist, dass sowohl das pH- als auch das Temperaturoptimum der GH5 deutlich unter denen der beiden anderen Enzyme lag. Zur genaueren Untersuchung des katalytischen Typs wurden von den Enzymen Produktbildungskinetiken auf Cellobiose, Cellotriose, Cellotetraose, Cellopentaose und Cellohexaose angefertigt.

Entgegen dem Ergebnis des Aminosäuresequenzvergleichs mit der Datenbank, welcher auf eine Endo-Glucanase hinwies, verhielt sich die GH5 in den Abbautests wie eine Exo-Glucanase (Cellobiohydrolase), welche die ß-1,4-glycosidische Bindung nach jedem zweiten Glucose-Molekül spaltet (Abbildung 49a). Cellobiose wurde nicht weiter abgebaut. Die GH9 hat eine sogenannte prozessive Endo-Glucanase Aktivität, welche innerhalb des Moleküls hydrolysiert und dann prozessiv Cellobiose abspaltet (Abbildung 49b). Die minimale Substratgröße ist Cellotriose. Die GH48 ist wiederum eine Cellobiohydrolase mit sehr schwacher Aktivität, welche die ß-1,4-glycosidische Bindung immer nach zwei Glucose-Molekülen spaltet. Die minimale Substratgröße ist die Cellotetrose. Cellobiose und Cellotriose wurden nicht abgebaut (Abbildung 49c).

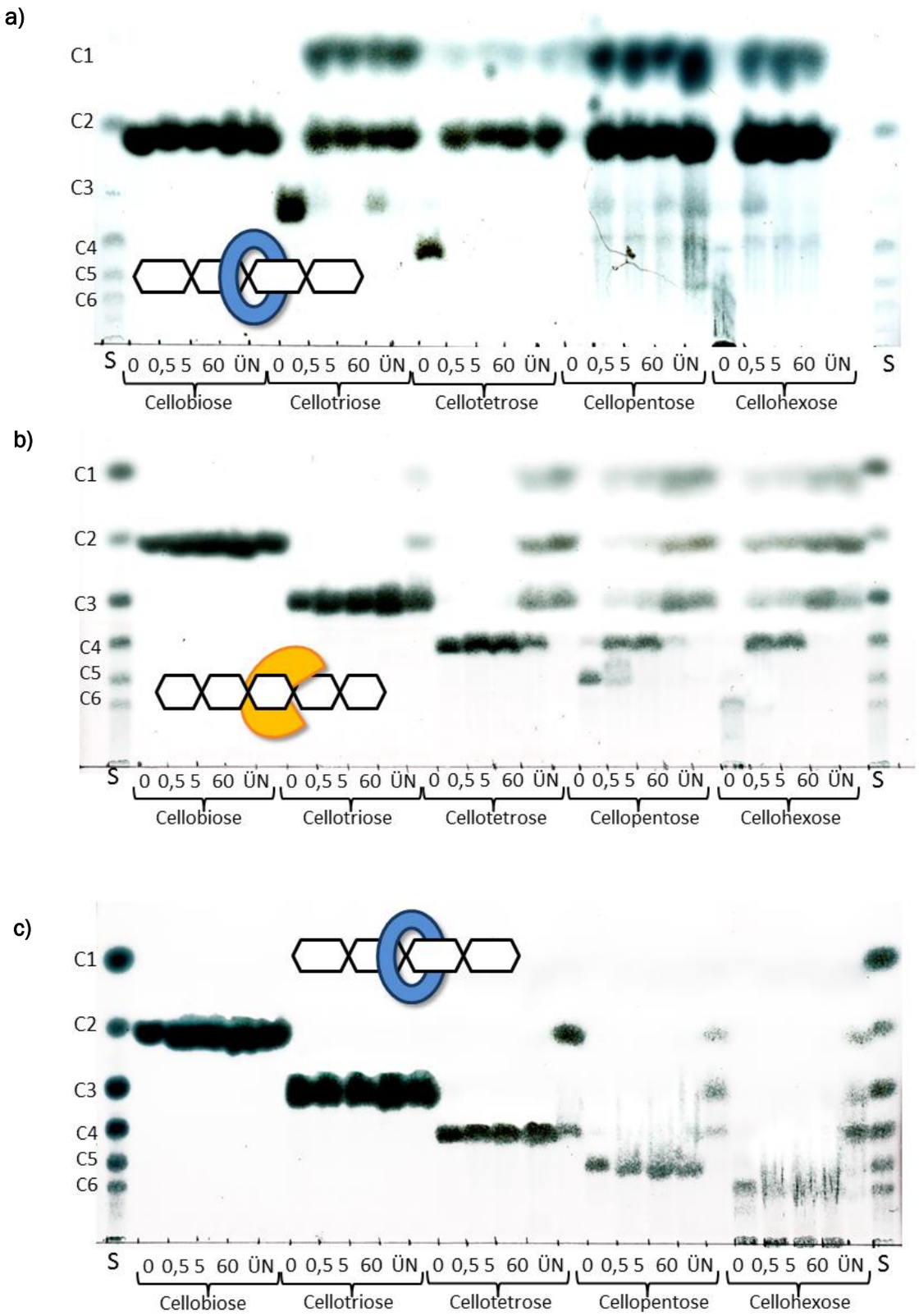


Abbildung 49: Dünnschichtchromatographien zur Darstellung der Enzymaktivität auf ß-1-4-glycosidisch verknüpften Oligosacchariden nach 0, 0,5, 5 und 60 Minuten sowie über Nacht (ÜN). a) Enzymkinetik T3/55 GH5; b) Enzymkinetik T3/55 GH9; c) Enzymkinetik T3/55 GH48

Um einen vermuteten Synergismus der Endo- und Exo-Glucanase (GH9 und GH48) von Isolat T3/55 beim Abbau von Cellulose aufzudecken, wurde die GH9 äquimolar mit der GH48 gemischt

und die Aktivitäten auf PASC mit den Aktivitäten der gleichen Konzentrationen der Einzelenzyme verglichen. Dabei wurde in einem Ansatz zusätzlich die ß-Glucosidase BglT aus *Thermus thermophilus* zugegeben, um die verbleibende Cellobiose vollständig zu Glucose abzubauen (Dion et al. 1999), da für Cellobiose in vielen Fällen eine Endprodukthemmung für Enzyme dieses Typs gezeigt worden ist.

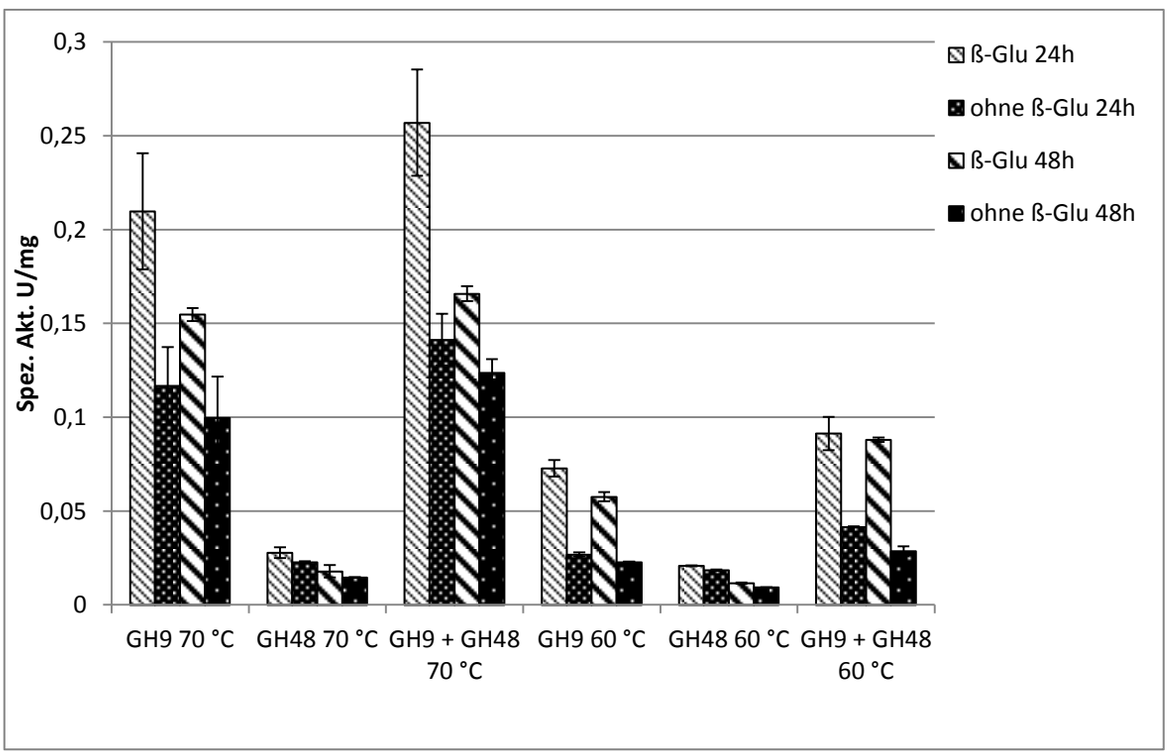


Abbildung 50: Spezifische Aktivitäten in Units pro Milligramm Enzym von GH9 und GH48 Enzymmischungen und Einzelenzymen aus Isolat T3/55 auf PASC mit und ohne ß-Glucosidase BglT (ß-Glu). Messung nach 24 bzw. 48 Stunden (h)

Die geringste hydrolytische Aktivität zeigte die GH48 bei 60 °C. Generell waren die Aktivitäten bei 60 °C geringer, da das Aktivitätsoptimum der Enzyme bei 70 °C liegt (Abbildung 50). Die höchste Aktivität wurde mit einer Kombination von GH9 und GH48 und Inkubation bei 70 °C erreicht. Die ß-Glucosidase hydrolysiert Cellobiose, das Endprodukt der Hydrolyse durch die Cellulasen, zu zwei Glucose-Einheiten, wodurch ein weiteres reduzierendes Ende entsteht. Dadurch konnte die spezifische Aktivität durch Zugabe von ß-Glucosidase noch einmal deutlich gesteigert werden.

3.2.3.4 Charakterisierung des hemicellulolytischen Systems von Isolat T3/55

Da der Zellüberstand von Isolat T3/55 besonders aktiv auf Xylanen war (Abbildung 40), wurde das Genom zusätzlich nach GH Modulen durchsucht, die potentielle Xylanase-Aktivität besitzen. Dabei wurden 12 mögliche Xylanasen identifiziert.

Tabelle 56: Potentielle Xylanasen im T3/55 Genom. GH Familie, Proteinblast (NCBI blastp) der Aminosäuresequenz und Lokalisation des Gens im Genom von T3/55.

Name	GH Familie	Max. Sequenzidentität (NCBI)	Lokalisation
X1	GH51	Alpha-L-Arabinofuranosidase (EC 3.2.1.55)	TUM355_rep1_1502
X2	GH39	Alpha-L-Arabinofuranosidase (EC 3.2.1.55)	TUM355_rep1_1501
X3	GH43	1,4-beta-Xylosidase (EC 3.2.1.37)	TUM355_rep1_1505
X4	GH51	Alpha-L-Arabinofuranosidase (EC 3.2.1.55)	TUM355_rep1_810
X5	GH11	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_1528
X6	GH10	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_1554
X7	GH43	Alpha-N-Arabinofuranosidase (EC 3.2.1.55)	TUM355_rep1_2049
X8	GH10	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_2050
X9	GH11	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_269
X10	GH10	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_698
X11	GH10	Endo-1,4-beta-Xylanase (EC 3.2.1.8)	TUM355_rep1_905
X12	GH43	1,4-beta-Xylosidase (EC 3.2.1.37)	TUM355_rep1_929

Analog zu den Cellulasen wurden die entsprechenden Leserahmen in *E. coli* kloniert und die Proteine rekombinant exprimiert und aufgereinigt. Von den gereinigten Proteinen wurde die Aktivität mit DNSA Tests auf diversen Substraten getestet.

Tabelle 57: DNSA Assay zur Ermittlung der spezifischen Aktivität der putativen Xylanasen von Isolat T3/55 auf diversen Substraten. Ausgedrückt als µmol Glucose Äquivalent pro mg Zellprotein und Minute (U/mg). Von jeder Proteinprobe wurden zwei Replikate gemessen. Auf den Substraten Arabinogalactan, Glucomannan, Galactomannan, Inulin, Laminarin, Lichenan, Pachyman, Pectin, Pullulan, Polygalacturonsäure, Xyloglucan war keines der Enzyme aktiv.

Substrat	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Arabinan (Zuckerrübe)	0.11	0	0	0.16	0	0	0	0	0	0	0	0
Arabinoxylan (medium viskos)	0	0	0	0	2.77	2.66	0	0.74	3.01	3.39	3.59	0
Arabinoxylan (unlöslich)	0	0	0	0	1.27	1.27	0	0.10	1.07	1.47	0.94	0
Mannan	0	0	0	0	0	0	0	0	0	0	0.13	0
Xylan (Birke, Hartholz)	0	0	0	0	3.31	3.05	0	0.10	2.32	3.41	3.12	0.26
Xylan (Haferspelzen)	0	0	0	0	3.51	5.18	0	0.20	1.68	3.26	3.19	0
Xylan (Lärche, Weichholz)	0	0	0	0	0	0.07	0	0	0	0	0.11	0

Die Enzyme X1 und X4 besitzen beide ein GH51 Modul und zeigen nur auf Arabinan Aktivität. Beide Enzyme wurden somit als alpha-L-Arabinofuranosidase annotiert. Diese hydrolysieren von den nicht reduzierenden Enden in Arabinan (1,3 und/oder 1,5 glycosidisch verknüpfte Arabinose) alpha-L-Arabinofuranosereste (Arabinose). Die beiden Arabinofuranosidasen sind allerdings nicht auf den Arabinanresten in Arabinogalactan oder den 2,3-verknüpften Arabinosen in Arabinoxylan aktiv. Die restlichen Enzyme besitzen ein katalytisches Modul der GH-Familie 10 oder 11 und wurden als 1,4–Xylanase annotiert. Diese Enzymklasse katalysiert die Hydrolyse von 1,4-ß-D-Xylanen durch die Spaltung der glycosidischen Bindung innerhalb des Moleküls. Wie aus Tabelle 57 ersichtlich ist, sind die Enzyme der beiden Enzymklassen aus Isolat T3/55 sowohl auf Arabinoxylanen als auch auf Xylanen aktiv. Die putativen Enzyme X2, X3 und X7 waren auf keinem der getesteten Substrate aktiv. Das katalytische Modul dieser Enzyme wurde entweder der GH43- oder der GH39-Familie zugeordnet. Einige Enzyme der Familie GH43 können sukzessiv D-Xylose Reste vom reduzierenden Ende der Xylane freisetzen oder spalten Oligosaccharide in Xylose. Von den drei Enzymen mit einem GH43 Modul zeigte allerdings nur X12 auf Birkenxylan geringe Aktivität.

3.2.4 Isolate aus der mesophilen Biogasanlage Pellmeyer (T3)

Als Modelbiogasanlage zur Isolierung möglichst vieler Arten/Stämme im mesophilen Bereich wurde der Pellmeyer Fermenter in Eggertshofen bei Freising (Deutschland) ausgewählt (vgl.2.3). In Tabelle 58 sind alle mit den verschiedenen Methoden isolierten mesophilen Bakterien aufgelistet. Insgesamt wurden 28 mesophile Reinisolate aus der Biogasanlage Pellmeyer mit den verschiedensten Isolierungsmethoden gewonnen. Die meisten Isolate konnten bereits bekannten Spezies zugeordnet werden (79 %). Die 3 (11 %) Isolate, welche in der Lage waren kristalline Cellulose abzubauen, gehören laut ihrer 16S rRNA Gensequenzähnlichkeit mit über 99 % zur gleichen Art. Mit 96,98 % 16S Sequenzähnlichkeit zu *Clostridium herbivorans* DSM 14428 sind sie der Gattung *Clostridium* zuzuordnen, gehören jedoch voraussichtlich zu keiner bisher bekannten Art. Ein Vertreter wurde für eine weitere Charakterisierung und Genomsequenzierung ausgewählt (in fett markiert). Auf diesen wird im Kapitel 3.2.5 näher eingegangen. Alle Isolate stammen aus dem Phylum Firmicutes und lassen sich wiederum drei Klassen (Clostridia, Bacilli, Negativicutes) zuordnen. Die 10 isolierten Hauptgattungen teilen sich überwiegend auf die Klassen der Clostridia und Bacilli. Insgesamt ließen sich davon 18 Isolate in die häufigste Ordnung Clostridiales, 3 Isolate in die Ordnung Lactobacillales und 5 Isolate in die Ordnung Bacillales einordnen. Nur das Isolat Iso2/2 37 aus der Familie Veillonellaceae gehört zur Klasse der Negativicutes und zur Ordnung Selenomonadales und wurde ebenfalls zur weiteren Charakterisierung ausgewählt.

Tabelle 58: Liste aller mesophilen Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten Sequenzähnlichkeit (laut NCBI Datenbankvergleich). [1]Inc. sed. = incertae sedis

Klasse	Ordnung	Familie	Art	Isolate	Max. Indent. (%)	Indent. Isolate (min. 98 %)	Cellulose-abbau
Clostridia	*Clostridiales*	*Clostridiaceae*	*Clostridium subterminale*	Iso2/37	99	1	-
			Clostridium saccharobutylicum	Iso11/1 37 °C	99	1	-
			Clostridium butyricum	Iso12/1; Iso12/2 37 °C	98	1	-
			Clostridium oryzae	Iso4/1 37 °C	99	1	-
			Proteiniclasticum ruminis	Iso4/37	98	1	-
		[1]*Inc. sed.*	*Sporanaerobacter acetigenes*	Iso2/45	98	1	-
		Ruminococcaceae	*Ruminococcus bromii*	Iso1/45	99	1	-
			Anaerofilum agile	37hoe/5	97	1	-
		Lachnospiraceae	*Clostridium celerecrescens*	Iso2/3; Iso1/1; Iso1/2; Iso8/1; Iso8/2 37 °C	99	5	-
			Clostridium amygdalinum	Iso4/2 37 °C	98	1	-
			Clostridium XlVa sp.	**37hoe/1; 37hoe/3; 37hoe/4**	97	3	**JA**
Bacilli	*Lactobacillales*	*Streptococcaceae*	*Streptococcus sp.*	Iso6/1 37 °C	96	1	-
		Lactobacillaceae	*Lactobacillus rhamnosus*	9/1 37 °C	99	1	-
		Carnobacteriaceae	*Trichococcus pasteurii*	Iso1/37	99	1	-
	Bacillales	*Paenibacillaceae*	*Paenibacillus barengoltzii*	Iso5/37, Iso5/45; Iso6/45; Iso4/45	99	4	-
		Bacillaceae	*Bacillus licheniformis*	37hoe/2	99	1	-
Negativicutes	*Selenomonadales*	*Veillonellaceae*	*Propionispora hippei*	**Iso2/2 37**	93	1	-
Summe						**28**	

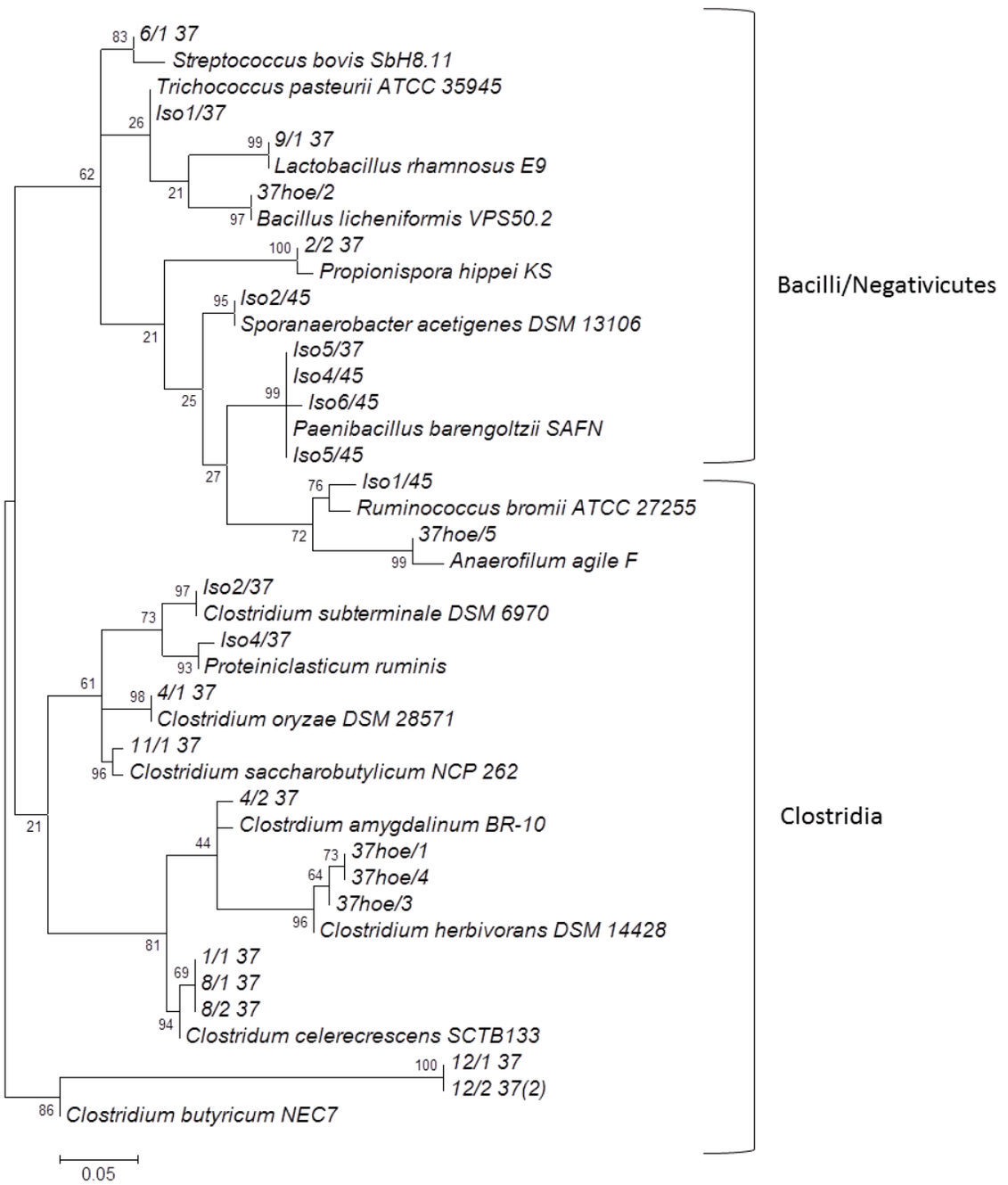


Abbildung 51: Ungewurzelter phylogenetischer Baum der 16S rRNA Gen-Sequenzen aller mesophilen Isolate und ausgewählter Referenzstämme unter Verwendung eines Maximum likelihood-basierten Algorithmus. Auf den einzelnen Ästen sind die einzelnen Bootstrap Werte angegeben.

Da der Hauptfokus in dieser Arbeit auf thermophilen cellulolytischen Isolaten lag, wurden wesentlich weniger mesophile (28) als thermophile Isolate (172) generiert. Dennoch ist verglichen mit den thermophilen Isolaten (Abbildung 34) die Diversität der mesophilen Isolate (Abbildung 51) höher. Der Shannon-Index (Hs) als Maß für die Diversität liegt bei 2,63 für die mesophilen Isolate und bei 2,48 für die thermophilen Isolate. Die Gattung *Ruminococcus* bildet im Phylogenetischen Baum ein eingenes Subcluster, abseits von anderen Vertretern der Klasse Clostridia.

3.2.5 *CHARAKTERISIERUNG AUSGEWÄHLTER MESOPHILER ISOLATE*

Insgesamt wurden zwei interessant erscheinende mesophile Isolate zur Genomsequenzierung und weiteren Charakterisierung ausgewählt, die in Tabelle 59 aufgeführt sind. Auswahlkriterien waren vor allem besondere Eigenschaften (z.B. sehr gute Cellulose-Hydrolyse) und Häufigkeit des Isolates. Beide Isolate stammen aus der mesophilen Biogasanlage T3 (Pellmayer). Isolat Iso2/2 37 wurde mit Strategie 1 aus einer Anreicherung in Mineralmedium mit 0,5 % (v/v) Arabinogalactan aus Lärche isoliert. Das Isolat 37hoe/3 wurden stellvertretend für alle 37hoe Isolate ausgewählt, welche mit Strategie 2 aus einer Verdünnungsreihe in flüssigem GS2 Medium mit 0,1 % (w/v) Hefeextrakt, 0,5 % (v/v) steril filtriertem Rumensaft und Filterpapier gewonnen und mit Agarplatten des selben Mediums mit MN301 Topagar gereinigt wurden.

Tabelle 59: Liste der zur genaueren Charakterisierung ausgewählten mesophile Isolate. BGA = Biogasanlage

Name Isolat	Max. Sequenzident. (NCBI)	Ursprung	EMBL/NCBI Zugangsnummern	DSMZ Nummer
37hoe/3	*Clostridium herbivorans (96,98 %)*	BGA T3 (Pellmeyer)	16S rRNA: LN868251	100335
Iso2/2 37	*Propionispora hippei (93 %)*	BGA T3 (Pellmeyer)	16S rRNA: LN868253	100628

3.2.5.1 Physiologische Charakterisierung ausgewählter mesophiler Isolate

Zur physiologischen Charakterisierung der mesophilen Isolate wurde zunächst Substratspektrum, Produktspektrum, pH- und Temperaturoptimum, sowie weitere wichtige morphologische und physiologische Eigenschaften bestimmt. Die Ergebnisse sind in Tabelle 60 zusammengefasst.

Tabelle 60: Wichtige morphologische und physiologische Eigenschaften von den Isolaten 37hoe/3, Iso2/2 37.

Characteristika	37hoe/3	Iso2/2 37
Temperatur (°C)	25 - 40	25 - 40
pH	6,5 – 7,5	5,5 – 8,2
Gram Reaktion	+	-
Zellmorphologie	Stäbchen	Stäbchen
Endosporen	+	+
Abbau von:		
Galactose	–	–
Arabinose	–	+
Xylan	+	+
Xylose	–	+
Cellulose*	+	–
Mannose	+	+
Fructose	+	+
Sucrose	+	–
Glucose	–	+
Cellobiose	+	+
Sorbitol	–	n.b.
Stärke	–	+
Haupt-Fermentationsprodukte	H_2, Butanol, Ethanol, Acetat, Butyrat	H_2, Propionat, Ethanol, Acetat, Butyrat

**Filterpapier (Whatman No. 1); n.b. = nicht bestimmt*

Das Isolat 37hoe/3 zeigte mit 96,98 % die höchste 16S rRNA Sequenzübereinstimmung mit dem *Clostridium herbivorans* Typstamm DSM 14428 und gehört somit voraussichtlich zu einer bisher nicht beschriebenen Art innerhalb der Gattung *Clostridium* und der Familie *Lachnospiraceae*. Da die Gattung *Clostridium* eine sehr große phänotypische und genotypische Diversität aufweist, wurde das neue Isolat nach der Einordnung von Collins et al. (1994) basierend auf den 16S rRNA Sequenzen dem *Clostridium* Cluster XIVa zugeordnet. Yutin und Galperin (2013) schlugen eine Umbenennung aller *Clostridium* spp. aus dem Cluster XIVa in das neue Genus *Lachnoclostridium* vor (Yutin and Galperin 2013). Diese Bezeichnung wurde allerdings in die LPSN (List of Prokaryotic names with Standing in Nomenclature) noch nicht aufgenommen und deshalb in dieser Arbeit noch nicht verwendet. Der nächste Verwandte *Clostridium herbivorans* hat sein Wachstumsoptimum ebenfalls im mesophilen Bereich und ist in der Lage kristalline Cellulose abzubauen. Auch das Substratspektrum der beiden Arten ist ähnlich, *C. herbivorans* konnte jedoch im Gegensatz zu 37hoe/3 nicht auf Fructose und Xylan wachsen. Beide bilden Butyrat und Ethanol als Endprodukte der Fermentation, *C. herbivorans* bildet aber im Gegensatz zu 37hoe/3 weder Acetat noch Butanol. Ein weiteres gemeinsames Merkmal ist die Notwendigkeit von Rumensaft als Wachstumsfaktor im Medium. *C. herbivorans* wurde erstmals aus dem Gastrointestinal-Trakt von Schweinen isoliert (Varel et al. 1995), 37hoe/3 stammt dagegen aus einer mesophilen Biogasanlage. Zudem wurde *C. herbivorans* bisher in Metagenomanalysen in keiner industriellen Biogasanlage nachgewiesen (Krause et al. 2008, Schlüter et al. 2008, Wirth et al. 2012).

Um das cellulolytische Potential des neuen Isolates mit dem des *C. herbivorans* Typstammes zu vergleichen, wurde die enzymatische Aktivität des Zellüberstandes auf verschiedenen Substraten bestimmt.

Tabelle 61: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat 37hoe/3 und C. herbivorans DSM 14428 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten

Isolat	CMC	PASC
37hoe/3	0,08 +/- 0,003	0,02 +/- 0,0007
C. herbivorans	0,18 +/- 0,005	0,07 +/- 0,002

Die Enzymaktivität des Zellüberstands des Isolats 37hoe/3 ist sowohl auf PASC als auch auf CMC messbar, aber signifikant geringer als die des verwandten *Clostridium herbivorans* Typstammes. Auf Avicel konnte bei beiden Stämmen keine Enzymaktivität gemessen werden, was auf das Fehlen eines kompletten cellulolytischen Enzymsystems hindeutet und dem Bakterium eine unterstützende Rolle beim Abbau ß-1,4-glukosidisch verknüpfter Substanzen zuweist.

Das Isolat Iso2/2 37 bildet, basierend auf den Ergebnissen der phylogenetischen Analyse der 16S rRNA Gensequenzen, eine neue Gattung innerhalb der Familie *Veillonellaceae* mit 92,92 %

Sequenzübereinstimmung mit dem 16S rRNA Gen von *Propionispora hippei* DSM 15287. Die Familie *Veillonellaceae* gehört zur Klasse der Negativicutes und zur Ordnung der Selenomonadales. Die Ordnung Selenomonadales wurde erst kürzlich auf Basis von phylogenetischen Analysen beschrieben und befindet sich derzeit noch im Umbruch. Gemeinsames Merkmal aller Spezies ist die gramnegativ färbende Zellwand innerhalb eines Phylums (Firmicutes) von grampositiven Bakterien (Marchandin et al. 2010). Die phylogenetische Position der Gattung *Propionispora* etc. innerhalb der *Veillonellaceae* ist umstritten. Marchandin et al. (2010) schlagen eine Neueinordnung der Gattungen in die Familie *Selenomonadales incertae sedis* vor. Eine genaue taxonomische Eingruppierung von Isolat Iso2/2 37 ist daher schwierig. Die nächste verwandte beschriebene Spezies *Propionispora hippei* bildet ebenso wie das neue Isolat Propionsäure und Essigsäure als Hauptfermentationsprodukte. Beide bilden gekrümmte Stäbchen mit terminalen Endosporen, pH und Temperaturoptimum sind ebenfalls vergleichbar (Tabelle 60).

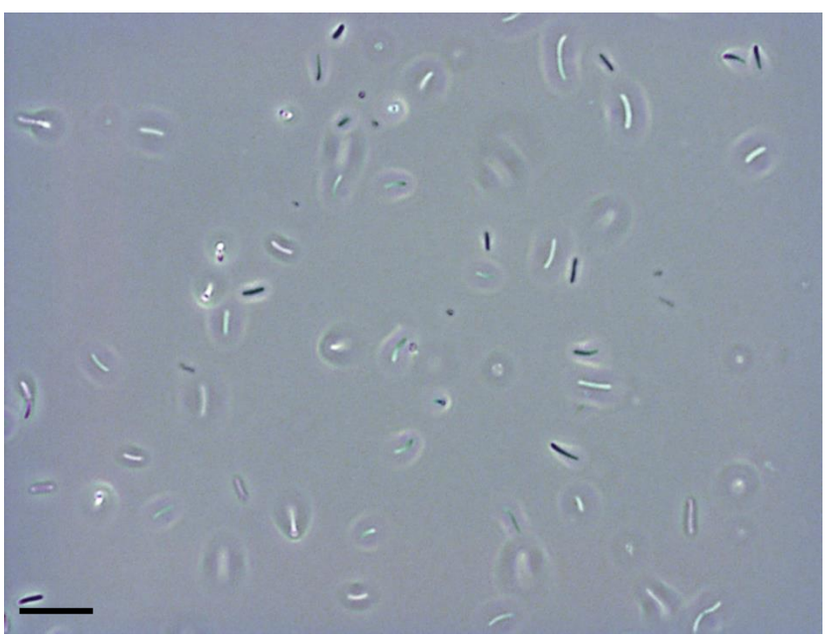


Abbildung 52: Zellen von Isolat Iso2/2 37 unter dem Phasenkontrastmikroskop (1000x Vergrößerung). Balken = 10µm

Propionispora hippei DSM 15287 wurde aus einer Anreicherung auf aliphatischen Polyestern isoliert (Abou-Zeid et al. 2004). Iso2/2 37 wurde dagegen als erster Vertreter der Ordnung Selenomonadales aus einer Biogasanlage isoliert.

Zur Bestimmung des enzymatischen Potentials wurde die spezifische Aktivität der Enzyme im Zellüberstand von Isolat Iso2/2 37 auf verschiedenen Substraten getestet.

Tabelle 62: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat Iso2/2 37 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten

Isolat	Xylan	ß-Glucan	CMC	PASC	Arabinogalactan
Iso2/2 37	0,014 +/- 0,0005	0,01 +/- 0,003	0,01 +/- 0,0005	0,01 +/- 0,0004	0,013 +/- 0,002

Wie erwartet waren die Enzyme des Zellüberstandes auf Arabinogalactan und Xylan aktiv, da auch das Isolat aus einer Anreicherung auf Arabinogalactan stammt. Zudem konnte geringe hydrolytische Aktivität auf CMC, PASC und ß-Glucan gemessen werden. Auf Avicel konnten die Enzyme keine quantifizierbaren Zucker frei setzen.
Neben dem Abbau von Hemicellulosen kann Isolat Iso2/2 37 auch eine Reihe von Mono- und Disacchariden verstoffwechseln und ist vermutlich als primärer Gärer in erster Linie an der Acidogenese im Biogasfermenter beteiligt.

3.2.5.2 Genomsequenzierung ausgewählter mesophiler Isolate

Vor allem um Erkenntnisse zur Ausstattung mit Glycosidhydrolasen für den Abbau von Biomasse und über die Beschaffenheit und Art des cellulolytischen Systems (falls vorhanden) zu erhalten, wurden von beiden mesophilen Isolaten die zur Charakterisierung ausgewählt wurden, die Genome zur Sequenzierung geschickt (Durchführung durch Dr. Andreas Schlüter, CeBiTec, Bielefeld). Die Genomsequenz von Isolat 37hoe/3 war bei Fertigstellung dieser Arbeit noch nicht verfügbar, daher sind in Tabelle 63 nur die grundlegenden Eigenschaften des sequenzierten Genoms von Isolat Iso2/2 37 zusammengefasst.

Tabelle 63: Eigenschaften des von Isolat Iso2/2 37 sequenzierten Genoms

Eigenschaft	Iso2/2 37
Größe [bp]	4.123.000
GC Gehalt [%]	45,59
Gesamtzahl Gene	3960
rRNA Operons	4
tRNAs	76

Die Genomsequenz wurde auf dem Illumina MiSeq System angefertigt und auf gleiche Weise wie bei den thermophilen Isolaten assembliert und annotiert. Da von der Biogasanlage T3 keine Metagenomanalyse durchgeführt wurde, wurden die Einzelsequenzen der Metagenomanalyse der Biogasanlage H2 (Viersen) mit dem Referenzgenom des mesophilen Isolats verglichen (Rusch et al. 2007). Obwohl das Isolat aus einer anderen mesophilen Biogasanlage stammt, konnten 0,51 % der 520.000 Metagenomsequenzen der Biogasanlage Viersen auf das Genom von Iso2/2 37 „gemappt“ werden (Einstellungen: 97 % Übereinstimmung, mindestens 40 bp Überlappung).

Tabelle 60 zeigt, dass die maximale Temperatur für das Wachstum mit 40 °C mehr als 10 °C unter der Betriebstemperatur des sequenzierten Reaktors liegt. Daher ist es unwahrscheinlich, dass die Sequenzen aus dem Metagenom zur selben Spezies gehören. Da als Schwellenwert nur 97 % Sequenzähnlichkeit gesetzt wurden, ist es wahrscheinlicher, dass die Sequenzen thermophilen Vertretern der gleichen Gattung zuzuordnen sind, welche wie Iso2/2 37 an der Acidogenese beteiligt sein könnten.

Die weitere Analyse und Interpretation der Genomdaten erfolgte mittels GenDB und mit Hilfe der CAZy Datenbank (Lombard et al. 2014) und dem Carbohydrate-active-enzyme Annotation Webserver dbCAN (Yin et al. 2012). Die Analyse zeigte, dass das Genom von Iso2/2 37 insgesamt für 136 auf Kohlenhydraten aktive Enzyme codiert (Tabelle 55). Diese gehören in erster Linie zu verschiedenen Familien der Glykosidhydrolasen, welche an der Hydrolyse und/oder Umordnung von glycosidischen Bindungen beteiligt sind. Darunter wurden keine Gene, die für den Abbau von kristalliner Cellulose nötig wären, identifiziert. Isolat Iso2/2 37 besitzt aber alle Enzyme die für den Abbau von Xylanen und Arabinogalactan nötig sind. Am häufigsten kommen Acetylxylan Esterasen der Familie 4 im Genom vor, welche die Deacetylierung von Xylanen und Xylo-Oligosacchariden katalysieren.

Tabelle 64: Kohlenhydrat aktive Enzyme die im Genom von Isolat Iso2/2 37 *codiert sind; CBM: Kohlenhydratbindemodule, CE: Kohlenhydratesterasen, GH: Glycosidhydrolasen, GT: Glycosidtransferasen*

CE	**Summe**	**25**	**CBM**	**Summe**	**13**
	CE1	1		CBM6	1
	CE3	3		CBM27	1
	CE4	12		CBM34	1
	CE6	2		CBM37	1
	CE7	2		CBM48	1
	CE9	2		CBM50	3
	CE10	1		CBM61	1
	CE11	2		CBM62	1
				CBM67	3
GH	**Summe**	**46**	**GT**	**Summe**	**54**
	GH2	1		GT2	13
	GH3	4		GT4	9
	GH4	2		GT5	1
	GH13	2		GT9	5
	GH18	4		GT17	2
	GH23	3		GT19	2
	GH25	2		GT26	1
	GH31	1		GT28	4
	GH43	1		GT30	1
	GH51	1		GT35	4
	GH53	1		GT51	3
	GH73	1		GT81	1
	GH77	1		GT83	5
	GH78	2			
	GH109	10			
	GH127	1			
	GH130	1			

3.3 „IN SACCO“ VERSUCHE

Für die Gewinnung von cellulolytischen Isolaten aus den „in sacco“ Anreicherungen wurden Säckchen nach 1 Tag (d) (1), 2d (2), 4d (3) und 7d (4) anaerob entnommen und mit steriler, 0,15 M Salzlösung gespült. Ein Teil des darin enthaltenen Filterpapiers wurde in alle Medien (vgl. 2.4.1) zur Gewinnung neuer Isolate inokuliert (Tabelle 65) und bei 55 °C inkubiert. Nach zwei Flüssig-Anreicherungen wurden Verdünngsausstriche auf den entsprechenden Festmedien angefertigt und mit PASC- und MN301-Topagar überschichtet (Isolierungsstrategie 1). In Basalsalz oder DNB Rumen (10 % v/v) fand kein Filterpapierabbau in Flüssigmedium statt.

Tabelle 65: Zusammenfassung und Ergebnis aller Medien die für die Gewinnung neuer Isolate aus den „in sacco“ Anreicherungen verwendet wurden. % Angaben in v/v. + Resultat positiv; - Resultat negativ

Medium	Filterabbau	Lysehöfe in Topagar	Isolate
Leitungswasser-R (40 % Rumen)	+	+	+
DNB Rumen (10 %)	-	-	-
DNB Gärrest (10%)	+	-	-
Komplex Rumen steril (1%)	+	+	-
Komplex GR steril (1%)	+	+	-
GS2 0,1 % Hefeextrakt	+	+	+
GS2	+	+	+
Basal Salz-Medium	-	-	-

Letztlich konnten nur von Leitungswasser-R und den beiden GS2 Medien cellulolytische Isolate mit hoher 16S rRNA Gen-Sequenzähnlichkeit (> 99 %) zu bereits bekannten Arten (*Clostridium cellulosi, Clostridium clariflavum*, *Acetivibrio cellulolyticus*) gewonnen werden (Tabelle 66). Für die weiteren Anreicherungen wurde GS2 Medium mit 10 % Gärrest aus dem Fermenter BM2T genutzt (GGR-Medium), um die Nährstoffbedingungen im Fermenter möglichst genau nachzustellen. Als Inokulum diente Filterpapier aus einer „in sacco“ Anreicherung, das an Tag 7 aus BM2T entnommen wurde. Die höchste Verdünnungsstufe mit Filterpapierabbau wurde in jedem Anreicherungsschritt erneut in demselben Medium verdünnt (Isolierungsstrategie 2). Schließlich wurden von der fünften Anreicherung die Verdünnungen 10^{-2} 10^{-3} 10^{-4} auf Platten desselben Mediums, sowie auf Platten mit Leitungswasser und 30 % Rumen mit MN301 Topagar ausgestrichen. Nach zwei Wochen Inkubation waren auf allen Platten Höfe im Topagar zu erkennen. Daraus konnten 11 Isolate mit weitgehend identischer 16S rRNA Gensequenz gewonnen werden (Tabelle 66).

Tabelle 66: Liste aller „in sacco“ Isolate mit der Zuordnung zu bekannten Spezies anhand der höchsten Sequenzähnlichkeit (laut NCBI Datenbankvergleich).

NCBI Datenbank Treffer	Isolate	Max. Indent. (%)	Ident. Isolate	Cellulose-abbau
Clostridium thermocellum	**GGR1-9**; RGR1, RGR2	94	11	**JA**
Clostridium clariflavum	diverse	99	10	**JA**
Acetivibrio cellulolyticus	GS2neu/4/2	99	2	**JA**
Clostridium cellulosi	K-GR Sacco	99	4	**JA**
Summe			**27**	

Insgesamt wurden 27 Isolate aus den „in sacco“ Anreicherungen gewonnen. Davon konnten alle, mit Strategie 1 gewonnenen Isolate bekannten Spezies zugeordnet werden. Generell gehörten alle Isolate zur Familie *Ruminococcaceae* und waren in der Lage kristalline Cellulose abzubauen. Die 11 Isolate, die mit Hilfe von Gärrest aus dem BM2T Fermenter und Strategie 2 gewonnen wurden, zeigten mit 94,36 % die höchste 16S rRNA Sequenzübereinstimmung mit dem *C. thermocellum* Typstamm ATCC 27405 und gehörten basierend auf den 16S rRNA Sequenzanalysen und den Kriterien gemäß Tindall et al. (2010) zu einer bisher nicht beschriebenen Gattung innerhalb der Familie *Ruminococcaceae*.

3.3.1 CHARAKTERISIERUNG AUSGEWÄHLTER „IN SACCO“ ISOLATE

Exemplarisch für die Gruppe dieser 11 Isolate, die eine neue Gattung innerhalb der Familie *Ruminococcaceae* darstellen, wurde Isolat GGR1 zur weiteren Charakterisierung und Genomsequenzierung ausgewählt. Die komplette 16S rRNA Gensequenz wurde in der NCBI Datenbank unter der Zugangsnummer LN868252 hinterlegt.

3.3.1.1 Physiologische Charakterisierung GGR1

Die wichtigsten Ergebnisse der morphologischen und physiologischen Charakterisierung von Isolat GGR1 sind in Tabelle 67 zusammengefasst.

*Tabelle 67: Wichtige morphologische und physiologische Eigenschaften von Isolat GGR1. *Filterpapier (Whatman No. 1)*

Characteristika	
Temperaturbereich (°C)	45 - 65
pH-Bereich	6,5 – 9,0
Gram Reaktion	+
Zellmorphologie	Stäbchen
Endosporen (terminal)	+
Abbau von:	
Galactose	+
Arabinose	–
Xylan	+
Xylose	+
Cellulose*	+
Mannose	–
Fructose	–
Sucrose	–
Glucose	+
Cellobiose	+
Sorbitol	+
Stärke	–
Haupt-Fermentationsprodukte	H_2, Ethanol, Acetat

Wie bereits erwähnt, gehört das Isolat GGR1 zur Familie *Ruminococcaceae*. Einige Vertreter der Familie *Ruminococcaceae,* die bereits gut charakterisiert wurden, sind ebenfalls in der Lage kristalline Cellulose abzubauen (Koeck et al. 2014a). *Ruminococcaceae* gehören neben den

Lachnospiraceae zu den zwei abundantesten Firmicutes Familien in Gastrointestinal-Trakten von Menschen und Tieren. Diese sehr diverse Gruppe hat als gemeinsames Merkmal den effektiven Abbau von pflanzlicher Biomasse in diversen mikrobiellen Gemeinschaften (Tap et al. 2009). Die cellulolytischen Vertreter fallen vor allem in das *Clostridium* Cluster III. Yutin und Galperin (2013) schlugen eine Umbenennung aller *Clostridium* spp. aus dem Cluster III in den neuen Genus *Ruminiclostridium* vor (Yutin and Galperin 2013). Diese Bezeichnung wurde allerdings in die LPSN (List of Prokaryotic names with Standing in Nomenclature) noch nicht aufgenommen und wird daher in dieser Arbeit noch nicht verwendet. Auch der nächste bekannte Verwandte des neuen Isolats, *C. thermocellum*, gehört zu dieser Gruppe.

Abbildung 53: Zellen von Isolat GGR1 unter dem Phasenkontrastmikroskop (1000x Vergrößerung). Balken = 10 µm

Die optimalen Wachstumsparameter, die Substrat- und Produktspektren und Zellmorphologien sind bei Isolat GGR1 und *C. thermocellum* ATCC 27405 sehr ähnlich (Abbildung 53). *C. thermocellum* kann allerdings Xylan und Xylose nicht direkt verstoffwechseln. Zur Bestimmung des enzymatischen Potentials wurde die spezifische Aktivität der Enzyme im Kulturüberstand (kultiviert in GS2 Medium mit Cellobiose und Avicel) von Isolat GGR1 auf verschiedenen Substraten getestet.

Tabelle 68: Spezifische Aktivität (U/mg) der Proteine im Kulturüberstand von Isolat GGR1 auf verschiedenen Substraten mit Standardabweichung von drei Messwerten

Isolat	Avicel	Xylan	ß-Glucan	CMC	PASC	Arabinoxylan	Xyloglucan
GGR1	0,05 +/- 0,003	4,57 +/- 0,12	5,88 +/- 0,16	0,17 +/- 0,009	0,18 +/- 0,034	5,83 +/- 0,16	4,33 +/- 0,16

Die spezifische Aktivität der Enzyme auf kristalliner Cellulose bewegt sich mit 0,05 U/mg in der gleichen Größenordnung, wie bei *C. thermocellum* Kulturüberstand (0,09 U/mg). Die ß-D-

glycosidische Bindung der Substrate PASC und CMC konnte von den Enzymen im Überstand in gleichem Maße hydrolysiert werden. Der Abbauwert von kristalliner Cellulose ist relativ hoch. Sehr gute Enzymaktivitäten konnten auch auf den Hemicellulosen Xylan, Arabinoxylan und Xyloglucan bestimmt werden (Tabelle 68). Anhand des Substratspektrums und der gemessenen Enzymaktivitäten kann geschlossen werden, dass das Isolat vermutlich eine zentrale Rolle in der hydrolytischen Stufe der Biogasanlage also beim Abbau von cellulose- und hemicellulosehaltiger Biomasse spielt.

3.3.1.2 Genomanalyse GGR1

Die Genomsequenz von Isolat GGR1 war bei Fertigstellung dieser Arbeit noch nicht verfügbar. Um dennoch Informationen über das cellulolytische Enzymsystem dieses Isolats zu erhalten, wurden die aufkonzentrierten Proteine des Kulturüberstandes mittels Gelfiltration aufgetrennt. Von den drei Fraktionen mit der höchsten Proteinkonzentration wurde ein Western Blot mit Anti-Cohesin-Antikörpern angefertigt. Falls Cohesine in der Fraktion vorhanden sind, könnte so auf die Anwesenheit von cellulosomalem Scaffoldinprotein, und damit auf die Anwesenheit von Cellulosomen geschlossen werden.

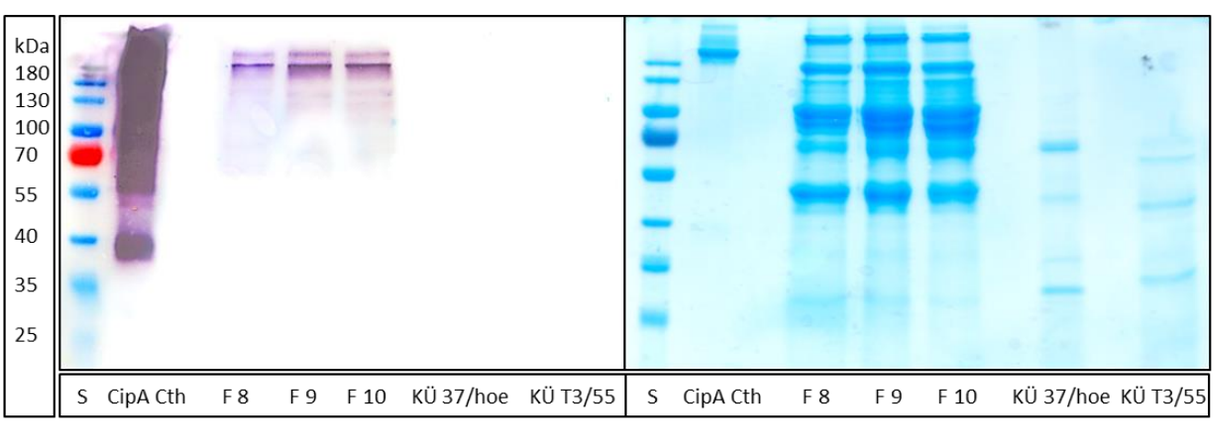


Abbildung 54 links: Western Blot mit Antikörpern gegen Cohesine des CipA Protein von C. thermocellum. Rechts: SDS PAGE analog zu dem Western Blot links. Cth = C. thermocellum. F = Fraktion der Gelfiltration von GGR1 Kulturüberstand. KÜ = Kulturüberstände anderer Isolate (37/hoe, T3/55) als Negativkontrolle. Es wurden jeweils 2 µg Protein aufgetragen.

In Abbildung 54 links ist deutlich zu erkennen, dass mit den Anti-Cohesin-Antikörpern ein makromolekulares Protein, in allen getesteten Fraktionen der Gelfiltration des GGR1 Kulturüberstands, nachgewiesen werden konnte. Da im Gegensatz zur Positivkontrolle kein gereinigtes CipA Protein, sondern vermutlich ein Gemisch aus cellulosomalen Komponenten aufgetragen wurde, ist das Signal wesentlich schwächer. Das nachgewiesene Protein scheint, mit ca. 200 kDa größer zu sein als CipA von *C. thermocellum*. Ein so großes Protein im Kulturüberstand ist ungewöhnlich und könnte, im Zusammenhang mit der Reaktion mit den CipA-Antikörpern und der negativen Reaktion mit den Kontrollüberständen von 37/hoe und T3/55, auf die Anwesenheit eines neuen Scaffoldins, und damit auf Cellulosomen, hinweisen. Ob das Isolat

GGR1 tatsächlich ein makromolekulares Cellulosom besitzt wird erst die Analyse der Genomsequenz klären.

3.3.2 *Amplikonsequenzierung der „in sacco“ Anreicherung*

Um zu überprüfen, ob mit der „in sacco“ Methode tatsächlich cellulolytische Bakterien „*in situ*“ (im Fermenter) angereichert werden konnten und um zu vergleichen ob tatsächlich die in der Anreicherung dominierenden Spezies isoliert wurden, wurde eine Amplikonsequenzierung der DNA-Proben durchgeführt. Dazu wurde das Filterpapier von jedem Entnahmezeitpunkt (1d (1), 2d (2), 4d (3) und 7d (4)) aus BM1T und BM2T homogenisiert und die DNA daraus isoliert. Die Proben von Tag 2 und 4 wurden für die Anlage BM2T vereinigt. Anschließend wurde das 16S rRNA Gen, mit spezifischen Primern gegen das eubakterielle 16S rRNA Gen (Tabelle 22), amplifiziert und in *E. coli* kloniert. Von jedem Probezeitpunkt wurden 15-20 Klone sequenziert. Um zu prüfen, ob sich durch einen Anreicherungsschritt im Labormedium die Zusammensetzung der Mikrobengemeinschaft ändert und in welchem Umfang dies geschieht, wurden zusätzlich Amplikonsequenzierungen des 2. und 5. Anreicherungsschritts durchgeführt.

3.3.2.1 Amplikonsequenzierung BM1T

Repräsentativ sind hier nur die detaillierten Ergebnisse der Sequenzierung von Tag 7 des Reaktors BM1T dargestellt (Tabelle 69).

Tabelle 69: Bester NCBI Datenbanktreffer für kultivierte Bakterien der klonierten 16SrRNA Gene einer DNA Probe des „in sacco“ Ansatzes: BM1T Entnahmetag 7. Sequenzen die über 98 % Übereinstimmung aufweisen sind in gleicher Farbe hinterlegt.

Klon	Sequenzident. %	Nächste verwandte kultivierte Spezies
7.1	93	*Halocella cellulosilytica strain Z-10151*
7.2	91	*Halocella cellulosilytica strain Z-10151*
7.4	91	*Halocella cellulosilytica strain Z-10151*
7.5	90	*Turicibacter sanguinis strain 111732/2010*
7.6	94	*Halocella cellulosilytica strain Z-10151*
7.7	96	*Clostridium thermocellum ATCC 27405*
7.9	92	*Halocella cellulosilytica strain Z-10151*
7.10	94	*Halocella cellulosilytica strain Z-10151*
7.11	93	*Halocella cellulosilytica strain Z-10151*
7.13	92	*Clostridium thermocellum ATCC 27405*
7.14	92	*Clostridium thermocellum ATCC 27405*
7.18	91	*Hydrogenispora ethanolica*
7.19	96	*Clostridium thermocellum ATCC 27405*
7.20	96	*Clostridium thermocellum ATCC 27405*
7.21	94	*Halocella cellulosilytica strain Z-10151*
7.23	94	*Halocella cellulosilytica strain Z-10151*
7.24	94	*Clostridium thermocellum ATCC 27405*
7.25	94	*Halocella cellulosilytica strain Z-10151*
7.27	96	*Clostridium thermocellum ATCC 27405*

Die restlichen Ergebnisse der anderen Probenahmezeitpunkte sind in Abbildung 55 zusammengefasst.

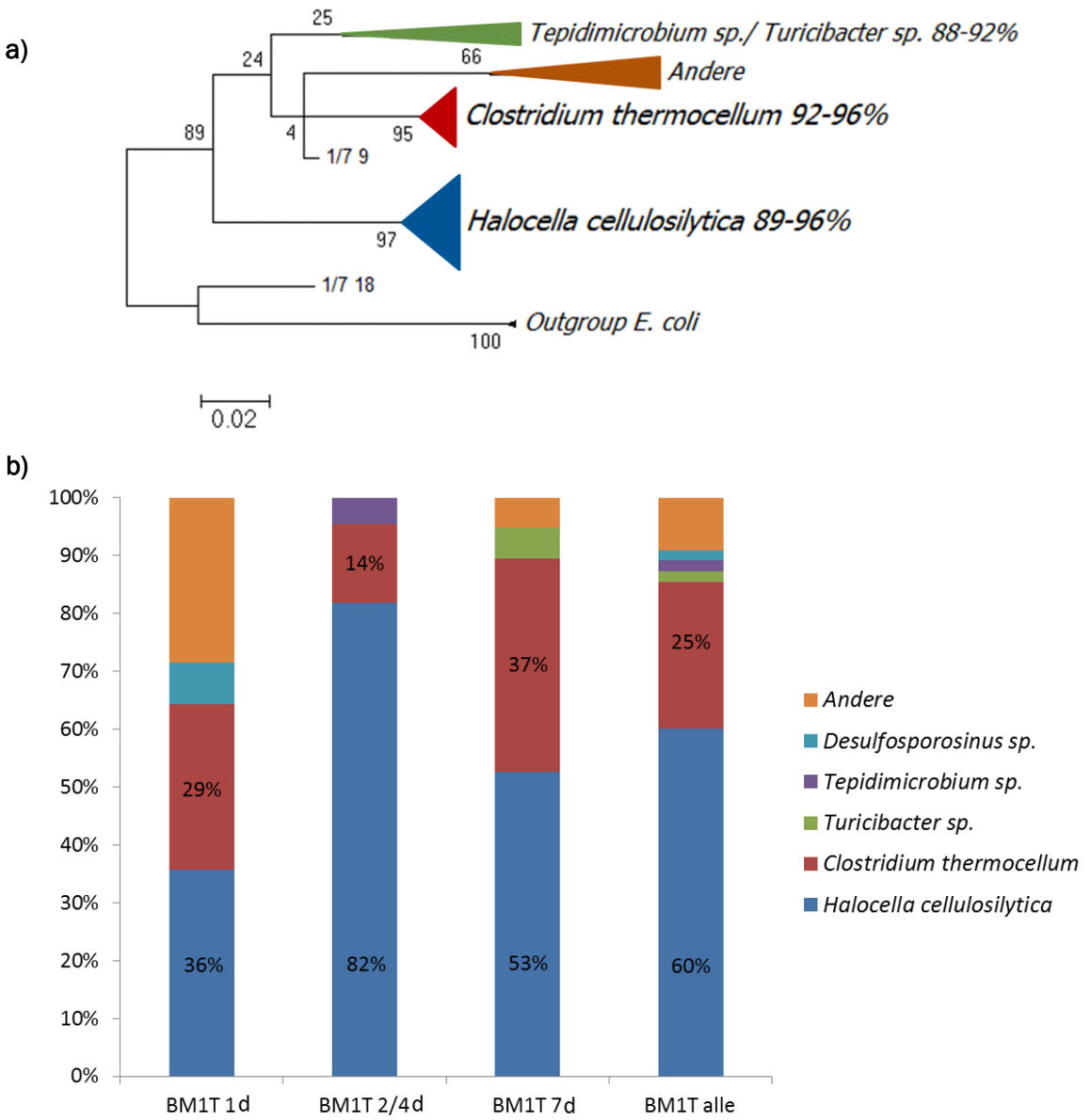


Abbildung 55: a) Phylogenetischer Baum der 16S rRNA Gensequenzen der Amplikonsequenzierung einer „in sacco“ Anreicherungen aus dem BM1T Reaktor unter Verwendung eines maximum likelihood-basierten Algorithmus. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in taxonomische Einheiten zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Sequenzen, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede. b) Prozentualler Anteil der wichtigsten taxonomischen Einheiten in der Amplikonsequenzierung der verschiedenen „in sacco“ Probenahmezeitpunkte von BM1T. Angegeben ist immer die nächste bekannte verwandte Spezies oder Gattung. $N_{gesamt} = 70$.

Bei der Erstellung der Klonbibliotheken aus den „in sacco“ Anreicherungen stammten zu über 80 % der sequenzierten Klone aus zwei definierten taxonomischen Einheiten, innerhalb derer die Sequenzen >98 % Ähnlichkeit aufwiesen. Diese haben die höchste 16S rRNA Gensequenzübereinstimmung bezüglich bereits isolierter Spezies mit cellulolytischen Bakterien. Die häufigste taxonomische Gruppe hatte als nächsten Verwandten *Halocella cellulosilytica* mit einer 16S rRNA Sequenzübereinstimmung von 89-96 %. Die zweit häufigste taxonomische

Gruppe wies mit 92-96 % Sequenzübereinstimmung den höchsten Verwandtschaftsgrad zu *C. thermocellum* auf (Abbildung 55a). Sowohl *C. thermocellum* als auch *H. cellulosilytica* Stämme sind in der Lage, kristalline Cellulose abzubauen (Koeck et al. 2014a). Somit war die Anreicherung cellulolytischer Bakteriengemeinschaften in situ erfolgreich. Bemerkenswert ist, dass generell die Sequenzähnlichkeit mit kultivierten Bakterien 97 % nicht übersteigt. Somit kann keiner der Klone direkt einer Art zugeordnet werden. Zwischen den einzelnen Probenahmezeitpunkten fand eine sukzessive Anreicherung der beiden Hauptpopulationen statt. Nach Tag 1 hatten Klone aus anderen Gruppen noch einen Anteil von 35 %, diese wurden bis Tag 7 auf 22 % reduziert. Die *„H. cellulosilytica“* Gruppe erreichte nach sieben Tagen einen Anteil von 53 % und die *„C. thermocellum“* Gruppe von 37 % (Abbildung 55a). Auch der Grad des Filterpapierabbaus nahm mit steigender Inkubationszeit immer weiter zu.

3.3.2.2 Amplikonsequenzierung BM2T

Wie auch schon bei Reaktor BM1T wurden repräsentativ für alle Probenahmezeitpunkte nur die detaillierten Ergebnisse der Sequenzierung von Tag 7 des Reaktors BM2T dargestellt (Tabelle 70).

Tabelle 70: Bester NCBI Datenbanktreffer für kultivierte Bakterien der klonierten 16SrRNA Gene einer DNA Probe des „in sacco“ Ansatzes: BM1T Entnahmetag 7. Sequenzen die über 98 % Übereinstimmung aufweisen, sind in gleicher Farbe hinterlegt.

Klon	Sequenzident. %	Nächste verwandte kultivierte Spezies
7.1	93	*Halocella cellulosilytica strain Z-10151*
7.2	92	*Turicibacter sanguinis strain 111732/2010*
7.3	90	*Halocella cellulosilytica strain Z-10151*
7.4	90	*Halocella cellulosilytica strain Z-10151*
7.6	95	*Clostridium thermocellum JGU1/2*
7.7	95	*Clostridium thermocellum JGU1/2*
7.8	94	*Halocella cellulosilytica strain Z-10151*
7.9	94	*Tepidimicrobium ferriphilum strain SB91*
7.11	95	*Halocella cellulosilytica strain Z-10151*
7.12	95	*Clostridium thermocellum ATB2/1*
7.13	95	*Clostridium thermocellum ATB2/1*
7.14	89	*Syntrophaceticus schinkii strain Sp3*
7.15	95	*Clostridium thermocellum ATB2/1*
7.16	91	*Halocella cellulosilytica strain Z-10151*
7.19	91	*Halocella cellulosilytica strain Z-10151*
7.20	97	*Tepidimicrobium xylanilyticum strain PML14*
7.21	95	*Clostridium thermocellum JGU1/2*
7.22	94	*Halocella cellulosilytica strain Z-10151*
7.23	92	*Halocella cellulosilytica strain Z-10151*
7.27	96	*Halocella cellulosilytica strain Z-10151*
7.28	94	*Halocella cellulosilytica strain Z-10151*
7.29	96	*Clostridium thermocellum JGU1/2*
7.30	95	*Halocella cellulosilytica strain Z-10151*
7.31	93	*Halocella cellulosilytica strain Z-10151*
7.32	94	*Halocella cellulosilytica strain Z-10151*
7.33	93	*Halocella cellulosilytica strain Z-10151*
7.34	92	*Halocella cellulosilytica strain Z-10151*
7.36	92	*Turicibacter sanguinis strain 111732/2010*
7.38	94	*Tepidimicrobium xylanilyticum strain PML14*
7.39	91	*Clostridium thermocellum JGU1/2*
7.40	90	*Bacillus mucilaginosus*

Die restlichen Ergebnisse der anderen Probenahmezeitpunkte sind in Abbildung 56 zusammengefasst.

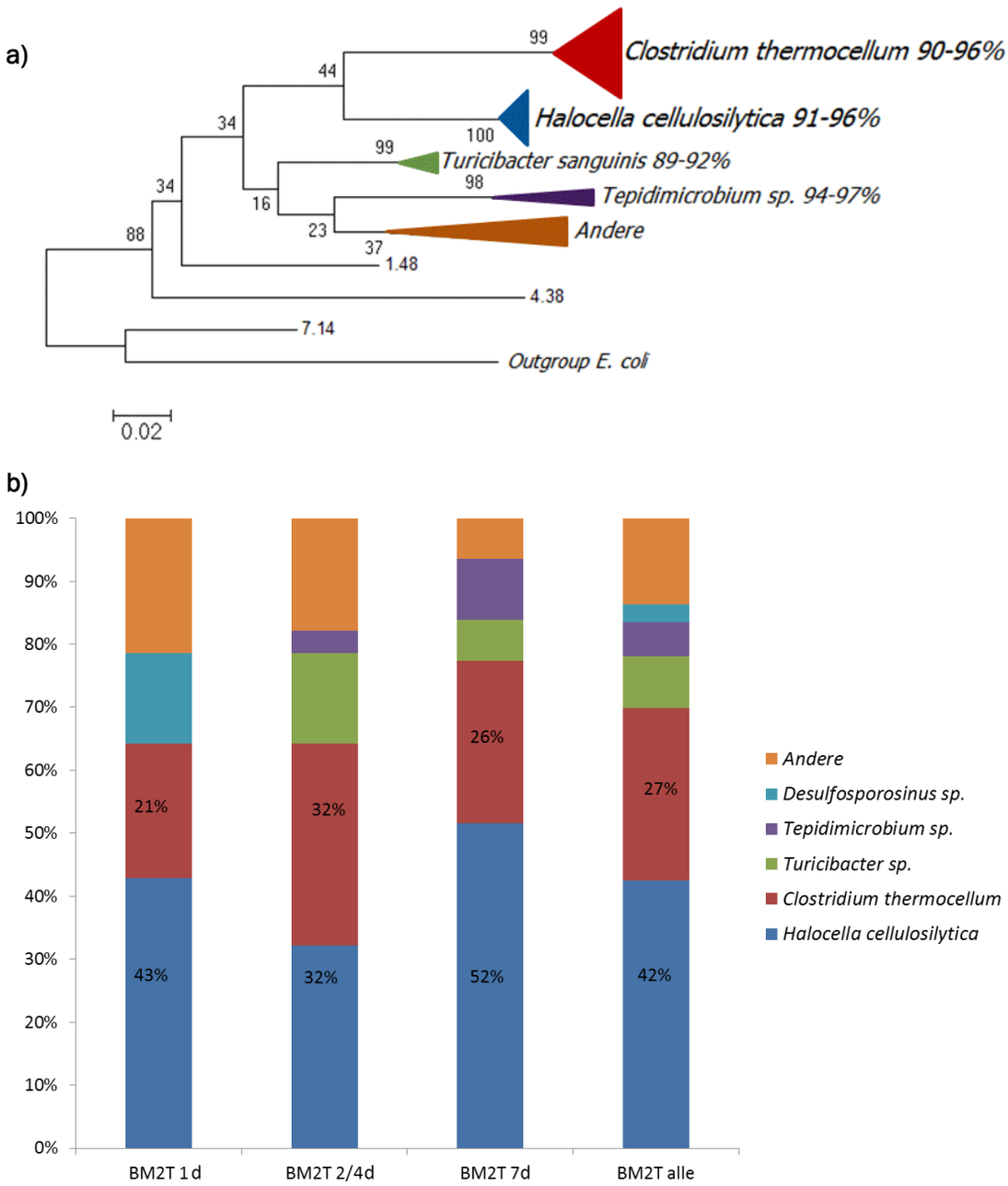


Abbildung 56: a) Phylogenetischer Baum der 16S rRNA Gen Sequenzen der Amplikonsequenzierung einer „in sacco“ Anreicherungen aus dem BM1T Reaktor unter Verwendung eines maximum likelihood-basierten Algorithmus. Subbäume von taxonomisch eng verwandten Gruppen mit mehr als 5 Isolaten sind in taxonomische Einheiten zusammengefasst. Die Höhe des Dreiecks ist abhängig von der Anzahl der Sequenzen, die Länge ist abhängig vom Maß der Sequenzunterschiede zwischen den Isolaten einer taxonomischen Einheit. Der Größenbalken indiziert 2 % Basenunterschiede. b) Prozentualler Anteil der wichtigsten taxonomischen Einheiten in der Amplikonsequenzierung der verschiedenen „in sacco“ Probenahmezeitpunkte von BM1T. Angegeben ist immer die nächste bekannte verwandte Spezies oder Gattung. N gesamt= 75.

Die meisten 16SrRNA Sequenzen der Klonbibliothek lieferten wiederum die höchste Übereinstimmung mit cellulolytischen Bakterien. Diese bilden zwei definierte taxonomische Einheiten innerhalb derer die Sequenzen >98 % Ähnlichkeit aufweisen. Die häufigste

taxonomische Gruppe hatte als nächsten Verwandten wieder *Halocella cellulosilytica* mit einer 16S rRNA Sequenzübereinstimmung von 91-96 %. Die zweit häufigste taxonomische Gruppe wies mit 90-96 % Sequenzübereinstimmung wieder den höchsten Verwandtschaftsgrad zu *C. thermocellum* auf (Abbildung 56a). Somit war auch die Anreicherung im Reaktor BM2T erfolgreich. Auch beim Vergleich der beiden Laborfermenter weisen die beiden häufigsten taxonomischen Einheiten über 98 % Sequenzähnlichkeit auf. Im Vergleich zu den ersten beiden Entnahmen nahm auch in den Säckchen aus BM2T der Anteil an den beiden Hauptpopulationen zu und erreichte nach sieben Tagen Inkubation einen Anteil von 76 % (Abbildung 56b). Dieser liegt etwas unter den Werten von Reaktor BM1T. Das Verhältnis zwischen den beiden Hauptgruppen blieb bei Reaktor BM2T während des gesamten Versuchs weitgehend konstant, während der Anteil aller Klone an den beiden Taxa bei Reaktor BM2T stärker schwankte.

3.3.2.3 Amplikonsequenzierung der Anreicherungen

Um die „in sacco“ angereicherten Taxa zu isolieren, wurde wie unter 3.3 beschrieben versucht, die an der Cellulose anhaftenden Bakterien unter Laborbedingungen zu kultivieren. Für die Anreicherungen wurde GS2 Medium mit 10 % Gärrest aus dem Fermenter BM2T genutzt. Als Inokulum diente Filterpapier, dass an Tag 7 aus BM2T entnommen wurde. Die höchste Verdünnungsstufe mit Filterpapierabbau wurde in jedem Anreicherungsschritt erneut im gleichen Medium verdünnt (Isolierungsstrategie 2). Von Anreicherung zwei und fünf wurden jeweils 10 -20 Klone sequenziert.

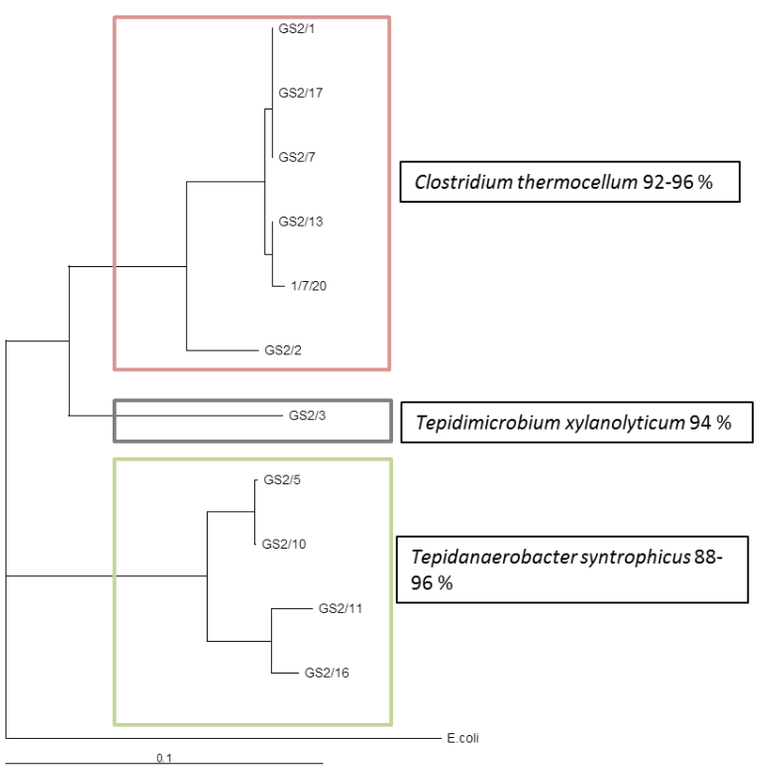


Abbildung 57: Phylogenetische Analyse der Amplikonsequenzierung des zweiten Anreicherungsschrittes

Bereits nach der ersten Anreicherung unter Laborbedingungen wurde die taxonomische Einheit „*Halocella sp.*“ nicht mehr nachgewiesen. Dafür tauchte eine neue Subpopulation auf, die den höchsten Verwandtschaftsgrad zu *Tepidanaerobacter syntrophicus* (88 – 96 %) aufwies. Die taxonomische Einheit mit der höchsten 16S rRNA Sequenzähnlichkeit zu *C. thermocellum* Stämmen blieb ebenso wie das „*Tepidimicrobium* Cluster“ erhalten. Die „*Halocella* sp. Gruppe“ ging also durch die Anreicherungsprozedur verloren.

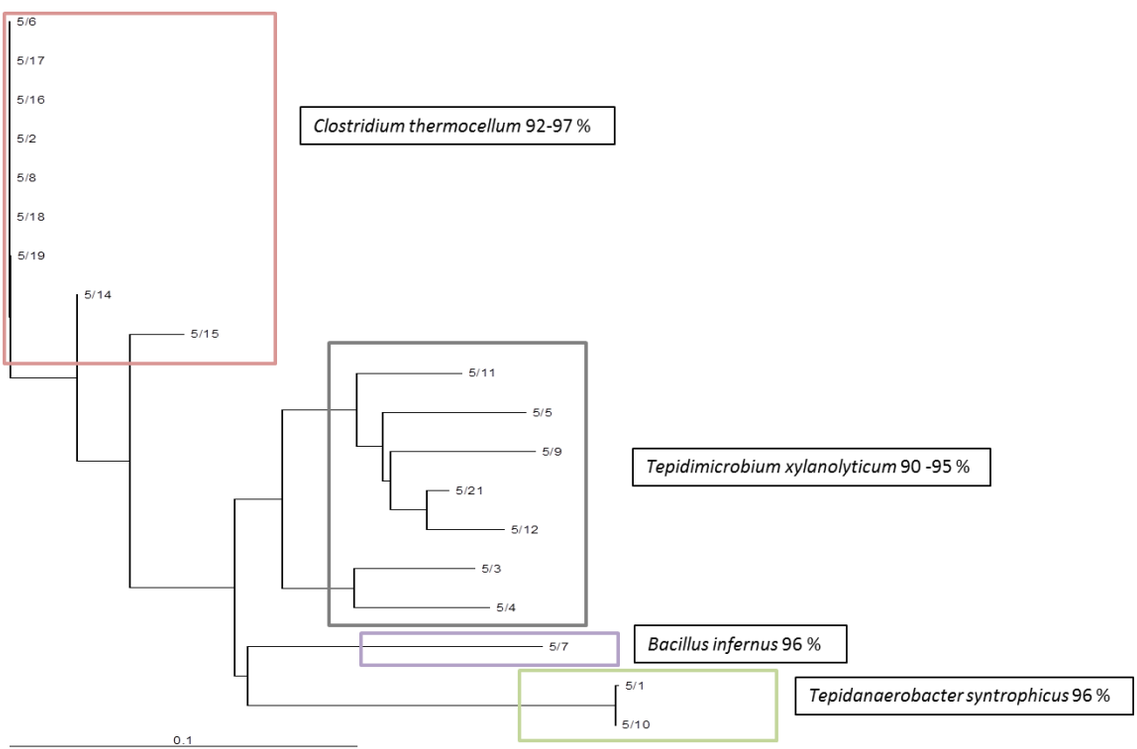


Abbildung 58: Amplikonsequenzierung des fünften Anreicherungsschrittes

In der Mischkultur nach fünf Verdünnungsreihen nahm die Häufigkeit der *Tepidanaerobacter syntrophicus* Gruppe ab. Dafür stieg die Anzahl der sequenzierten Klone des *Tepidimicrobium* Clusters stark an. Die *C. thermocellum* Gruppe blieb unverändert abundant. Ein multiples Alignment von Sequenzen aus mit Klonen aus der „in sacco“ Bibliothek von Tag 7 zeigte, dass es sich immer noch um dieselbe taxonomische Gruppe handelte.

3.3.3 SCHLUSSFOLGERUNGEN

Vergleicht man die Isolate aus den „in sacco“ Versuchen mit den Klonbibliotheken, so fällt auf, dass mit den zunächst eingesetzten Medium und Zusätzen zwar cellulolytische Bakterien angereichert und isoliert werden konnten, diese sind allerdings in den Amplikonsequenzierungen nach Anreicherung unter Laborbedingungen nicht wieder zu finden. Umgekehrt wurde keiner der

16S-Klone in Reinkultur isoliert. Diese Isolate scheinen im Laborfermenter nicht abundant zu sein und für den Abbau von Biomasse eher eine untergeordnete Rolle zu spielen. Erst unter Laborbedingungen wurden diese Populationen gezielt angereichert. Dass es sich durchwegs um bereits mehrfach isolierte und gut charakterisierte Arten handelte, lässt den Schluss zu, dass besonders diese sehr gut kultivierbar sind und andere schwer kultivierbare Spezies bei den gewählten Bedingungen verdrängen. Erst eine Änderung der Isolierungsstrategie und Zugabe von 10 % Gärrest aus dem betreffenden Reaktor lieferte erste Isolate aus einer der taxonomischen Hauptgruppen der Klonbibliothek, welche die höchste Sequenzübereinstimmung mit *C. thermocellum* ausweist. Die Gruppe *Halocella sp.* entzog sich weiterhin allen Isolierungsversuchen. Eine weitere Amplikonsequenzierung zeigte, dass diese Gruppe bereits nach der ersten Anreicherung unter Laborbedingungen nicht mehr wiedergefunden werden konnte. Somit lässt sich die Hypothese aufstellen, dass eine wichtige Gruppe von Schlüsselbakterien für den Celluloseabbau in dieser Anlage mit den bisherigen Kultivierungsversuchen nicht erfasst werden konnte. Allerdings ist eine dieser wichtigen Gruppen nun identifiziert und kann daher in Zukunft mit gezielten Strategien in weitergehende Untersuchungen mit einbezogen werden. Auch steht nun eine wirksame Methode zur Verfügung, die es erlauben sollte, auch aus anderen Fermentern mit anderen Substraten und Fermentationsbedingungen Bakterien zu identifizieren, die am Biomasseabbau entscheidenden Anteil haben.

3.4 Vergleich der hydrolytischen Enzyme aller sequenzierten Genome neuer Isolate

Abschließend wurden die Abundanzen der auf Kohlehydraten aktiven Enzyme zwischen den verschiedenen Isolaten mit sequenziertem Genom verglichen. In Abbildung 59 wurde zunächst die Verteilung, der in der CAZy Datenbank aufgeführten Kategorien, gegenübergestellt. Auffällig ist, dass im Genom von *C. thermocellum* BC1 eine signifikant höhere Anzahl an Glycosidhydrolasen und Kohlenhydratbindemodulen codiert ist. Die Glycosidtransferasen sind dagegen bei Iso2/2 37 am abundantesten. Ein Isolat der Familie *Lachnospiraceae* (T3/55) besitzt die meisten Kohlenhydratesterasen. Im Gegensatz zu den Ergebnissen von Biddle et al. (2013) konnte die allgemeine These, dass generell die Familien *Ruminococcaceae* und *Lachnospiraceae* signifikant höhere Abundanzen an GHs und CBMs als andere Familien aufweisen, nicht bestätigt werden (Biddle et al. 2013).

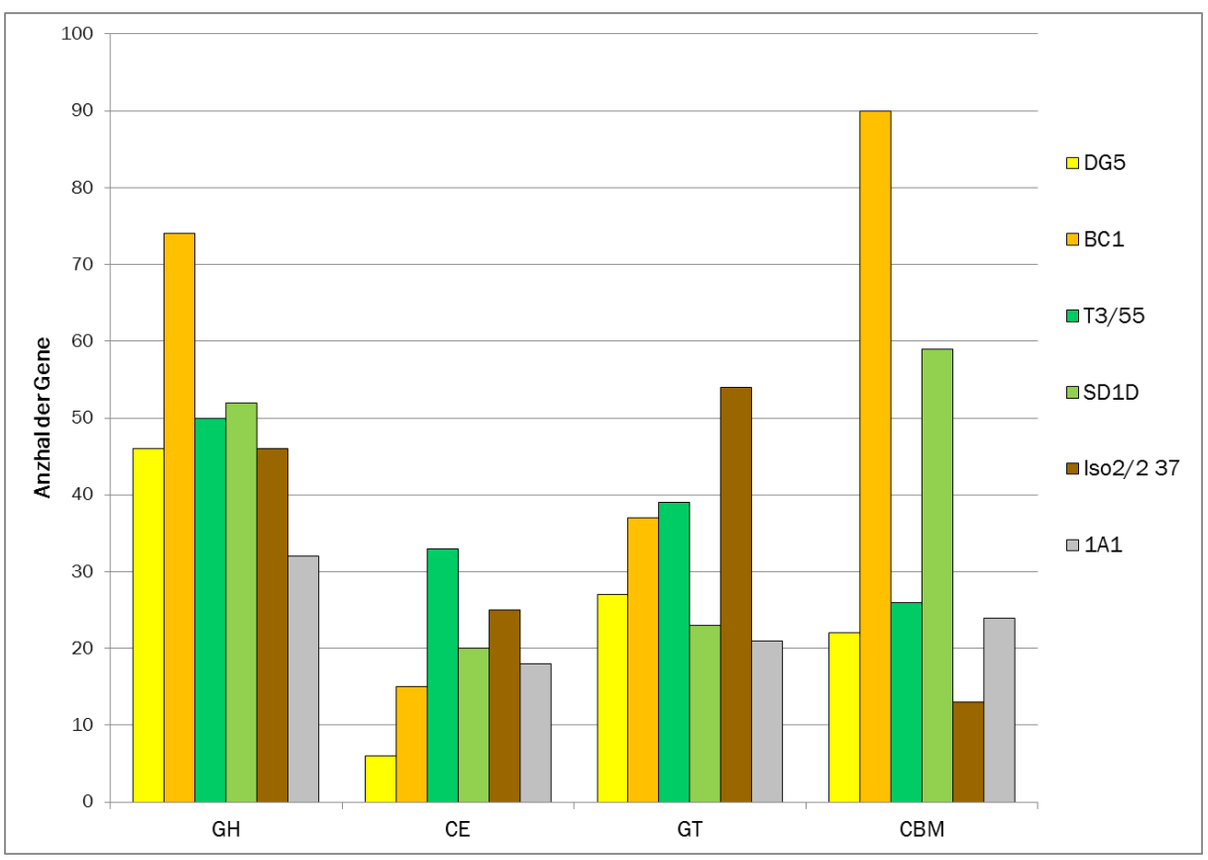


Abbildung 59: Vergleich der Abundanzen von Kohlenhydrataktiven Enzyme (CAZy), die in den Genomen der neuen Isolate codiert sind. GH = Glycosidhydrolasen, GT = Glucosyltransferasen, CE = Carbohydratesterasen, PL = Polysaccharidlyasen.

Ein Vergleich der Häufigkeiten wichtiger Glycosidhydrolasen zwischen den verschiedenen Isolaten zeigte, dass BC1 neben den meisten Glycosidhydrolasen auch signifikant mehr Cellulasen besitzt. Diese verteilen sich auf zwei Cellulasen der Familie 48, 10 potentielle Cellulasen der Familie 5 und 16 Cellulasen der Familie 9 (Abbildung 60). Die Verteilung der Cellulasen korreliert mit der

Beschaffenheit des cellulolytischen Systems und der Celluloseabbau-Effizienz. Unter den sequenzierten Isolaten besitzen nur BC1 und SD1D ein Cellulosom und daher auch signifikant mehr Cellulasen und die höchste hydrolytische Aktivität auf Cellulose (Avicel) (Tabelle 51, Tabelle 46). Die beiden Isolate (1A1, Iso2/2 37), die nicht in der Lage sind kristalline Cellulose abzubauen, haben im Genom weder ein GH5 Enzym noch ein GH48 Enzym codiert. Das amylolytische Isolat *Bacillus thermoamylovorans* 1A1 weist allerdings eine signifikant höhere Anzahl an GH13 Genen auf, welche für mögliche Amylasen codieren.

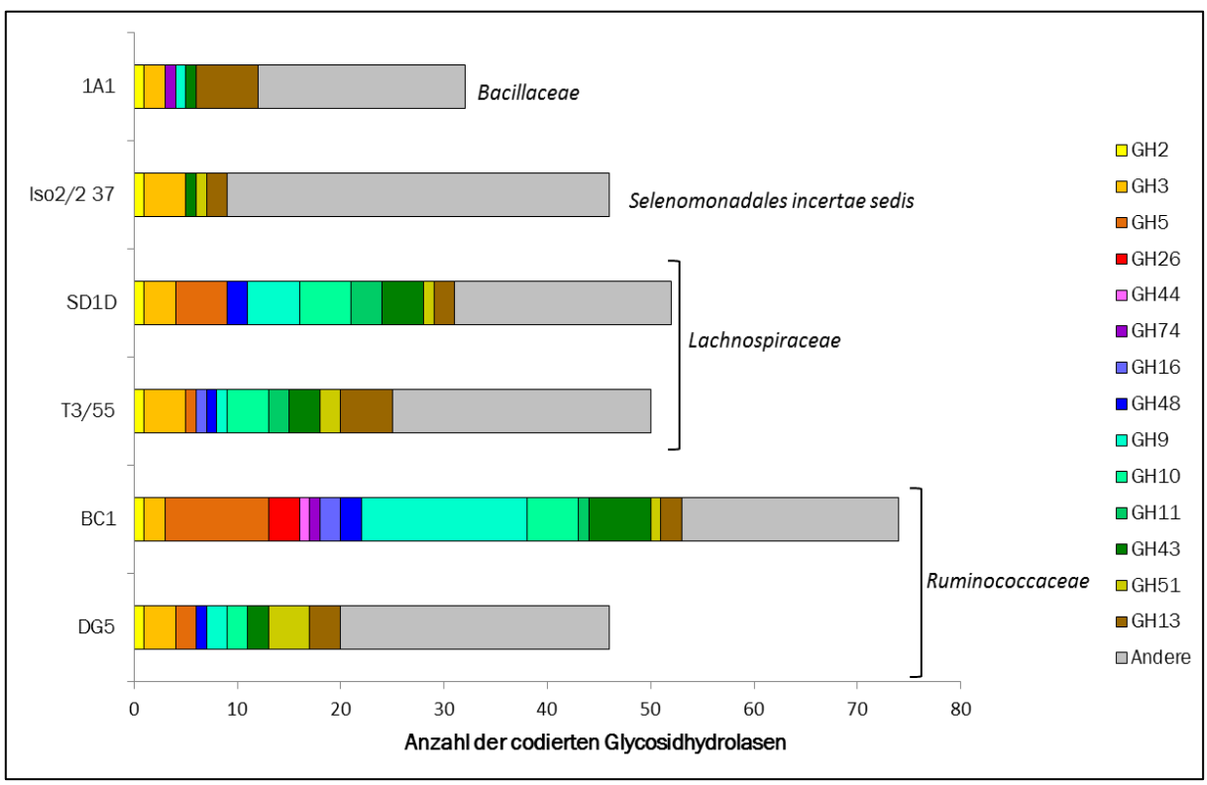


Abbildung 60: Glycosidhydrolase Profile der Genome aller sequenzierten Isolate. Gezeigt sind ausgewählte Familien welche typischerweise Xylanasen (GH5, GH10, GH11), Cellulasen (GH5, GH9, GH48), Mannanasen (GH5, GH26), β-Glucanasen/Xyloglucanasen (GH16, GH44, GH74), Arabinofuranosidasen/Xylosidasen (GH51, GH43), Glycosidasen (GH2, GH3), und Amylasen (GH13) enthalten. Die funktionelle Zuordnung der GH Familien wurde aus der CAZy Website entnommen (http://www.cazy.org).

Besonders die Stämme aus den Familien *Lachnospiraceae* und *Ruminococcaceae,* welche zu den zwei abundantesten Firmicutes Familien in Gastrointestinal-Trakten und Biogasanlagen zählen, sind mit zahlreichen cellulolytischen und hemicellulolytischen Enzymen ausgestattet (Biddle et al. 2013, Tap et al. 2009). Dies ermöglicht ihnen im Gegensatz zu den anderen Isolaten den effektiven Abbau von pflanzlicher, lignocellulosehaltiger Biomasse. Die durch diese hydrolytischen Bakterien gebildeten Zwischenprodukte werden dann in der zweiten Phase des Biomasseabbaus durch säurebildende Mikroorganismen, wie die Isolate 1A1 und Iso2/2 37, in kurzkettige Carbonsäuren (z.B. Butter-, Essig-, Ameisen-, Propionsäure), niedermolekulare Alkohole, Wasserstoff, CO_2 und Wasser umgewandelt.

4 DISKUSSION

Das Hauptziel dieser Arbeit war die Isolierung neuer cellulolytischer Bakterien aus einer definierten thermophilen und mesophilen Biogasanlage. Dabei wurden auch einige bis dahin unkultivierte, cellulolytische mesophile und thermophile Bakterienarten und Gattungen isoliert. Ein Isolat aus jeder dieser Gruppen wurde für weitere physiologische Untersuchungen, Genomsequenzierung und Charakterisierung des cellulolytischen Systems ausgewählt. Zudem wurden in dieser Studie erstmals *C. thermocellum* Stämme aus verschiedenen Biogasanlagen isoliert und die genetische und phänotypische Variation zwischen den Stämmen bestimmt.

4.1 GENOTYPISIERUNG

Bisher wurde die Isolierung von *C. thermocellum* Reinkulturen aus Biogasanlagen nicht beschrieben, obwohl diese Art häufig als wichtiger Bestandteil des mikrobiellen Konsortiums in Biogasanlagen nachgewiesen wurde (Krause et al. 2008, Wirth et al. 2012). Im Rahmen dieser Arbeit wurden erfolgreich je zwei *C. thermocellum* Stämme aus neun verschiedenen Biogasanlagen gewonnen und deren genetische und metabolische Unterschiede, sowie deren cellulolytisches Potential ermittelt.

Eine verlässliche Technik für die Genotypisierung von *C. thermocellum* Stämmen sollte dabei folgende Kriterien erfüllen: Eignung für Hochdurchsatzanalysen, geringe Kosten, Reproduzierbarkeit, Möglichkeit zur digitalen Datenspeicherung und Archivierung (Sauer et al. 2008). Alle getesteten PCR-basierten Methoden lassen sich schnell und einfach implementieren. Diese getesteten etablierten Genotypisierungstechniken hatten dabei aber ihre Vor- und Nachteile: für den RADP-Assay muss von dem Zielorganismus vorher keinerlei Sequenzinformation vorliegen; für eine Ribotypisierung muss zumindest die Sequenz des rRNA Operons bekannt sein; und für VNTR oder RIME sollte für die Primer-Entwicklung die Genomsequenz bekannt sein. Dadurch, dass die genomische DNA-Sequenz über einen längeren Zeitraum stabil bleibt und sich nicht abhängig von Umwelteinflüssen ändert, können PCR-basierte Identifizierungs- und Genotypisierungsmethoden verlässliche und reproduzierbare Informationen von bakteriellen Proben generieren (Sauer et al. 2008). Ein großer Nachteil dieser auf Bandenmustern basierenden Techniken ist allerdings die schlechte Übertragbarkeit der generierten Daten (Zaiß et al. 2009). In dieser Arbeit wurde gezeigt, dass anders als bei *C. difficile* die Sequenz von jedem der vier rRNA Operone zwischen den verschiedenen *C. thermocellum* Stämmen identisch ist. Nur die vier verschiedenen Operone eines Stammes unterscheiden sich in Länge und Sequenz. Daher ist diese Methode nicht geeignet, um verschiedene *C. thermocellum* Stämme zu unterscheiden. Der RADP und der VNTR Assay ergaben zwar verschiedene Bandenmuster bzw. Amplikonlängen, konnten aber nicht alle Isolate aus unterschiedlichen Biogasanlagen voneinander abgrenzen. Die in dieser Arbeit entwickelte

Verwendung des mobilen Elements IS1447 für die Genotypisierung (RIME Assay) zeigt die beste Differenzierung zwischen verschiedenen *C. thermocellum* Stämmen und lieferte einen Genotyp für jede Biogasanlage. Verbunden mit der Suche nach *C. thermocellum* Mutanten mit unvollständiger Cellulosehydrolyse wurde dieses neue und aktive Insertionselement (IS) IS1447 entdeckt, welches zur IS3 Familie gehört und in den *cipA* Leserahmen integriert ist (Zverlov et al. 2008). Die IS-Sequenz besteht aus einem Transposase-Gen und den invertierten Wiederholungen IRR und IRS an den Enden. Die Insertion führt zu einer Vervielfältigung von 3 Basenpaaren innerhalb der nicht-spezifischen Zielsequenz. Dieser Mangel an Spezifität ermöglicht die Integration an variablen Positionen im *C. thermocellum* Genom (Zverlov et al 2008). *In silico* konnten 18 Kopien den mobilen Elements IS1447 in der Genomsequenz von *C. thermocellum* ATCC 27405 identifiziert werden (Zverlov et al. 2008), sowie 15 vollständige und zwei fragmentierte Mobilelemente im Genom von Stamm DSM 1313. Kopien des mobilen Elements wurden auch in allen, im Rahmen dieser Arbeit, daraufhin untersuchten *C. thermocellum* Stämmen mittels PCR nachgewiesen. Dies macht IS1447 interessant für die Entwicklung eines einfachen und günstigen genetischen Markers zur Genotypisierung verschiedener *C. thermocellum* Stämme. Kopien desselben IS1447 wurden auch in den Genomen von eng verwandten Bakterien wie *C. stercorarium* und *C. clariflavum* gefunden. Daher sollte die einfache Übertragung dieser Methode auf andere Spezies möglich sein. Es sollte sogar möglich sein durch die Verwendung von anderen, zufällig integrierbaren und noch aktiven mobilen Elementen die Methode auf andere taxonomische Gruppen zu übertragen. Da die Sequenz des Transposons bekannt ist, können die generierten Daten zudem zwischen verschiedenen Laboren übertragen und verglichen werden.

Die Genotypisierung anhand der Zellproteinprofile würde ebenfalls all diese Kriterien erfüllen und offenbarte sich auch als gute und einfache Methode, um Bakterien bis zum Speziesebene zu unterscheiden. Ähnlich wie die Ribotypisierung ist diese Methode aber nicht geeignet, um *C. thermocellum* Isolate verlässlich bis auf Subspeziesebene zu unterscheiden. Zum einen sind in vielen Fällen die Klassifizierungsergebnisse aus Massenspektraldaten und 16S ribosomalen Gensequenzen durch die Koevolution ribosomaler Proteine und ribosomaler Nukleinsäuren zu ähnlich (Sauer et al. 2008). Zum anderen können Variationen der Kultivierungsbedingungen oder Extraktionsmethoden auch in denselben Bakterien unterschiedliche Spektren erzeugen (Sauer and Kliem 2010).

Generell unterschieden sich die Stämme verschiedener Biogasanlagen anhand ihrer Genotypisierungsergebnisse und physiologischen Eigenschaften. Beide Stämme einer Biogasanlage zeigten aber immer identische Bandenmuster mit jeder der angewandten Genotypisierungsmethoden. Diese Ergebnisse wurden durch die Berechnungen der Bandenmusterähnlichkeit mit dem „Coincidence index“ bestätigt. Bezüglich der optimalen Wachstumsbedingungen, cellulolytischen Aktivität und Substratspezifität zeigten die Isolate einer

Biogasanlage ebenfalls ähnliche Charakteristika. Die Resultate lassen vermuten, dass alle Isolate einer Biogasanlage aufgrund der ähnlichen Eigenschaften zur selben *C. thermocellum* Subspezies (Stamm) gehören. Jeweils ein einziger Stamm scheint demnach in jeder Biogasanlagen dominant vorzukommen und somit angereichert und isoliert zu werden. Dieser ist vermutlich an die Prozess-Parameter der entsprechenden Biogasanlage am besten angepasst und hat sich durchgesetzt. Dies lässt die Vermutung zu, dass dies auch auf andere Spezies der mikrobiellen Lebensgemeinschaft in Biogasanlagen zutrifft. Ähnlich wie andere *C. thermocellum* Stämme aus vorangegangenen Studien (z.B. JW20, Freier et al. 1988) konnten einige Isolate im Gegensatz zum Typstamm (ATCC 27405) Glucose und Fructose verwerten. Dies wurde auch in vorangegangenen Arbeiten für die Stämme BC1 und NB2 nachgewiesen (Köllmeier 2013). Laut statistischer Analysen zeigten besonders die beiden Isolate aus Biogasanlage J1 auch eine signifikant höhere Cellulaseaktivität als der Typstamm *C. thermocellum* ATCC 27405. Für den Stamm BC1 wurde ebenfalls eine signifikant höhere Aktivität des Cellulosoms auf Cellulose bei niedrigen pH-Werten nachgewiesen. Allerdings lassen sich aus der komparativen Analyse der Genome von BC1 und ATCC 27405 keine Rückschlüsse auf die Unterschiede in der Abbaueffizienz von Cellulose bei niedrigen pH-Werten ziehen. Ein vorangegangener Vergleich der Cellulosomen mittels MALDI-TOF MS Analyse zeigte, dass bei den beiden Stämmen mit besserem Biomasseabbau (BC1 und NB2), mehr Xylanasen in deren Cellulosompräparationen detektiert werden konnten (relativ zum Typstamm). Dies ist ein Indiz für die zentrale Bedeutung dieser Enzymklasse beim Abbau pflanzlicher Biomasse. In einer anderen Arbeit wurde zum einen die Anzahl der Cohesine zwischen den verschiedenen cipA Genen von BC1, NB2 und ATCC 27405 und zum Anderen die Aminosäuresequenz verschiedener CelR Exoglucanasen verglichen. Allerdings könnte bereits aus diesen Ergebnissen abgeleitet werden, dass die Aminosäure- oder DNA-Sequenz cellulosomaler Komponenten nicht geeignet ist, die cellulolytische Effizienz verschiedener *C. thermocellum* Isolate zu vergleichen oder gar vorherzusagen (Koeck et al. 2015a).

Solche Stämme mit hohem cellulolytischen Potential scheinen sehr gut an die Bedingungen in Biogasanlagen angepasst zu sein und könnten als gezieltes Inokulum in Biogas-Fermentern zum Einsatz kommen, um den Lignocelluloseabbau zu verbessern. Vorangegangene Untersuchungen belegen, dass durch gezielte Inokulation mit einem besonders effizient hydrolysierenden Bakterienstamm die Abbaurate und der Biogasertrag verbessert werden können (Köllmeier 2013). Um das Inokulum optimal zu etablieren und so die Biogasausbeute zu erhöhen, ist allerdings eine kontinuierliche Bakterienzugabe der einmaligen Inokulation vorzuziehen (Schmack and Reuter 2010). Da in dieser Arbeit vor allem *C. thermocellum* aus fast jeder Biogasanlage angereichert und isoliert werden konnte, scheint diese Spezies eine wichtige Rolle bei Celluloseabbau während der hydrolytischen Stufe in Biogasanlagen zu spielen. Diese Aussage wird durch die Metagenomanalyse unterstützt (siehe nächsten Abschnitt).

4.2 ISOLATE VS. METAGENOMANALYSE

Für ein vertieftes Verständnis der Hydrolyse lignocellulosereicher Biomasse, ist es wichtig, detaillierte Kenntnisse über cellulolytische Bakterien und deren Anforderungen zu gewinnen, um damit eine Effizienzsteigerung der Biogasproduktion zu erreichen. Mehrere Forschungsarbeiten beschäftigten sich mit der Bakterienzusammensetzung in Biogasanlagen; trotzdem muss die Biogas-Mikrobiologie im Allgemeinen noch als „black box“ bezeichnet werden (Cirne et al. 2007, Ziganshin et al. 2013). Metagenomanalysen liefern zwar einen Überblick über die taxonomische Zusammensetzung, können aber meist das Artspektrum aufgrund von fehlenden Referenzdaten nicht bis in die Tiefe auflösen. Da der überwiegende Anteil der am Biomasseabbau beteiligten Mikroben-Gemeinschaft aus dem Phylum Firmicutes stammt, sind die Einzelsequenzen zwar verschiedenen Ursprungs aber durch die hohe Sequenzähnlichkeit schwer abzugrenzen und zuzuordnen. Zudem gibt der Nachweis einer Art allein noch keinen Aufschluss über deren Rolle bzw. Bedeutung bei der Substratumsetzung. Neueste Studien beschäftigen sich mit der Analyse des Metatranskriptoms von mikrobiellen Lebensgemeinschaften und vergleichen dieses mit der entsprechenden Metagenomanalyse derselben Anlage. Zakrzewski et al. (2012) haben gezeigt, dass von den abundantesten Spezies in der Gemeinschaft auch die meisten Transkripte stammen (Zakrzewski et al. 2012). Dies lässt darauf schließen, dass diese Bakterien auch maßgeblich am Abbau beteiligt sind. Allerdings repräsentierten nur 2,6 % der Metatranskriptomsequenzen mRNA Stücke und 90 % wurden als rRNA klassifiziert (Zakrzewski et al. 2012). Zudem kann wiederum aufgrund mangelnder Referenzgenome die Zuordnung der einzelnen Sequenzen nur unvollständig erfolgen (Wirth et al. 2012).

Um tatsächlich die Schlüsselorganismen für den Abbau zu identifizieren und deren Wege zur Substratumsetzung bis hin zum Methan aufzuklären, muss das gesamte Methodenspektrum ausgenutzt werden. Eine Kombination von klassischer Stammisolierung, Charakterisierung und anschließender Genomsequenzierung mit Metagenom- und Metatranskriptomanalysen erscheint als vielversprechender Ansatz. Die Genome neuer Isolate können direkt mit den Sequenzen der Metagenome und Metatranskriptome verglichen werden, um so direkt Aussagen über Abundanz und Abbauleistung des Bakteriums zu treffen. In dieser Arbeit wurde erstmals von einer thermophilen Biogasanlage eine Metagenomanalyse durchgeführt und aus derselben Anlage Isolate gewonnen.

Die Metagenomanalysen zeigten zunächst, dass die Gemeinschaft der Bakterien mit über 60 % der Sequenzen vom Phylum Firmicutes dominiert wird. Innerhalb der Firmicutes ist die Familie Clostridia am abundantesten und erst mit größerem Abstand gefolgt von den Thermotogae und Bacilli. Auch in früheren Metagenomanalysen thermophiler Fermenter gehörte die überwiegende Mehrheit der identifizierten Spezies zum Phylum Firmicutes bzw. zur Klasse Clostridia (Wirth et al. 2012; Krause et al. 2008). Wo eine Auflösung der Metagenomdatensätze bis auf Speziesebene

möglich ist, wurde gezeigt, dass meist *Clostridium thermocellum* die abundanteste cellulolytische Spezies ist (Wirth et al. 2012; Krause et al. 2008). Dies konnte auch für die in dieser Arbeit analysierte Biogasanlage bestätigt werden. *C. thermocellum* scheint einer der Hauptakteure beim Abbau Cellulose-haltiger Biomasse zu sein. Dabei muss aber beachtet werden, dass *C. thermocellum* mit mittlerweile 5 sequenzierten Genomen die beste Referenzdatenbank aller cellulolytischen Bakterien aufweist, was die Zuordnung der Metagenomsequenzen im Vergleich zu anderen nicht sequenzierten Spezies erst ermöglicht bzw. erleichtert und durch die erhöhte Wiedererkennung eine in Wirklichkeit nicht in diesem Maße vorhandene Abundanz hervorruft.

Auch alle aus den untersuchten Biogasanlagen gewonnenen Isolate stammen aus dem Phylum Firmicutes und lassen sich wiederum in zwei Klassen (Clostridia, Bacilli) einordnen. Von den 8 isolierten Hauptgattungen gehört jedoch nur *Bacillus* in die Klasse der Bacilli. Aus den anderen Phyla, die in der Metagenomanalyse präsent sind, wurde keine Spezies isoliert. Dies kann zum einen daran liegen, dass auch bei molekularbiologischen Methoden eine Verschiebung der Populationszusammensetzung auftreten kann. Alle Schritte der Probenaufarbeitung, von den Probenahmen bis zur PCR der extrahierten DNA, können als potentielle Ursache für eine solche Verschiebung in Frage kommen. Bei der Extraktion der DNA kann es zu einer selektiven Anreicherung der DNA leicht zu lysierender Zellen im DNA-Extrakt kommen. Der Zellaufschluss bei Gram-positiven Bakterien ist aufgrund der dicken Zellwand deutlich schwieriger als der Aufschluss von Gram-negativen Bakterien. Ein unzureichender Zellaufschluss wird also vorrangig die DNA leicht zu lysierender Zellen enthalten, während bei zu harschen Extraktionsbedingungen die DNA geschert werden kann, was zu Artefakten führt. Allerdings wurde mit verschiedensten DNA-Extraktionsmethoden aus der Biogasanlage H1 ein ähnliches taxonomisches Profil erhalten.

Zum anderen sind im anaeroben Milieu wie es in Biogasanlagen vorherrscht überwiegend Vertreter der Firmicutes für den Abbau Lignocellulosehaltiger Biomasse verantwortlich (Koeck et al. 2014a). Da gezielt auf Cellulosen und Hemicellulosen angereichert wurde, ist es daher nachvollziehbar, dass ausschließlich Vertreter der Firmicutes isoliert wurden. Der Anteil der Gattung *Clostridium* ist bei den Isolaten (70 %) mehr als doppelt so hoch wie in der Metagenomanalyse (29 %) (Abbildung 61).

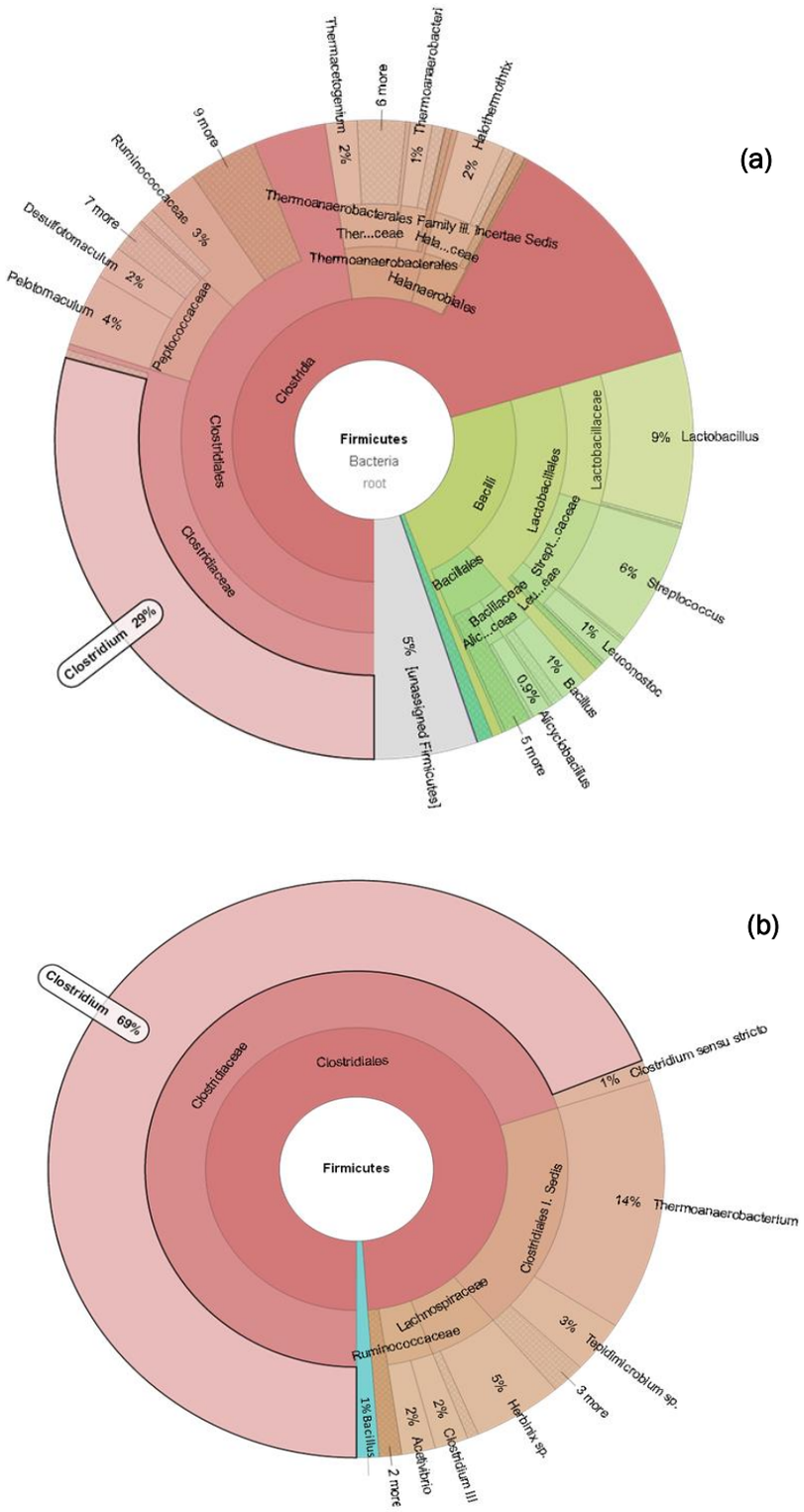


Abbildung 61: Kronaplotvergleich des Phylum Firmicutes zwischen der Metagenomanalyse der Biogasanlage H1 (a) und der Isolate aus derselben Biogasanlage (b)

Obwohl es auch innerhalb der Bacilli cellulolytische Spezies gibt, waren diese unter den isolierten Spezies nicht vertreten. Möglicherweise sind die in der Metagenomanalyse gefundenen Bacillus Sequenzen von Spezies, die primär an der Acido- und Acetogenese beteiligt sind und weniger an der Hydrolyse von cellulosehaltigen Substraten. Falls sich innerhalb der Bacilli-

Metagenomsequenzen ein cellulolytischer Vertreter befindet, wurde dieser mit den angewandten Isolierungsstrategien nicht erfasst. Vor allem die Ergebnisse von Metatranskriptomanalysen und Metaproteomanalysen anderer Biogasanlagen sprechen aber eher dafür, dass vor allen die Clostridien Glycosidhydrolasen bereitstellen und daher auch auf Cellulose vorwiegend angereichert und isoliert werden (Lu et al. 2014, Zakrzewski et al. 2012). Die Isolierungsergebnisse werden damit durch diese Experimente gestützt. Sowohl in der Metagenomanalyse als auch unter den späteren Isolaten ist *Clostridium thermocellum* die abundanteste Spezies. Es ist deshalb nicht erstaunlich (und unterstützt ebenfalls die Isolierungsergebnisse), dass auch unter den im Rahmen dieser Arbeit sequenzierten Genomen das Genom von C. *thermocellum* mit 0,8 % eine der besten Abdeckungen mit dem Metagenomdatensatz der Biogasanlage H1 aufweist (vgl. 3.1.7).

Sicherlich spielt dieses Bakterium wie bereits erwähnt eine zentrale Rolle beim Biomasseabbau. Die häufige Isolierung aus Biogasanlagen liegt aber nach mikrobiologischer Erfahrung in erster Linie nicht an einer hohen Abundanz in der Ausgangsprobe, sondern auch an der einfachen und guten Kultivierbarkeit unter Laborbedingungen. Die gute Kultivierbarkeit unter Laborbedingungen ist sicher auch der Grund, warum generell am häufigsten bereits bekannte Spezies isoliert wurden und selten neue Gattungen bzw. Arten. 91 % der thermophilen Isolate aus Biogasanlage H1 konnten bereits bekannten Spezies zugeordnet werden. Bei der mesophilen Biogasanlage T3 waren es 79 %. Mit Hilfe der einfachen Anreicherung und Reinigung (Strategie 1) wurden auch auf diversen Medien nur bereits bekannte Spezies angereichert. Diese scheinen unter Laborbedingungen selbst in geringen Ausgangsdichten andere schwieriger zu kultivierende Spezies schnell zu überwachsen. Erst das Arbeiten mit Verdünnungsreihen (Strategie 2 und 3) brachte den gewünschten Erfolg und lieferte bisher unbekannte Isolate. Die schlechte Kultivierbarkeit liegt also nicht nur an einer Spezies selbst, sondern an den unter den gewählten Kultivierungsbedingungen schneller wachsenden, konkurrierenden Arten, die sich noch im Ausgangsmaterial befinden. Schlechter anwachsende Bakterien mit längeren Generationszeiten auf den labortypischen Isolierungsmedien werden überwachsen und verdrängt. Für die erfolgreiche Isolierung neuer Spezies ist also weniger die Wahl des Mediums, sondern die Wahl der Isolierungsstrategie entscheidend. Als Kennzahl für die Biodiversität der Isolate wurde der Shannon-Wiener-Index herangezogen. Der höchste Wert (größte Diversität) zeigte sich bei den thermophilen Isolaten mit Strategie 2 und 3. Die mit Strategie 1 angereicherten Kulturen haben den niedrigsten Wert und somit die geringste Diversität – sie repräsentieren damit am wenigsten die vorhandene Population. Ca. 10 % der Metagenomsequenzen können keinem bisher bekannten Phylum zugeordnet werden. Zur vollständigen Interpretation der Daten ist es daher weiterhin unverzichtbar, die Isolierungsstrategien weiter zu entwickeln und neue Isolate zu gewinnen, die bisher unbekannten Phyla angehören.

Ein weiterer Aspekt, der untersucht wurde, war der Einfluss des Substrats auf die Bakterienzusammensetzung und die Isolate. Dabei zeigte sich, dass vor allem Mischsubstrate (z.B. Maissilage) durch die Vielfalt an unterschiedlichen organischen Polymeren eine höhere Diversität der Isolate fördern. Wie erwartet ist auch die taxonomische Diversität der mesophilen Isolate wesentlich höher als die Diversität der thermophilen Isolate. Im mesophilen Bereich wurden 12 Gattungen isoliert, im thermophilen Bereich dagegen nur 8 Gattungen. Dies ist jedoch nicht erstaunlich, da insgesamt die biologische Diversität in thermophilen Habitaten eingeschränkt ist.

4.3 „IN SACCO“ ANREICHERUNG VS. AMPLIKON

Für Amplikonsequenzierungen muss bei Mischkulturen das exponentielle Verhalten der PCR-abhängigen Amplifikation beachtet werden. Demzufolge verschiebt sich das Verhältnis zwischen den verschiedenen bakteriellen Spezies mit Zunahme der Zyklen zugunsten der häufiger vertretenen Spezies. Dadurch wird der Nachweis von seltenen bakteriellen Spezies schwieriger. Dies muss bei der Auswertung der Sequenzen durch eine ausreichende Probenzahl kompensiert werden. Um eine gleichmäßige Amplifikation aller DNA-Moleküle in einem Gemisch zu erhalten, müssen zusätzlich verschiedene Voraussetzungen erfüllt sein. Alle DNA-Moleküle müssen gleich gut mit den eingesetzten Primern hybridisieren, wobei die Bindungshäufigkeit und die Bindungsstärke bei allen Molekülen vergleichbar sein und die DNA-Polymerase bei allen DNA-Molekülen mit der gleichen Effizienz arbeiten muss (Suzuki et al. 1995). Da es sich bei dem gewählten Primer-System um bereits vielfach etablierte und eingesetzte Universalprimer handelte, welche in den konservierten Bereichen der bakteriellen 16S rRNA Gene binden, war dies gewährleistet (Loy et al. 2005). Die Amplifikation der DNA in der PCR kann vor allem bei Umweltproben durch Huminsäuren in der DNA-Probe behindert werden. Viele Polymerasen werden durch die Anwesenheit von Huminsäuren gehemmt (Tebbe and Vahjen 1993). Wenn versucht wird, die inhibierende Wirkung der Huminsäuren durch Verdünnung des DNA-Extrakts zu umgehen, besteht die Gefahr, dass niedrig konzentrierte DNA-Moleküle nicht mehr über die PCR nachgewiesen werden können.

Sowohl in der Metagenomanalyse der Biogasanlage Viersen als auch in der Amplikonsequenzierung der „in sacco“ Anreicherungen der thermophilen Laborfermenter der LfL sind neben der Hauptgattung *Clostridium* die Gattungen *Halocella* und *Tepidimicrobium* abundant. Aus den Gattungen *Clostridium* und *Tepidimicrobium* konnten auch diverse Isolate gewonnen werden. Vertreter der Gattung *Halocella* scheinen ebenfalls eine zentrale Rolle im Biomasseabbau thermophiler Biogasanlagen zu spielen, entziehen sich aber trotz mehrfacher Wiederholungen jeglichen Isolierungsversuchen. Bei der Erstellung der Klonbibliotheken aus den „in sacco“ Anreicherungen stammten über 80 % der sequenzierten Klone aus zwei definierten taxonomische Einheiten, innerhalb derer die Sequenzen >98 % Ähnlichkeit aufweisen. Diese

Einheiten haben mit 90-97 % die höchste 16S rRNA Gensequenzübereinstimmung mit cellulolytischen Bakterien. Vorangegangene Arbeiten haben gezeigt, dass ein taxonomisches Profil ohne vorheriger „in sacco“ Anreicherung wesentlich diverser ist und überwiegend nicht-cellulolytische Spezies umfasst (Köllmeier 2013). Somit war eine *in situ* Anreicherung auf Cellulose erfolgreich. Dafür spricht auch, dass aus den DNA-Proben mit spezifischen Primern diverse Cellulasen der Familie 48 amplifiziert werden konnten (Pechtl et al., in Vorbereitung). Die 16S rRNA Amplikonsequenzierungen der *in situ* Versuche zeigen aber, dass anscheinend wenige Arten beim Abbau cellulosehaltiger Biomasse dominieren. Die Diversität cellulolytischer Bakterien in den thermophilen Laborfermentern BM1T und BM2T ist nicht so groß wie erwartet.

Im Rahmen einer anderen Arbeit wurden die „in sacco“ Anreicherungen auch in mesophilen Laborfermentern durchgeführt und lieferten ein breiteres Artenspektrum mit geringerem Anteil an cellulolytischen Spezies (Pechtl et al., in Vorbereitung). Somit war die Anreicherung im thermophilen Fermenter aufgrund des geringeren Artenspektrums erfolgreicher als im mesophilen Fermenter. Bemerkenswert ist, dass keine der Amplikonsequenzen einer bisher bekannten Spezies zugeordnet werden konnte. Die 16S rRNA Sequenzähnlichkeit betrug maximal 97 %. Im Gegensatz zur Metagenomanalyse einer Probe aus dem Fermenter H1 konnte *C. thermocellum* in den „in sacco“ Anreicherungen der Laborfermenter nicht als Schlüsselbakterium identifiziert werden. Dies spricht wiederum dafür, dass einige für den Biomasseabbau in Biogasanlage wichtigen Bakteriengruppen bisher noch nicht erfasst und charakterisiert werden konnten.

Letztlich konnte wieder durch mehrere Verdünnungsreihen (Strategie 2) und mit Hilfe von Gärrest aus dem BM2T Fermenter ein cellulolytisches Isolat aus einer der taxonomischen Einheiten gewonnen werden. Dieses zeigte mit 94,36 % die höchste 16S rRNA Sequenzübereinstimmung mit dem *Clostridium thermocellum* Typstamm ATCC 27405 und gehört basierend auf diesen 16S rRNA Sequenzanalysen voraussichtlich zu einer bisher nicht beschriebenen Gattung innerhalb der Familie *Ruminococcaceae* (Kim et al. 2012). In einem ersten Isolierungsversuch wurde mit demselben Medium mit Strategie 1 nur *Clostridium clariflavum* isoliert. *Clostridium clariflavum* taucht allerdings in keiner der Amplikonbanken auf und ist daher vermutlich im Ausgangsmaterial nur in sehr geringen Abundanzen vorhanden. Hier wurde wiederum die Hypothese bestätigt, dass für die erfolgreiche Isolierung neuer Spezies weniger die Wahl des Mediums sondern die Wahl der Isolierungsstrategie entscheidend ist.

Die Gruppe *Halocella* sp. entzog sich weiterhin allen Isolierungsversuchen. Eine weitere Amplikonsequenzierung der Laboranreicherungen zeigte, dass diese Bakteriengruppe bereits nach der ersten Anreicherung unter Laborbedingungen nicht mehr wieder gefunden werden konnte. Die nächste Verwandte *Halocella cellulosilytica* wurde aus einer hypersalinen Lagune des Sivash See isoliert und benötigt, um zu wachsen 5 bis 20 % NaCl im Medium (Simankova et al.

1993). Dessen optimale Wachstumsbedingungen wurden im Labor nachgestellt, führten aber ebenfalls nicht zum Erfolg. Mit lediglich 96 % 16S rRNA Gensequenzübereinstimmung ist auch nicht gesichert, ob die taxonomische Gruppe aus der Biogasanlage unter denselben Bedingungen wie *Halocella cellulosilytica* wachsen kann.

Somit lässt sich die Hypothese aufstellen, dass eine wichtige Gruppe von Schlüsselbakterien für den Cellulose-Abbau in dieser Anlage mit den bisherigen Kultivierungsversuchen nicht erfasst werden konnte. Es ist auch nicht gesichert, dass durch die *in situ* Anreicherung alle cellulolytischen Populationen, die am Biomasseabbau beteiligt sind, erfasst wurden. Allein auf der Basis von Metagenomanalysen lässt sich dies bisher nicht nachweisen.

Der „in sacco“ Ansatz, welcher ursprünglich für Futtermittelanalysen in fistulierten Kühen entwickelt wurde (Mohamed and Chaudhry 2008), wurde im Rahmen dieser Arbeit erstmals in Biogasanlagen eingesetzt. Diese neue Methode war für die Isolierung neuer cellulolytischer Bakterien sehr erfolgreich und sollte deshalb weiter verfolgt werden. Hierbei werden die Bakterien in ihrem natürlichen Habitat, in dem alle nötigen Wachstumsfaktoren zur Verfügung stehen, bereits vorangereichert. Es ist durchaus vorstellbar, dass diese Methode auch auf andere unlösliche Modellsubstrate übertragen werden kann. Dieser Ansatz bietet eine Möglichkeit, gezielt Bakterien in ihrer natürlichen Umgebung auf dem Wunschsubstrat anzureichern und Kulturen zu gewinnen, die dieses Substrat besser abbauen als bisher bekannte Spezies. Von diesen kann wiederum auf Basis der Genomsequenz das für den Abbau verantwortliche Enzymsystem durch Klonierung und rekombinante Expression gewonnen, biochemisch charakterisiert und gegebenenfalls kommerziell genutzt werden. Letztlich bleibt für die physiologische Charakterisierung der Mikroorganismen deren Reinigung und Kultivierung unter Laborbedingungen extrem wichtig. Für „nicht kultivierbare Spezies“ kann hier nur auf alternative Methoden wie optische Pinzetten oder Mikromanipulatoren zurückgegriffen werden. Für syntrophe Organismen können zur Reinigung Diffusionskammern, zur Simulation der natürlichen Umwelt, eingesetzt werden (Kaeberlein et al. 2002). Da aber unter den cellulolytischen Bakterien kein direkter Syntrophismus bekannt ist, wurde in dieser Arbeit auf eine solche Methode verzichtet.

4.4 Cellulolytische Systeme vs. funktionale Einordnung der Isolate

Durch die phylogenetische Einordnung der Isolate konnten erste Informationen über eine mögliche Funktion der verschiedenen Arten im Reaktor gewonnen werden. Diese Informationen wurden mit den Ergebnissen physiologischer Experimente an den Reinkulturen ergänzt. Es bleibt allerdings noch die Frage zu klären, inwieweit die verschiedenen Isolate am Biomasseabbau beteiligt sind oder sich dabei ergänzen. Die anfängliche Hydrolyse lignocellulosehaltiger Biomasse wird vermutlich in erster Linie durch cellulolytische Clostridia bewerkstelligt, von denen auch

diverse Vertreter isoliert und charakterisiert wurden. Dabei wurde bereits in vorangegangenen Studien gezeigt, dass eine Kombination verschiedener Enzyme mit unterschiedlichen Substratspezifitäten effizienter im Substratabbau ist als die Einzelenzyme (Zverlov et al. 2010). Ein Beispiel ist der Abbau von Cellodextrinen zu Glucose durch extrazelluläre ß-Glucosidasen. Die meisten saccharolytischen Bakterien scheinen ein Aufnahmesystem für Cellodextrine zu besitzen, welche von intrazellulären Enzymen weiter verwertet werden, entweder durch Phosphorylasen und Bildung von Glucose-1-Phosphat (wie für *C. thermocellum* beschrieben) oder durch intrazelluläre ß-Glucosidasen (Lynd et al. 2002). Im Genom von *Bacillus thermoamylovorans* 1A1 wurden alle Gene für das Wachstum auf Cellobiose identifiziert (Koeck et al. 2014c). *B. thermoamylovorans* wächst auch unter aeroben Bedingungen und betreibt unter Sauerstoffmangel einen Gärungsstoffwechsel. Das Vorhandensein typischer Gärungsenzyme wie der Lactat-Dehydrogenase (EC 1.1.1.27) und der Alkohol-Dehydrogenase (EC 1.1.1.1) wurde ebenfalls im Genom von Isolat 1A1 nachgewiesen. Deshalb wird vermutet, dass 1A1 zum Teil die Cellobiose, welche von cellulolytischen Bakterien gebildet wird, weiter zu organischen Säuren verstoffwechselt und somit einen Teil der Ausgangsmetabolite für die methanogenen Archaea bereit stellt. Dafür spricht, das 1A1 zunächst nur in Kokultur mit *C. thermocellum* isoliert werden konnte. Andererseits konnte beim Abbau von Maissilage kein Synergismus zwischen diesen beiden Bakterien festgestellt werden. Im Gegensatz dazu wurde gezeigt, dass der Cellulose-Spezialist *C. thermocellum* und der Hemicellulose-Spezialist *C. stercorarium* Maissilage in Kokultur effizienter abbauen (Zverlov et al. 2010).

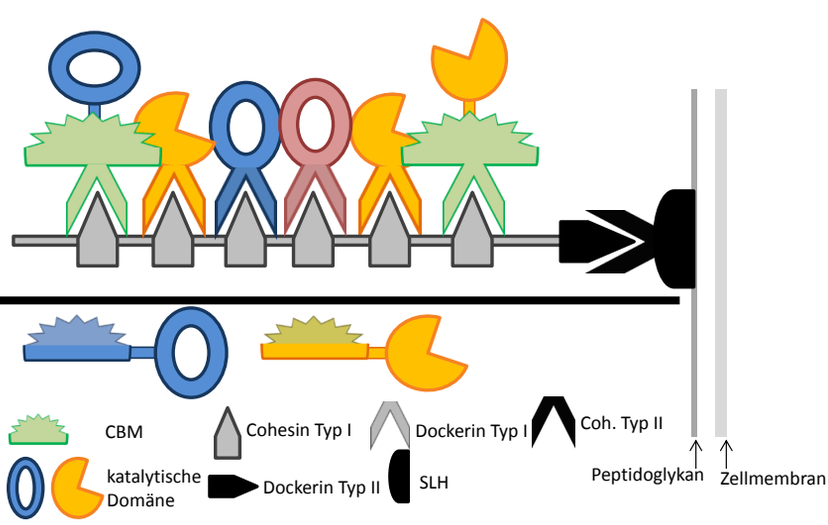
Bei einigen Isolaten wirft der Vergleich des Substratspektrums mit der Enzymausstattung weitere Fragen auf. So besitzen z.B. die Isolate SD1D und BC1 (*C. thermocellum*) Xylanasen, können aber Xylan nicht verstoffwechseln. Möglicherweise dienen die Xylanasen nur dazu, um die Cellulose von den Hemicellulosen freizulegen. Die freigesetzten Oligomere würden dann anderen Bakterien zur Verfügung stehen. Um dies zu bestätigen, müsste erst die Funktionalität und Substratspezifität der Enzyme geklärt werden.

Das beste Beispiel für die Kooperation verschiedener Enzyme beim Biomasseabbau ist der Aufschluss kristalliner Cellulose. Wie eingangs erwähnt, wurden in der Natur im anaeroben Bereich bisher drei dominante Typen mikrobieller Enzymsysteme zum Abbau von Cellulose gefunden. Dies sind entweder freie Einzelenzyme, multifunktionale Enzyme oder Multienzymkomplexe (Cellulosomen). Die Art des Systems kann bereits aus der Genominformation der betreffenden Bakterien abgeleitet werden (Himmel et al. 2010). So konnten bei den in dieser Arbeit isolierten Bakterien nur zwei der Systeme identifiziert werden. Die Isolate T3/55 (*Herbinix hemicellulosilytica*), DG5 und 37hoe/1 besitzen jeweils ungebundene Einzelenzyme. Die Isolate BC1 (*C. thermocellum*), SD1D (*Herbinix sp.*) und GGR1 besitzen dagegen ein makromolekulares Scaffoldin-Protein, CipA mit Typ I Cohesinen, was für das Vorhandensein eines Cellulosoms spricht. Vergleicht man die Systeme durch eine Bestimmung

der Enzymaktivität auf kristalliner Cellulose (Avicel) im Zellüberstand der verschiedenen Isolate, zeigen zudem, bis auf Isolat SD1D, die Spezies mit potentiellen Cellulosomen durch die räumliche Nähe der einzelnen Cellulasen eine signifikant höhere Abbaueffizienz (Tabelle 71). Das Isolat SD1D trotz Cellulosom auf Avicel eine cellulolytische Aktivität, die vergleichbar mit Einzelenzymsystemen ist, aufweist, könnte mehrere Gründe haben. Zum einen sind in der Genomsequenz von *C. thermocellum* BC1 74 Leserahmen mit einem Dockerin Modul enthalten und im Genom von SD1D dagegen nur 4 Leserahmen. Damit besitzt BC1 wesentlich mehr cellulosomale Komponenten und Variationen des Cellulosoms, die zum synergistischen Abbau von Cellulose beitragen. Zum anderen ist das Scaffoldinprotein von SD1D mit nur 140 kDa wesentlich kleiner als bei *C. thermocellum* BC1 und Isolat GGR1 und somit sind vermutlich auch die fertig assemblierte Cellulosomen kleiner und weniger divers.

Tabelle 71: Spez. Akt. (U/mg) der Zellüberstände ausgewählter Isolate auf PASC (Phosphoric acid swollen cellulose)

Isolat	Spez. Akt. (U/mg)
GGR1	0,18 +/- 0,034
BC1	0,56 +/- 0,04
SD1D	0,06 +/- 0,0012
37hoe/3	0,07 +/- 0,002
T3/55	0,02 +/- 0,0007



Bemerkenswert ist, dass meist die Isolate mit einem Cellulosom taxonomisch zur Familie *Ruminococcaceae* gehören und die Isolate mit freien Enzymen zur Familie *Lachnospiraceae*. Es gibt allerdings auch Vertreter der Familie *Ruminococcaceae*, welche nur freie Enzyme besitzen, z.B. *C. stercorarium* (Poehlein et al. 2013). Umgekehrt gab es jedoch bisher keinen bekannten Vertreter der Familie *Lachnospiraceae*, der ein Cellulosom besitzt (Koeck et al. 2014a). Mit Stamm SD1D wurde im Rahmen dieser Arbeit erstmals ein Vertreter der *Lachnospiraceae* mit funktionalem Cellulosom isoliert und charakterisiert. Innerhalb der Familie *Clostridiaceae* sind nur wenige echt cellulolytische Bakterienspezies bekannt.

Das Isolat *Bacillus thermoamylovorans* 1A1 besitzt zwar eine Cellulase der Familie GH9 und ist daher auf CMC aktiv. Für den Abbau von kristalliner Cellulose fehlt allerdings eine Exoglucanase (z.B. GH48). Jedoch reicht die Genominformation nicht aus um zu beurteilen, ob ein Bakterium Cellulose abbauen kann. Selbst die vermeintliche Ausstattung mit allen Genen für ein Cellulosom ist dafür keine Garantie. Im Genom von *C. acetobutylicum* ATCC 824 wurde z.B. ein großes

cellulosomales Gencluster identifiziert (Nolling et al. 2001). Sequenzanalysen haben gezeigt, dass dieses Cluster Gene für das Scaffoldin Protein CipA, die prozessive Endoglucanase Cel48A und einige Endoglucanasen der Familie 5 und 9 codiert. Der Stamm wächst zwar auf Cellobiose und amorpher Cellulose, kann aber kristalline Cellulose nicht verwerten. Obwohl Proteomanalysen gezeigt haben, dass wenige Kopien eines Cellulosoms exprimiert werden, scheinen diese nicht funktionell zu sein (Sabathe et al. 2002). *Clostridium bornimense* M2/40^T wurde kürzlich aus einem mesophilen zwei-phasigem Biogasreaktor isoliert (Hahnke et al. 2014a). Die vollständig sequenzierte und annotierte Genomsequenz offenbarte auch hier ein cellulosomales Gencluster auf einem zweiten Replikon (Chromid) (Hahnke et al. 2014b). Aber weder die Kultur noch die Enzyme im Kulturüberstand zeigten hydrolytische Aktivität auf kristalliner Cellulose (Avicel) oder amorpher Cellulose (PASC). Die rekombinant exprimierten Cellulasen Glycosidhydrolase CelK (GH9) und CelA (GH48) aus dem cellulosomalen Gencluster waren ebenfalls nicht auf Cellulose aber auf Xylanen aktiv (Tomazetto et al. 2015). Bisher ist nicht klar, ob M2/40 die Enzyme selbst überhaupt exprimieren und sekretieren kann, da Signalpeptide in den abgeleiteten Aminosäuresequenzen fehlen. Die Funktion und Funktionalität des Genclusters bleibt, bis weitere Proteom- und Transkriptomanalysen durchgeführt werden, unklar.

4.5 AUSBLICK

Die Identifizierung der wichtigsten Akteure für den Celluloseabbau im Biogasfermenter wird durch das begrenzte Wissen über echt cellulolytische („truly cellulolytic“) Bakterien beschränkt. Einige wichtige Cellulose abbauenden Bakterien in der Natur scheinen immer noch unentdeckt, insbesondere in Biogasfermentern. Die Isolierung und umfassende Charakterisierung neuer cellulolytischer Bakterien von anaerob aufgeschlossenem Pflanzenmaterial im Rahmen dieser Arbeit wird wesentlich dazu beitragen, Verfahren zur Überwachung und Steuerung der Hydrolyse in den Fermentern zu entwickeln. Das Wissen um die Schlüsselorganismen kann auch dazu beitragen, die optimalen Bedingungen für die Substrathydrolyse zu definieren und besonders geeignete Bakterienkulturen zur Inokulation von Biogasfermentern zu ermitteln, mit dem Ziel einer besseren Substratverwertung und einer höheren Raumzeitausbeute.

Die sequenzierten Genome bieten zusätzlich einen Zugang zu bisher nicht bekannten Enzymen mit neuen Eigenschaften für den Biomasseabbau. Dieses Wissen wird helfen, die kommerziell hergestellten Cellulasemischungen zu verbessern und damit die Biotechnologie der zweiten und dritten Generation, welche sich auf die Verwertung von lignocellulosehaltigem Material konzentriert, einen entscheidenden Schritt nach vorne zu bringen.

5 REFERENZEN

Abou-Zeid DM, Biebl H, Sproer C, Muller RJ (2004). *Propionispora hippei* sp nov., a novel Gram-negative, spore-forming anaerobe that produces propionic acid. *International journal of systematic and evolutionary microbiology* **54:** 951-954.

Adelsberger H, Hertel C, Glawischnig E, Zverlov VV, Schwarz WH (2004). Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the in vivo system from recombinant enzymes. *Microbiol-Sgm* **150:** 2257-2266.

Antoni D, Zverlov VV, Schwarz WH (2007). Biofuels from microbes. *Applied microbiology and biotechnology* **77:** 23-35.

Bae JW, Rhee SK, Park JR, Kim BC, Park YH (2005). Isolation of uncultivated anaerobic thermophiles from compost by supplementing cell extract of *Geobacillus toebii* in enrichment culture medium. *Extremophiles : life under extreme conditions* **9:** 477-485.

Bauer C, Korthals M, Gronauer A, Lebuhn M (2008). Methanogens in biogas production from renewable resources--a novel molecular population analysis approach. *Water science and technology : a journal of the International Association on Water Pollution Research* **58:** 1433-1439.

Bayer EA, Belaich JP, Shoham Y, Lamed R (2004). The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* **58:** 521-554.

Bayer EA, Setter E, Lamed R (2013). *The Prokaryotes: Lignocellulose-Decomposing Bacteria and Their Enzyme Systems*. Springer Verlag: Berlin.

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27(2):** 573-580.

Biddle A, Stewart L, Blanchard J, Leschine S (2013). Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. *Diversity* **5:** 627-640.

Bidet P, Barbut F, Lalande V, Burghoffer B, Petit J-C (1999). Development of a new PCR-ribotyping method for *Clostridium dificile* based on ribosomal RNA gene sequencing. *FEMS Microbiology Letters* **175:** 261-266.

Blom J, Albaum SP, Doppmeier D, Puhler A, Vorholter FJ, Zakrzewski M *et al* (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC bioinformatics* **10:** 154.

Boone DR, Chynoweth DP, Mah RA, Smith PH (1993). Ecology and microbiology of biogasification. *Biomass Bioenerg* **5:** 191-202.

Brazier JS (2001). Typing of *Clostridium difficile*. *Clin Microbiol Infect* **7:** 428-431.

Bronnenmeier K, Rucknagel KP, Staudenbauer WL (1991). Purification and properties of a novel type of exo-1,4-beta-glucanase (avicelase II) from the cellulolytic thermophile *Clostridium stercorarium*. *European journal of biochemistry / FEBS* **200:** 379-385.

Bronnenmeier K, Staudenbauer WL (1993). Molecular biology and genetics of substrate utilization in clostridia. *Biotechnology* **25:** 261-309.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37:** D233-238.

Caspi J, Irwin D, Lamed R, Li YC, Fierobe HP, Wilson DB *et al* (2008). Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: Comparative impact on cellulose-degrading activity. *Journal of biotechnology* **135:** 351-357.

Caspi J, Barak Y, Haimovitz R, Irwin D, Lamed R, Wilson DB *et al* (2009). Effect of Linker Length and Dockerin Position on Conversion of a *Thermobifida fusca* Endoglucanase to the Cellulosomal Mode. *Applied and environmental microbiology* **75:** 7335-7342.

Cirne DG, Lehtomaki A, Bjornsson L, Blackall LL (2007). Hydrolysis and microbial community analyses in two-stage anaerobic digestion of energy crops. *Journal of applied microbiology* **103:** 516-527.

Claassen PAM, van Lier JB, Contreras AML, van Niel EWJ, Sijtsma L, Stams AJM *et al* (1999). Utilisation of biomass for the supply of energy carriers. *Applied microbiology and biotechnology* **52:** 741-755.

Combet-Blanc Y, Ollivier B, Streicher C, Patel BK, Dwivedi PP, Pot B *et al* (1995). *Bacillus thermoamylovorans* sp. nov., a moderately thermophilic and amylolytic bacterium. *International journal of systematic bacteriology* **45:** 9-16.

Coorevits A, Logan NA, Dinsdale AE, Halket G, Scheldeman P, Heyndrickx M *et al* (2011). *Bacillus thermolactis* sp. nov., isolated from dairy farms, and emended description of *Bacillus thermoamylovorans. International journal of systematic and evolutionary microbiology* **61:** 1954-1961.

Davies G, Henrissat B (1995). Structures and mechanisms of glycosyl hydrolases. *Structure* **3:** 853-859.

Dice LR (1945). Measures of the amount of ecological association between species. *Ecology* **26:** 297-302.

Dion M, Fourage L, Hallet JN, Colas B (1999). Cloning and expression of a beta-glycosidase gene from *Thermus thermophilus*. Sequence and biochemical characterization of the encoded enzyme. *Glycoconjugate J* **16:** 27-37.

Dische Z (1962). Color reactions of carbohydrates. *Methods Carbohydr Chem* **1:** 477-512.

Divya D, Gopinath LR, Christy PM (2015). A review on current aspects and diverse prospects for enhancing biogas production in sustainable means. *Renew Sust Energ Rev* **42:** 690-699.

Eder B, Schulz H (2007). *Biogas Praxis; Grundlagen, Planung, Anlagenbau, Beispiele, Wirtschaftlichkeit*. Staufen: Freiburg.

Euzéby J (2010). List of new names and new combinations previously effectively, but not validly, published. *International journal of systematic and evolutionary microbiology* **60:** 1009-1010.

Fliegerova K, Mrazek J, Hoffmann K, Zabranska J, Voigt K (2010). Diversity of anaerobic fungi within cow manure determined by ITS1 analysis. *Folia microbiologica* **55:** 319-325.

FNR (2013). Leitfaden Biogas -Von der Gewinnung zur Nutzung. Fachagentur Nachwachsende Rohstoffe e. V.: Gülzow-Prüzen.

Freier D, Mothershed CP, Wiegel J (1988). Characterisation of *Clostridium thermocellum* JW20. *Applied and environmental microbiology* **54:** 204-211.

Görisch U, Helm M (2006). *Biogasanlagen*: Stuttgart.

Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE *et al* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PloS one* **6:** e17447.

Gupta P, Samant K, Sahu A (2012). Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential. *International journal of microbiology* **2012:** 578925.

Hahnke S, Striesow J, Elvert M, Mollar XP, Klocke M (2014a). *Clostridium bornimense* sp. nov., isolated from a mesophilic, two-phase, laboratory-scale biogas reactor. *International journal of systematic and evolutionary microbiology* **64:** 2792-2797.

Hahnke S, Wibberg D, Tomazetto G, Puhler A, Klocke M, Schlüter A (2014b). Whole genome sequence of *Clostridium bornimense* strain M2/40 isolated from a lab-scale mesophilic two-phase biogas reactor digesting maize silage and wheat straw. *Journal of biotechnology* **184:** 199-200.

He YL, Ding YF, Long YQ (1991). Two cellulolytic Clostridium species: *Clostridium cellulosi* sp. nov. and *Clostridium cellulofermentans* sp. nov. *International journal of systematic bacteriology* **41:** 306-309.

Henrissat B, Davies G (1997). Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology* **7:** 637-644.

Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA (2010). Microbial enzyme systems for biomass conversion: emerging paradigms. *Biofuels* **1:** 323-341.

Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH (2012). Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels* **5**.

Irwin D, Shin DH, Zhang S, Barr BK, Sakon J, Karplus PA *et al* (1998). Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* **180:** 1709-1714.

Janssen PH, Yates PS, Grinton BE, Taylor PM, Sait M (2002). Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Applied and environmental microbiology* **68:** 2391-2396.

Johnson EA, Madia A, Demain AL (1981). Chemically Defined Minimal Medium for Growth of the Anaerobic Cellulolytic Thermophile *Clostridium thermocellum*. *Applied and environmental microbiology* **41:** 1060-1062.

Johnson MJ, Thatcher E, Cox ME (1995). Techniques for controlling variability in gram staining of obligate anaerobes. *Journal of clinical microbiology* **33:** 755-758.

Kaeberlein T, Lewis K, Epstein SS (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296:** 1127-1129.

Kazamia E, Smith AG (2014). Assessing the environmental sustainability of biofuels. *Trends in plant science* **19:** 615-618.

Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H *et al* (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International journal of systematic and evolutionary microbiology* **62:** 716-721.

Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI *et al* (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PloS one* **8:** e47879.

Klemm D, Heublein B, Fink HP, Bohn A (2005). Cellulose: Fascinating biopolymer and sustainable raw material. *Angew Chem Int Edit* **44:** 3358-3393.

Koeck DE, Wibberg D, Koellmeier T, Blom J, Jaenicke S, Winkler A *et al* (2013). Draft genome sequence of the cellulolytic *Clostridium thermocellum* wild-type strain BC1 playing a role in cellulosic biomass degradation. *Journal of biotechnology* **168:** 62-63.

Koeck DE, Pechtl A, Zverlov VV, Schwarz WH (2014a). Genomics of cellulolytic bacteria. *Curr Opin Biotech* **29:** 171-183.

Koeck DE, Wibberg D, Maus I, Winkler A, Albersmeier A, Zverlov VV *et al* (2014b). Complete genome sequence of the cellulolytic thermophile *Ruminoclostridium cellulosi* wild-type strain DG5 isolated from a thermophilic biogas plant. *Journal of biotechnology* **188C:** 136-137.

Koeck DE, Wibberg D, Maus I, Winkler A, Albersmeier A, Zverlov VV *et al* (2014c). First draft genome sequence of the amylolytic *Bacillus thermoamylovoran*s wild-type strain 1A1 isolated from a thermophilic biogas plant. *Journal of biotechnology* **192 Pt A:** 154-155.

Koeck DE, Zverlov VV, Liebl W, Schwarz WH (2014d). Comparative genotyping of *Clostridium thermocellum* strains isolated from biogas plants: Genetic markers and characterization of cellulolytic potential. *Systematic and applied microbiology* **37:** 311-319.

Koeck DE, Koellmeier T, Zverlov VV, Liebl W, Schwarz WH (2015a). Differences in biomass degradation between newly isolated environmental strains of *Clostridium thermocellum* and heterogeneity in the size of the cellulosomal scaffoldin. *Systematic and applied microbiology*.

Koeck DE, Ludwig W, Wanner G, Zverlov VV, Liebl W, Schwarz WH (2015b). *Herbinix hemicellulosilytica*, gen. nov., sp. nov., a thermophilic cellulose-degrading bacterium isolated from a thermophilic biogas reactor. *International journal of systematic and evolutionary microbiology*.

Koeck DE, Maus I, Wibberg D, Winkler A, Zverlov VV, Liebl W *et al* (2015c). Draft genome sequence of *Herbinix hemicellulosilytica* T3/55, a new thermophilic cellulose degrading bacterium isolated from a thermophilic biogas reactor. *Journal of biotechnology*.

Köllmeier T (2013). Analyse spezialisierter Bakteriengemeinschaften und deren cellulolytische Hauptvertreter für die Hydrolyse in thermophilen Biogasanlagen. PhD thesis, Technische Universität München, München.

Krakat N, Westphal A, Satke K, Schmidt S, Scherer P (2010). The microcosm of a biogas fermenter: Comparison of moderate hyperthermophilic (60 degrees C) with thermophilic (55 degrees C) conditions. *Engineering in Life Sciences* **10:** 520-527.

Krause L, Diaz NN, Edwards RA, Gartemann K-H, Krömeke H, Neuwger H *et al* (2008). Taxonomic composition and gene content of a methane producing microbial community isolated from a biogas reactor. *J Biotech* **136:** 91-101.

Krauss J, Zverlov VV, Schwarz WH (2012). In vitro reconstitution of the complete *Clostridium thermocellum* cellulosome and synergistic activity on crystalline cellulose. *Applied and environmental microbiology* **78:** 4301-4307.

Krober M, Bekel T, Diaz NN, Goesmann A, Jaenicke S, Krause L *et al* (2009). Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. *Journal of biotechnology* **142:** 38-49.

Kruus K, Wang WK, Ching J, Wu JH (1995). Exoglucanase activities of the recombinant *Clostridium thermocellum* CelS, a major cellulosome component. *J Bacteriol* **177:** 1641-1644.

Kuberl A, Schneider J, Thallinger GG, Anderl I, Wibberg D, Hajek T *et al* (2011). High-quality genome sequence of Pichia pastoris CBS7435. *Journal of biotechnology* **154:** 312-320.

Lakaniemi AM, Hulatt CJ, Thomas DN, Tuovinen OH, Puhakka JA (2011). Biogenic hydrogen and methane production from *Chlorella vulgaris* and *Dunaliella tertiolecta* biomass. *Biotechnol Biofuels* **4:** 34.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490-495.

López Contreras AM (2003). *Utilization of lignocellulosic substrates by solvent-producing clostridia*: Wageningen, Netherlands.

Lottspeich F, Zorbas H (1998). *Bioanalytik*. Spektrum Akademischer Velag: Heidelberg.

Loy A, Schulz C, Lucker S, Schopfer-Wendels A, Stoecker K, Baranyi C *et al* (2005). 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". *Applied and environmental microbiology* **71:** 1373-1386.

Lu F, Bize A, Guillot A, Monnet V, Madigou C, Chapleur O *et al* (2014). Metaproteomics of cellulose methanisation under thermophilic conditions reveals a surprisingly high proteolytic activity. *The ISME journal* **8:** 88-102.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363-1371.

Ludwig W, Schleifer K-H, Whitman WB (2009). *Bergey's Manual® of Systematic Bacteriology. Revised road map to the phylum Firmicutes*, vol. 3. Springer: New York,USA.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66:** 739.

Madden RH (1983). Isolation and Characterization of *Clostridium stercorarium* Sp.Nov., Cellulolytic Thermophile. *International journal of systematic bacteriology* **33:** 837-840.

Marchandin H, Teyssier C, Campos J, Jean-Pierre H, Roger F, Gay B *et al* (2010). *Negativicoccus succinicivorans* gen. nov., sp nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov and Acidaminococcaceae fam. nov in the bacterial phylum Firmicutes. *International journal of systematic and evolutionary microbiology* **60:** 1271-1279.

Maus I, Wibberg D, Stantscheff R, Cibis K, Eikmeyer FG, Konig H *et al* (2013). Complete genome sequence of the hydrogenotrophic Archaeon *Methanobacterium* sp Mb1 isolated from a production-scale biogas plant. *Journal of biotechnology* **168:** 734-736.

Mc Bee R (1954). The characteristics of *Clostridium thermocellum*. *J Bacteriol* **67:** 505-506.

Medie FM, Davies GJ, Drancourt M, Henrissat B (2012). Genome analyses highlight the different biological roles of cellulases. *Nat Rev Microbiol* **10:** 227-U.

Meehan CJ, Beiko RG (2014). A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome biology and evolution* **6:** 703-713.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J *et al* (2003). GenDB - an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187-2195.

Miller GI (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31:** 426-428.

Mohamed R, Chaudhry AS (2008). Methods to study degradation of ruminant feeds. *Nutr Res Rev* **21:** 68-81.

Morais S, Heyman A, Barak Y, Caspi J, Wilson DB, Lamed R *et al* (2010). Enhanced cellulose degradation by nano-complexed enzymes: Synergism between a scaffold-linked exoglucanase and a free endoglucanase. *Journal of biotechnology* **147:** 205-211.

Morais S, Barak Y, Lamed R, Wilson DB, Xu Q, Himmel ME *et al* (2012). Paradigmatic status of an endo- and exoglucanase and its effect on crystalline cellulose degradation. *Biotechnology for Biofuels* **5**.

Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH *et al* (2011). Release LTPs104 of the All-Species Living Tree. *Systematic and applied microbiology* **34:** 169-170.

Ng TK, Zeikus JG (1981). Comparison of Extracellular Cellulase Activities of *Clostridium thermocellum* LQRI and *Trichoderma reesei* QM9414. *Applied and environmental microbiology* **42:** 231-240.

Nissila ME, Tahti HP, Rintala JA, Puhakka JA (2011). Effects of heat treatment on hydrogen production potential and microbial community of thermophilic compost enrichment cultures. *Bioresource technology* **102:** 4501-4506.

Noike T, Endo G, Chang JE, Yaguchi JI (1985). Characteristics of carbohydrate degradation and the rate-limiting step in anaerobic-digestion. *Biotechnol Bioeng* **27:** 1482–1489.

Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng QD, Gibson R *et al* (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology* **183:** 4823-4838.

O'Neill GL, Ogunsola FT, Brazier JS, Duerden BI (1996). Modification of a PCR ribotyping method for application as a routine typing scheme for *Clostridium dificile*. *Anaerobe* **2:** 205-209.

Parte AC (2014). LPSN-list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* **42:** D613-D616.

Perez J, Munoz-Dorado J, de la Rubia T, Martinez J (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *International microbiology : the official journal of the Spanish Society for Microbiology* **5:** 53-63.

Pham VH, Kim J (2012). Cultivation of unculturable soil bacteria. *Trends in biotechnology* **30:** 475-484.

Podosokorskaya OA, Bonch-Osmolovskaya EA, Beskorovaynyy AV, Toshchakov SV, Kolganova TV, Kublanov IV (2014). *Mobilitalea sibirica* gen. nov., sp nov., a halotolerant polysaccharide-degrading bacterium. *International journal of systematic and evolutionary microbiology* **64:** 2657-2661.

Poehlein A, Zverlov VV, Daniel R, Schwarz WH, Liebl W (2013). Complete Genome Sequence of *Clostridium stercorarium subsp. stercorarium* Strain DSM 8532, a Thermophilic Degrader of Plant Cell Wall Fibers. *Genome announcements* **1:** e0007313.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590-D596.

Rademacher A, Zakrzewski M, Schlüter A, Schonberg M, Szczepanowski R, Goesmann A *et al* (2012). Characterization of microbial biofilms in a thermophilic biogas system by high-throughput metagenome sequencing. *Fems Microbiol Ecol* **79:** 785-799.

Ratanakhanokchai K, Waeonukul R, Pason P, Tachaapaikoon C, Kyu KL, Sakka K *et al* (2013). *Biomass Now - Cultivation and Utilization*. InTech.

Reveneau C, Adams SE, Cotta MA, Morrison M (2003). Phenylacetic and phenylpropionic acids do not affect xylan degradation by *Ruminococcus albus*. *Applied and environmental microbiology* **69:** 6954-6958.

Riedel K, Ritter J, Bronnenmeier K (1997). Synergistic interaction of the *Clostridium stercorarium* cellulases Avicelase I (CelZ) and Avicelase II (CelY) in the degradation of microcrystalline cellulose. *Fems Microbiology Letters* **147:** 239-243.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al* (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *Plos Biol* **5:** 398-431.

Sabathe F, Belaich A, Soucaille P (2002). Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum. Fems Microbiology Letters* **217:** 15-22.

Sauer S, Freiwald A, Maier T, Kube M, Reinhardt R, Kostrzewa M *et al* (2008). Classification and identification of bacteria by mass spectrometry and computational analysis. *PloS one* **3:** e2843.

Sauer S, Kliem M (2010). Mass spectrometry tools for the classification and identification of bacteria. *Nature Reviews* **8:** 74-82.

Schlüter A, Bekel T, Diaz NN, Dondrup M, Eichenlaub R, Gartemann KH *et al* (2008). The metagenome of a biogas-producing microbial community of a production-scale biogas plant fermenter analysed by the 454-pyrosequencing technology. *Journal of biotechnology* **136:** 77-90.

Schmack D, Reuter M (2010). *Einsatz von Hilfsmitteln zur Steigerung der Effizienz und Stabilität der Biogasprosesses*.

Schoenborn L, Yates PS, Grinton BE, Hugenholtz P, Janssen PH (2004). Liquid serial dilution is inferior to solid media for isolation of cultures representative of the phylum-level diversity of soil bacteria. *Applied and environmental microbiology* **70:** 4363-4366.

Schwarz WH (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl Microbiol Biotechnol* **56:** 635-649.

Schwarz WH, Zverlov VV, Bahl H (2004). Extracellular glycosyl hydrolases from clostridia. *Advances in applied microbiology* **56:** 215-261.

Schwarz WH, Bello OJG, De Jong W, Leahy JJ, Oyaas K, Sorum L *et al* (2011). Strategic Energy Technology Plan. In: Baxter D, Tzimas E (eds). *Scientific Assessment in support of the Materials Roadmap enabling Low Carbon Energy Technologies*. Joint Research Centre – Institute for Energy and Transport Luxembourg.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ *et al* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**.

Simankova MV, Chernych NA, Osipov GA, Zavarzin GA (1993). *Halocella cellulolytica* Gen. Nov., Sp-Nov, a New Obligately Anaerobic, Halophilic, Cellulolytic Bacterium. *Systematic and applied microbiology* **16:** 385-389.

Spellerberg IF, Fedor PJ (2003). A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon-Wiener' Index. *Global Ecol Biogeogr* **12:** 177-179.

Stamatakis A (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688-2690.

Stams AJM, Vandijk JB, Dijkema C, Plugge CM (1993). Growth of Syntrophic Propionate-Oxidizing Bacteria with Fumarate in the Absence of Methanogenic Bacteria. *Applied and environmental microbiology* **59:** 1114-1119.

Stieglmeier M, Wirth R, Kminek G, Moissl-Eichinger C (2009). Cultivation of anaerobic and facultatively anaerobic bacteria from spacecraft-associated clean rooms. *Applied and environmental microbiology* **75:** 3484-3491.

Studier FW (2005). Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41:** 207-234.

Svetlichnyi VA, Svetlichnaya TP, Chernykh NA, Zavarzin GA (1990). *Anaerocellum thermophilum* Gen. Nov. Sp. Nov. an Extremely Thermophilic Cellulolytic Eubacterium Isolated from Hot-Springs in the Valley of Geysers. *Microbiology*+ **59:** 598-604.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30:** 2725-2729.

Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP *et al* (2009). Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11:** 2574-2584.

Tebbe CC, Vahjen W (1993). Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Applied and environmental microbiology* **59:** 2657-2665.

Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *International journal of systematic and evolutionary microbiology* **60:** 249-266.

Tomazetto G, Hahnke S, Koeck DE, Wibberg D, Maus I, Pühler A *et al* (2015). Complete genome analysis of *Clostridium bornimense* strain M2/40: A new acidogenic Clostridium species isolated from a mesophilic two-phase laboratory-scale biogas reactor. *Journal of biotechnology*.

Tschech A, Pfennig N (1984). Growth-Yield Increase Linked to Caffeate Reduction in Acetobacterium-Woodii. *Archives of microbiology* **137:** 163-167.

Ueno Y, Haruta S, Ishii M, Igarashi Y (2001). Microbial community in anaerobic hydrogen-producing microflora enriched from sludge compost. *Applied microbiology and biotechnology* **57:** 555-562.

Varel VH, Tanner RS, Woese CR (1995). *Clostridium herbivorans* Sp. Nov., a cellulolytic Anaerobe from the Pig Intestine. *International journal of systematic bacteriology* **45:** 490-494.

Vartoukian SR, Palmer RM, Wade WG (2010). Strategies for culture of 'unculturable' bacteria. *Fems Microbiology Letters* **309:** 1-7.

Wang JY, Stabnikova O, Ivanov V, Tay ST, Tay JH (2003). Intensive aerobic bioconversion of sewage sludge and food waste into fertiliser. *Waste management & research : the journal of the International Solid Wastes and Public Cleansing Association, ISWA* **21:** 405-415.

Wanner G, Formanek H, Galli D, Wirth R (1989). Localization of aggregation substances of *Enterococcus faecalis* after induction by sex pheromones. An ultrastructural comparison using immuno labelling, transmission and high resolution scanning electron microscopic techniques. *Archives of microbiology* **151:** 491-497.

Watanabe H, Tokuda G (2010). Cellulolytic systems in insects. *Annual review of entomology* **55:** 609-632.

Weiland P (2001). *Grundlagen der Methangärung - Biologie und Substrate*. VDI Verlag.

Weiland P (2006). Biomass Digestion in Agriculture: A Successful Pathway for the Energy Production andWaste Treatment in Germany. *Eng Life Sci* **6:** 302-309.

Wenzel W (2002). Mikrobiologische Charakterisierung eines Anaerobreaktors zur Behandlung von Rübenmelasseschlempe. PhD thesis, Technische Universität Berlin, Berlin.

White BA, Lamed R, Bayer EA, Flint HJ (2014). Biomass Utilization by Gut Microbiomes. *Annual Review of Microbiology, Vol 68* **68:** 279-296.

Wibberg D, Blom J, Jaenicke S, Kollin F, Rupp O, Scharf B *et al* (2011). Complete genome sequencing of *Agrobacterium* sp H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *Journal of biotechnology* **155:** 50-62.

Widdel F, Bak F (1992). *The Prokaryotes: a handbook on the biology of bacteria*: New York

Wietschel M, Arens M, Dötsch C, Herkel S, Krewitt W, Markewitz P *et al* (2010). *Energietechnologien 2050. Politikbericht*. Fraunhofer Verlag: Stuttgart.

Wirth R, Kovacs E, Maroti G, Bagi Z, Rakhely G, Kovacs KL (2012). Characterization of a biogas-producing microbial community by short-read next generation DNA sequencing. *Biotechnology for Biofuels* **5:** 41.

Wood TM (1988). Preparation of Crystalline, Amorphous, and Dyed Cellulase Substrates. *Method Enzymol* **160:** 19-25.

Yin YB, Mao XZ, Yang JC, Chen X, Mao FL, Xu Y (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40:** W445-W451.

Yutin N, Galperin MY (2013). A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol* **15:** 2631-2641.

Zaiß HN, Rupnik M, Kuijper EJ, Harmanus C, Michielsen D, Janssens K *et al* (2009). Typing *Clostridium difficile* strains based on tandem repeat sequences. *BMC Microbiology* **9:6**.

Zakrzewski M, Goesmann A, Jaenicke S, Junemann S, Eikmeyer F, Szczepanowski R *et al* (2012). Profiling of the metabolically active community from a production-scale biogas plant by means of high-throughput metatranscriptome sequencing. *Journal of biotechnology* **158:** 248-258.

Zakrzewski M, Bekel T, Ander C, Pühler A, Rupp O, Stoye J *et al* (2013). MetaSAMS--a novel software platform for taxonomic classification, functional annotation and comparative analysis of metagenome datasets. *Journal of biotechnology* **167:** 156-165.

Zhai Y, Tchieu J, Saier MH, Jr. (2002). A web-based Tree View (TV) program for the visualization of phylogenetic trees. *Journal of molecular microbiology and biotechnology* **4:** 69-70.

Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y *et al* (2003). *Gemmatimonas aurantiaca* gen. nov., sp. nov., a gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *International journal of systematic and evolutionary microbiology* **53:** 1155-1163.

Zhang K, Chen X, Schwarz WH, Li F (2014). Synergism of glycoside hydrolase secretomes from two thermophilic bacteria cocultivated on lignocellulose. *Applied and environmental microbiology* **80:** 2592-2601.

Ziganshin AM, Liebetrau J, Proter J, Kleinsteuber S (2013). Microbial community structure and dynamics during anaerobic digestion of various agricultural waste materials. *Applied microbiology and biotechnology* **97:** 5161-5174.

Zverlov V, Mahr S, Riedel K, Bronnenmeier K (1998). Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology+* **144 (Pt 2):** 457-465.

Zverlov V, Klupp M, Krauss J, Schwarz WH (2008). Mutations in the Scaffoldin Gene, cipA, of *Clostridium thermocellum* with Impaired Cellulosome Formation and Cellulose Hydrolysis: Insertions of a New Transposable Element, IS1447, and Implications for Cellulase Synergism on Crystalline Cellulose. *Journal of Bacteriology* **190:** 4321–4327.

Zverlov VV, Velikodvorskaya GA, Schwarz WH (2003). Two new cellulosome components encoded downstream of cell in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP. *Microbiology+* **149:** 515-524.

Zverlov VV, Kellermann J, Schwarz WH (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* **5:** 3646-3653.

Zverlov VV, Schwarz WH (2008). Bacterial cellulose hydrolysis in anaerobic environmental subsystems, *Clostridium thermocellum* and *Clostridium stercorarium*, thermophilic plant-fiber degraders. *Annals of the New York Academy of Sciences* **1125:** 298-307.

Zverlov VV, Hiegl W, Köck DE, Kellermann J, Köllmeier T, Schwarz WH (2010). Hydrolytic bacteria in mesophilic and thermophilic degradation of plant biomass. *Engineering in Life Sciences* **10:** 528-536.

VIII. DANKSAGUNG

Als erstes möchte ich größte Dankbarkeit an Herrn Prof. Dr. rer. nat. Wolfgang Liebl richten, für die Möglichkeit meine Dissertation am Lehrstuhl für Mikrobiologie durchzuführen und für die zahlreichen und hilfreichen Ratschläge während der Bearbeitung.

Mein ganz besonderer Dank gilt Dr. Wolfgang Schwarz und Dr. Vladimir Zverlov, die mich herzlich in ihre Arbeitsgruppe aufgenommen haben und mir die Möglichkeit gegeben haben an einem aktuellen und interessanten Thema zu arbeiten. Vielen Dank für die engagierte Betreuung und die geduldige Beantwortung aller meiner Fragen.

Bei Dr. Wolfgang Ludwig bedanke ich mich ganz herzlich für die zahlreichen Hilfestellungen bei taxonomischen Fragestellungen.

Weiterhin möchte ich mich sehr bei Dr. Claudia Held, Dr. Benedikt Leis, Dr. Björn Andreeßen, Dr. Petra Kornberger, Alexander Pechtl, Jannis Bröker, Matthias Mechelke, Simon Heinze, Jonathan Herlet für die sehr gute Zusammenarbeit, das tolle Arbeitsklima und anregende Diskussionen bei dem einen oder anderen Feierabendbier bedanken.

Natürlich gilt mein Dank auch allen weiteren Mitarbeitern des Lehrstuhls, die mich bei meiner Arbeit unterstützt haben.

Den Studenten Marie von Wrisberg, Florian Krabichler und Kristina Scheitz danke für ihre Beiträge zu den Ergebnissen im Rahmen von Praktikums-, Bachelor- und Masterarbeiten.

Irena Maus, Dr. Andreas Schlüter und Daniel Wibberg vom CeBiTec Bielefeld danke ich ganz besonders für die Durchführung der Genomsequenzierung und der großen Unterstützung bei der bioinformatischen Auswertung. Außerdem danke ich dem Projektkoordinator Dr. Michael Klocke und allen anderen Projektpartnern für die gute Zusammenarbeit und der FNR für die Projektfinanzierung.

Dr. Carmen Fahn von Lehrstuhl für Tierernährung danke ich für die gute Zusammenarbeit und die Lieferung von Rumenflüssigkeit.

Dr. M. Lebuhn und Bernhard Munk von der LfL-AQU danke ich für die gute Zusammenarbeit und die Zurverfügungstellung einer nicht unbeträchtlichen Anzahl von Fermenterproben.

Größte Dankbarkeit gebührt auch meiner Familie und meinen Freunden, die mich durch alle Höhen und Tiefen der Promotion begleitet haben.

IX. ERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit eigenständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht.

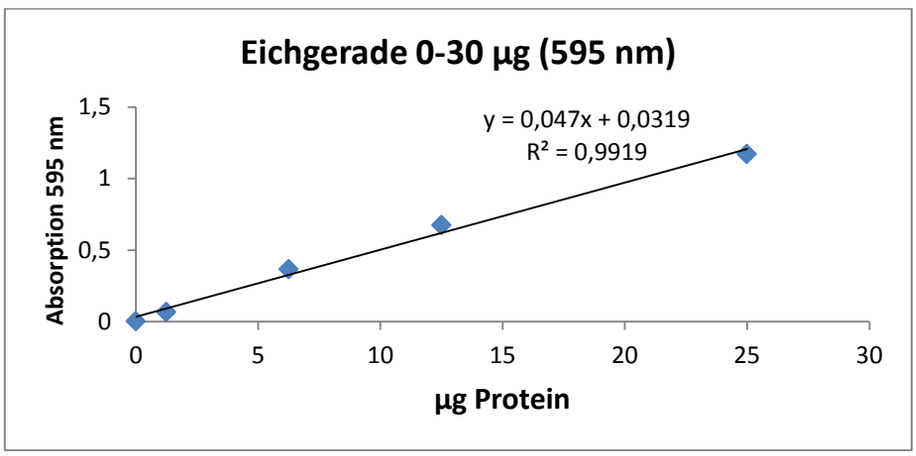
_____ _____

Ort, Datum Unterschrift

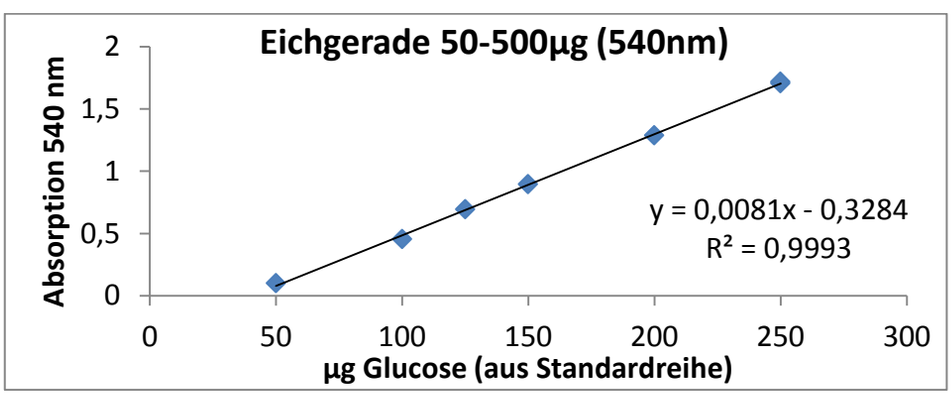
X. ANHANG

i. KALIBRIERGERADEN

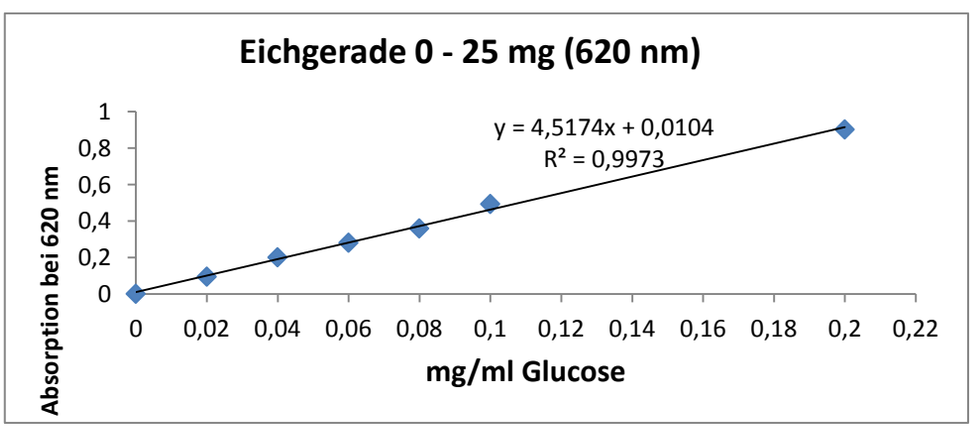
BRADFORD ASSAY



DNS ASSAY



ANTHRON ASSAY



ii. GENETISCH VERÄNDERTE ORGANISMEN UND VEKTOREN

Vektor	Nukleinsäure	GVO	Spender	Empfänger
pET24(+)	TUM355_438	pET24(+)-TUM355_438	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_2138	pET24(+)-TUM355_2138	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_2139	pET24(+)-TUM355_2139	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_1502	pET24(+)-TUM355_1502	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_1501	pET24(+)-TUM355_1501	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_1505	pET24(+)-TUM355_1505	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_810	pET24(+)-TUM355_810	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_1528	pET24(+)-TUM355_1528	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_1554	pET24(+)-TUM355_1554	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_2049	pET24(+)-TUM355_2049	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_2050	pET24(+)-TUM355_2050	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_269	pET24(+)-TUM355_269	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_698	pET24(+)-TUM355_698	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_905	pET24(+)-TUM355_905	*H. hemicellulosilytica*	*E. coli*
pET24(+)	TUM355_929	pET24(+)-TUM355_929	*H. hemicellulosilytica*	*E. coli*

iii. 16S rRNA Gensequenzen mesophile Biogasanlage T3

>Iso1/37
TGCAAGTCGAACGGTCTTTTCTATGGAAGCTTGCTTCCACTGAGAAGATAGTGGCGAACGGGTGAGTAACACGTGGGTAACCTGCCCATAAGAGGGGGA
TAACATCCGGAAACGGGTGCTAATACCGCATAGGTTTCTTGGTCGCATGATCGAGAAAGGAAAGACGGCCTTTGTGCTGTCGCTTATGGATGGACCCGCG
GCGTATTAGTTAGTTGGTGAGGTAACGGCTCACCAAGACGATGATACGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGGACTGAGACACGGCCCAA
ACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTAAAACTC
TGTTGTCAGAGAAGAACAAGTCGGAGAGTAACTGCTCCGGCCTTGACGGTATCTGACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATA
CGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTCCTTAAGTCTGATGTGAAAGCCCACGGCTCANCCGTGGAAGGT
CATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAGAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAGGAACACCAGTGGCGAANGC
GACTCTCTGGTCTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCANACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGT
GTTNGGAGGGTTTCCACCCTTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGG
GACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTTTGACAATCCTAGAGATAGGACTTTCC
CTTCGGGGACAAAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTT
GCCAGCATTCAGTTGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTAC
ACACGTGCTACAATGGATGGTACAACGAGCAGCAAGACCGCGAGGTCAAGCGAATCTCTTAAAGCCATTCTCAGTTCGGATTGCAGGCTGCAACTCGCCT
GCATGAAGCCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTA
ACACCCGAAGTCGGTGAGGTAACCTTTTTGGAGCCAGCTCGCNTAAG
>Iso2/37
GTCGAGCGATGAAGCCCTTCGGGGTGAATTAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTTATAGAGGGGGATAGCCTTCCGAAAGGAAG
ATTAATACCGCATGAAATGTTTTTATCGCATGGTAGAAACATCAAAGGAGTAATCCGCTATAAGATGGACCCGCGGCGCATTAGCTAGTTGGTGAGGTAAC
GGCCCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
GAATATTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGAATGAAGAAGGCCTTAGGGTTGTAAAGTTCTGTTTTCTGGGACGATAATGACGG
TACCAGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTACTGGGCGTAAAGGATGCGTA
GGCGGATGTTTAAGTGAGATGTGAAATACCCGAGCTCAACTTGGGTGCTGCATTTCAAACTGGATATCTAGAGTGCAGGAGAGGAAAGCGGAATTCCTAG
TGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGCTGANGCATGAAAGCGTGGGGAGCAA
ACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTAGGAGGTATCGACTCCTTCTGTGCCGCAGTTAACACAATAAGTATTCCG
CCTGGGGAGTACGATCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACC
TTACCTAGACTTGACATCTCCTGAATTACCCTTAATCGGGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCATTAGTTGCTACCATTAAGTTGAGCACTCTAGTGAGACTGCCACGGTTAACGTGGAGGAAGGT
GGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGTGAATACAAATAGATGCAATACCGCGAGGTGGAGCCAAACT
ATAAAATTCATCCCAGTTCGGATTGTAGTCTGAAACTCGACTACATGAAGCCGGAGTTGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGGTAACACCCGAAGTCCGTGAGGTAACCGTAAGGAGCCAGCGGCCGAAGGTGGGATTAGTGA
TCGGGGTGAAGTCGTAACA
>Iso4/37
CATGCAAGTCGAGCGGAGAATGCAGAAATGTTTACATGGAAGCGTTCTTAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTCAAAGTGGGGG
ATAGCCTTCCGAAAGGAAGATTAATACCGCATAAGCCTACAGTGCCGCATGGCACAGCAGGAAAAGGAGCAATCCGCTTTGAGATGGACCCGCGGCGCA
TTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGAACTGAGACACGGTCCAGACTC
CTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCGCGTGAGTGATGAAGGCCTTCGGGTCGTAAAGCTCTTT
GATCAGGGATGATAATGACAGTACCTGAAAAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTA
CTGGGCGTAAAGGATGCGTAGGTGGATACTTAAGTGGGATGTGAAATCCCCGGGCTCAACCCGGGAACTGCATTCCAAACTGGGTATCTAGAGTGCAGG
AGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACACTGAGG
CATGAAAGCGTGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGGTACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGT
CGTTAACACAATAAGTACCCCGCCTGGGGAGTACGGTCGCAAGACTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTA
ATTNGAAGCAACGNGAAGACCNTTACCTAGACTTGACNTNTCNTGAATTACCCTTAACCGGGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGT
TGTCGTCAGTTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTTTTAGTTGCTACCATTCAGTTGAGCACTTTAAAGAGACTGC
CCGGGTTAACCGGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTTTAGGGCTACACACGTGTTACAATGGTGAGAACAAAGAGACGGG
AGACCGCGAGGTGGAGCAAATTTCATAAAATTCATCCCAGTTCGGATTGCAGGCTGAAATTCGCCTGCATGAAGCCGGAGTTGTTAGTAATCGGGAATCA
GCATGTCGCGGTGAATACGTTCCCGGCCTTTGTACACACCGCCCGTCACACCATGAGAGTTGGCAATACCCGAAGTCCGTGGGGCAACCAGTAATGGAG
CCAGCGGCCGAAGGTAGGGTCA
>Iso5/37
ATTGGGCGTAAAGCNCGCGCAGNCGGCTGTTTAANTCTGGTNTTTAATNNGGGGNTCAACCCCGGGTNGCNCTGGAAACTNNNCAGCTGAGTGCAGA
AGAGGAGAGTGGAATTCCACGNGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGGCTGTAACTGACGNTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCNTGGTAGTCCACGCCGTAAACGATNNAATGCTAGGTGTTAGGGGTTTCGATACCCTTGGTG
CCGAAGTTAACACATTAAGCATTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGTGGAGTATGTG
GTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCNNNTGACCGGTACAGAGATGTACCTTTCCTTCGGGACAGGGGAGACAGGTGGT
GCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGACTTTAGTTGCCAGCAGGTAAGGCTGGGCACTCT
AGAGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGGTAC
AACGGGAAGCGAAGGAGCGATCTGGAGCGAATCCTATAAAGCCGGTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAATTGCTAGT
AATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCACGAGAGTTTACAACACCCGAAGTCGGTGAGGTAA
CCGCAAGGAGCCAGCCGCCGAAGGTGGGGTAGANGAT
>Iso1/45
GAGCTTGCTTCTTCTTAACTTAGTGGCGGACGGGTGAGTAACGCGTGAGTAACCTGCCTTTCAGAGGGGGATAACGTCTGGAAACGGACGCTAATACCG
CATGACATTTCTGTGCCTCTTGACATAGAAATCAAAGGAGCAATCCGCTGAGAGATGGATTCGCGTCCGATTATCTAGTTGGTGAGATAAAGGCCCACCAA
GGCGACGATCGGTAGCCGGACTGAGAGGTTGAACGGCCACATTGGGACTGATACACGGCCCAGACTCCTACGGGAGGCAGCNNTGGGGGATATTGTT
CNATGGAGGAAACTCTGATGCATCTTCGCCGCGTGAGGGAATAAGGTCTTCGGATTGTAAACCTTTGTCCTTGGTGAAGAAAAGAATGACTGTNTTTTTG
ANGAAGCTCCGGCTAACTACGTGCCAGCATCCGCGGTAATACNTAGGGAGCAAGCGTTGTCCGGATTTNNTGGGTGTAAAGGGTGCGTAGGCGGCTCT
GCAAGTCTNGCGTGAAAATCATGGGCTTAACCCATTGNATTGCGTTTGAAACTGTGGAGCTTGAGTGANNTANANGTAGGCGGAATTCCCGGTGNAGCG
NTGAAATGCGTAGAGNTCGGGGAGGAACACCCAGTGGNGAAGGCGGCTTANTGGGCTTTAACNGACGCTGAGGCANGAAAGCATGGGTAGCAAAACA
NGATTAGATACCCTTGGTAGTCCATGCCGTAAAACGANGATTANTAGGTGTGGGGGGGTCTGACCCCTTCCGTGCCGGGAGTTAACACAATAAGTANTC
CACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGA
ACCTTACCAGGTCTTGACATCCAACTAACGAAGCAGAGATGCATTAGGTGCCCTTCGGGGAAAGTTGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTG
TCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTGATTAGTTGCTACGCAAGAGCACTCTAATCAGACTGCCGTTGACAAAACGGAGGAA
GGTGGGGACGACGTCAAATCATCATGCCCTTTATGACCTGGGCTACACACGTACTACAATGGCTGTTAACAGAGGGAGGCAAGCCCGCGAGGGGGAGC
AAAACCCTAAAAACAGTCTCAGTTCGGATCGCAGGCTGAAACCCGCCTGCGTGAAGTTGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATAC
GTTCCCGGNCCTTGTACACACCGCCCGTCACTCCATTGGAGCTGGCAACNTCTGAAGTCAACCTCCTTTCTGTCTCACGTCATTTANGGTTCGCTTCAATT
>Iso2/45

GATGTGAAATCTCACGGCTTAACCGTGGTAAAGCATTTGAAACTGTGAAGGCTTGAGTNCAGGAGAGGAGAGTGGAATTCNTAGTGTAGCGGTGAAATG
CGTAGATTATTAGGAGGAATACCAGTGGCGAANGGCGACTNTNTGGACTGTANCTGACGCTGAGGCACGAAAAGCGTGGGGAGCAAAACAGGATTAGA
TACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTCGGGGGGGTCAAACCTCGGTGCCGCAGTTAACACATTAAGCACTCCNCCTGGGGAGTAC
GGTGGCAACACTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTT
GACATCCCCCGGACGGAACTAGAGATAGTTCATTCCCTTCGGGGACTGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGG
GTTAAGTCCCGCAACGAGCGCAACCCTTGTCTTTAGTTGCAAACAGTCAGGCTGTGCACTCTAAAGAGACTGCCGATGACAAATCGGAGGAAGGTGGGG
ATGACGTCAAATCATCATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAACGGGCAGCCACTTAGTGATAAGGAGCTAATCCCACAA
AACCGATCCCAGTTCGGATTGTAGGCTGCAACCCGCCTACATGAAGCCGGAGTTGCTAGTAATCGCGGATCAGAATGCCGCGGTGAATGCGTTCCCGGG
TCTTGTACACACCGCCCGTCACTCCATGGGAGTTGACAATACCTGAAGTTGGTGAGCTATCTGNATGAAGTCAGCT
>Iso4/45
TTGTCCNNNAATTATTGGGCGTAAAGCNGCGNNCANGGCGGCTGTTTTAAGTCTGGTGNTTTANTCCTGGGGNTCAANCCNNNGGTCGCACTGGAAAC
TGGNCAGCTTGAGNGCAGNAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACNCCAGTGGGCGAAGGCGACTCTC
TGGGCTGTAACTGACGCTGAGGCGCGAAGCGTGGGGAGCAANCAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATGCTAGGTGTTAGGG
GTTTCGATACCCTTGGTGCCGAAGTTAACACATTAAGCATTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGACCCGCAC
AAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCNTNTGACCGGTACAGAGATGTACCTTTCCTTCGGGAC
AGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGACTTTAGTTGCCAGCAGG
TAAGGCTGGGCACTCTAGAGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTA
CTACAATGGCCGGTACAACGGGAAGCGAAGGAGCGATCTGGAGCGAATCCTTTAAAGCCGGTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGA
AGTCGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCACGAGAGTTTACAACACCCG
AAGTCGGTGAGGTAACCGCAAGGAGCCAGCCGCCGAAGGTGGGGTAGATGATTGGGGNGAAGTCGTAACAA
>Iso5/45
TTAAGTCTGGTGTTTAATCCTGGGNCTCAACCCCNGGNTCGCACTGGAAACTGGGCAGCTTGANTGCAGAAGAGGAGAGTGGAANTTCCACGTGTAGC
GGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTCTCTGGGCTGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGG
ATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATGCTAGGTGTTAGGGGTTTCGATACCCTTGGTGCCGAAGTTAACACATTAAGCATTCCGCCTGG
GGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC
AGGTCTTGACATCCNNCTGACCGGTACAGAGATGTACCTTTCCTTCGGGACAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATG
TTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGACTTTAGTTGCCAGCAGGTAAGGTTGGGCACTCTAGAGTGACTGCCGGTGACAAACCGGAGGAAGG
TGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGGTACAACGGGAAGCGAAGGAGCGATCTGGAGCGAAT
CCTATAAAGCCGGTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTT
CCCGGGTCTTGTACACACCGCCCGTCACACCACGAGAGTTTACAACACCCGAAGTCGGTGAGGTAACCGCAAGGAGCCAGCCGCCGAAGGTGGGGTAG
ATGATTGGGG
>Iso6/45
TGTAAACTGACGTTGAGGCGCGAAAGCGTGGGGAGCAAACACGATTAGATCCCNTGGTAGTCCACGCCGTAAANNANGAATTNTAAGTGTTAGGGGTTT
CGATACCCTTGGTGACAGAAGTTAACAACATTAAGCATTCCGCNTGGGGAGTACGNCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGACCCGCACAA
GCANTGGAGTATGTGGTTTAATTNGAAGCAACGCGAAGAACCTTTCCAGGTCTTGACATCCCCCTGACCGGTACAGAGATGTACCTTTCCTTCGGGACAG
GGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGNGAGATTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGACTTTAGTTGCCAGCAGGTAA
GGNTGGGCACTNTAGAGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTA
CAATGGCCGGTACAACGGGAAGCGAAGGAGCGATCTGGAGCGAATCCTATAAAGCCGGTNTCAGTTCGGAAGCAGGCTCCAATTCGCNTGCATGTAGTT
GGAAATCATAGTTATTGTGGATCACCATNTCGCGGTCAACTTCNNNCGGATTCACCTCCATTTNCGATCCCCTCCAATTGGATCACATCCTNTGTGATCAC
CTCCTTTATGGATCACGTACTTTAAGGTTCGCGTCA
>4/1/37
TGCAAGTCGAGCGAGAGACTTCGGTCTCTAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTCATAGAGGGGGATAGCCTCCCGAAAGGGAGA
TTAATACCGCATAAGAAGTAGATACCGCATGGTATCAGCTTTAAAGGAGAAATCCGCTATGAGATGGACCCGCGGCGCATTAGCTAGTTGGTGAGGTAAA
GGCTCACCAAGGCGACGATGCGTAGCCGGCCTGAGAGGGCGAACGGCCACATTGGAACTGAGAGACGGTCCAGACTCCTACGGGAGGCAGCAGTGGG
GAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAAGGAAGAAGGCCTTCGGGTTGTAAACTTCTGTTAATGGGGACGATAATGACGG
TACCCATAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGATGCGTA
GGCGGATAATTAAGTGAGATGTGAAATACCCGGGCTTAACCTGGGGGGTGCATTTCAAACTGGTTGTCTAGAGTGCAGGAGAGGAAAGTGGAATTTCTAG
TGTAGCGGTGAAATGCGTAGATATTAGGAAGAACACCAGTGGCGAAGGCGACTTTCTGGACTGTAACTGACGCTGAGGCATGAAAGCGTGGGGAGCAAA
CAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTGGGAGTTGTCATGACTTCCGTGCCGCCGTAAACACATTAAGTATTCCGC
CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCT
TACCTGGACTTGACATCTCCTGAATAAATCTGTAATGGATGAAGCTCTTCGGAGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATAGTTAGTTGCCAGCACGTTATGGTGGGCACTCTAGCGAGACTGCCGGGGTTAACTCGGAGGAA
GGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCAGGGCGACACACGTGCTACAATGGCGAGAACAAAGAGAAGCAAGACCGTGAGGTGGAGCAA
AACTGATAAAACTCGTCCCAGTTCGGATTGTAGGCTGAAACTCGCCTACATGAAGCCGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACG
TTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTGGCAATACCCGAAGTCCGTGAGTTAACCGAAAGGGGACAGCGGCCGAAGGTAGG
>4/237
TGCAGTCGAGCGAAGCACTTTAAAGGAAGTTTTCGGATGGAATTTAAAGTGACTGAGCGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACA
GGGGGATAACAGTTGGAAACGACTGCTAATACCGCATAAGCGCACAGTGCTGCATAGCACAGTGTGAAAAACTCCGGTGGTATGAGATGGACCCGCGTC
TGATTAGGTAGTTGGTGGGGTAACGGCCCACCAAGCCGACGATCAGTAGCCGACCTGAGAGGGTGACCGGCCACATTGGGACTGAGACACGGCCCAAA
CTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGAAACCCTGATCCAGCGACGCCGCGTGAGTGAAGAAGTGTTTCGGCATGTAAAGCTC
TATCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGAT
TTACTGGGTGTAAAGGGAGCGTAGACGGCGATGCAAGTCTGGAGTGAAAGCCCGGGGCTCAACCCCGGGACTGCTTTGGAAACTGTGTTGCTAGAGTG
CAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACGTT
GAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCNGGTAGNCCACGCCGTAAACGATGAATACTAGGTGTTGGGGAGCAAAGCTCTTCGGTG
CCGCCGCTAACGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGG
TTTAATTCGAAGCAACGCGAAGAACCTTACCAAGTCTTGACATCGGAATGACCGTCCCGTAACTGGGGCTTCCCTTCGGGGCATTCCAGACAGGTGGTGC
ATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTAGTAGCCAGCAGTTCGGCTGGGCACTCTGGG
GAGACTGCCAGGGATAACCTGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGATTTGGGCTACACACGTGCTACAATGGCGTAAACAAA
GGGAAGCGAAGGAGTGATCCGGAGCAAATCTCAAAAATAACGTNTCAGTTCGGACTGTAGTCTGCAACTCGACTACATGAAGCTGGAATCGCCTAGTAAT
CGCGGATCAGAATNCCGCGGTGAATACGTTNCCGGGTCTTGTACACACCCCCCGTCACACCATGCGAGTTGGTAACGCCCGCAGTCAGTGACCCCACC
GCAAGGAGGGAGCTGNCGC
>2/237
GGCAGCGTAATATGAGCGAGAGAAAAGGAGCTCATCACGCAAAATAAAAGTGCCAACACCAAACAAGAGAAAATGAAAGCTGANTGTTGAGCGTTAGATA
GTGGCGAACGGGTGAGTAACGCGGAAACAACCTGCCTCATAGTTGGGGACAACACCGCTAAAGTGGTGCTAATACCNAATGAGATGTATCTTCCGCATG
GGAGGAACAAGAAAGGTGGCCTCTTAACAATGCTACCGCTAAGAGAGGGGTTTGCGTCTGATTAGCTGGTTGGTGGGGTAAGGGCCTACCAAGGCGAC
GATCANTAGCCGGTCTGAGAGGATGAACGGCCACACTGNGACTGAGACACGGCCCANACTCCTACNGGAGGCAGCGNTGGGGAATCTTCCGCAATGG
ACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAAGAAGGTTTTCGGATCGTANAGCTCTGTCTTCAAGGNAGAACGTGAGCAGTGTGAAAAATGCTGC
TTANTGACGGTACTTGANGACGAAGCCACGGCTAACTACNTGCCANCAGCCGCGGTAATACGTACGTGGNGAGCGTTGTCCAGAATTATTGAGCGTAAA

GGGCGCGTANGNNGATGGTTAAGTCCAACCTAAAAGTGCGGGGCTCAACCCCGTGANGGGTTGGAAACTAACCATCTTGAGTGCAGNAGAGGAAAGTG
GAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGACTGTGTCTGACGCTGAGGCGCGAAAGCCA
GGGGAGTGAACGGGATTAGATACCCCGGTAATCCTGGCCGTAAACGATGGGTACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACGCAA
TAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGTATGTGGTTTAATTCGACGCA
ACGCGAAGAACCTTACCAGGGCTTGACATTGAGTGAAAGGGCTAGAGATAGCCCGGTCTCCTTCTGGAGACACGAAAACAGGTGGTGCATGGCTGTCGT
CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTATGTTGCCAGCACGTAAAGGTGGGAACTCATAGGAGACTGCCG
CAGAGAATGCGGAGGAAGGCGGGGATGACGTCAAGTCATCATGCCCCATATGTCCTGGGCTACACACGTACTACAATGGGCTTAAATAGAGGGAAGCGA
GGCCGCGAGGCGGAGCGAACCCCAGAAACAAGCTCTCAGTTCGGATCGCAGGCTGCAACTCGCCTGCGTGAAGTCGGAATCGCTAGTAATCGCAGATC
AGCATGCTGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTGGAAACACCCGAAGCCGGTGGGGTAACCTGAAAAGG
AACTAGCCGTCGAAGGTGGGGCCGATGATTGGGG
>12/237
GAGGGGGGGCCGCACAAGCACAGGAGCATATGTTGTNTAATGANNCAACACGAAAANCNTTTCACAGACTTGACATCTCTCGAGATATTCTGTAATGGA
GGAAAACACATCGGGGTCACGAAGACAGNTGGGGCACGGTGGTGTTCAGCTCGTGTNGTGAGAGGTTGGGGTAAGTCTCGCACCGCGAGCACCCCTTA
TTTTTATTTGTTATCACTTATTTGAGCACTCTAGCGCGAGTGCCCGGGTTTTCCGGGGGGAAGGGGGGGAGGAGGTCAAATCTTCACGTCCCCTATATGT
AGGGGTACACACGTGCTACAAAGGTGGGGACACAGAGATGCACCCTCGCGAGAGAGAGCAAAAATATAAAACCGATATCACTTTGGATAGTAGGGGGAA
AATCTCNTACACGAAAAGGGAGATGTTAGTAATATCGAATCAGAAAATGGCGGGGAATACGTTNCCGGGCCCTGTGCACACCGCCCCTCACACCACGAG
AGATGGCACTATCCAAAATTTTTGAGATATCGCNTCAGCGAGGCAGCGACCTAAGGTAGGGTCAGCGATTGGGGNGAAGTCGTAA
>11/137
ATACTCGGGCTTAACCTGGNTGCTGCATNCCAAACTGGATATCTAGAGTGCAGGAGAGGAAAGTAGAATTCCTAGTGTAGCGNTGAAATGCGTAGAGATT
AGGAAGAATACCAGTGGCGAAGGCGACTTTCTGGACTGTAACTGACACTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
CGCCGTAAACGATGAATACTAGGTGTGGGGGTTGTCATGACCTCCGTGCCGCCGCAAACGCATTAAGTATTCCGCCTGGGGAGTACGGTCGCAAGATTA
AAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTAGACTTGACATCTCCTGAAT
TACTCTTAATCGAGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGC
GCAACCCTTATTGTTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGAGGAAGGTGGGGATGACGTCAAATCATCATGCC
CCTTATGTCTAGGGCTACACACGTGCTACAATGGCTGGTACAGAGAGATGCTAAACCGTGAGGTGGAGCCAAACTTTAAAACCAGTCTCAGTTCGGATTGT
AGGCTGAAACTCGCCTACATGAAGCTGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACA
CCATGAGAGTTGGCAATACCCAAAGTTCGTGAGCTAACGCGTAAGCGAGGCAGCGACCTAAGGTAGGGTCAGCGATTGGGGNGAAGTCGTA
>6/437
AACCTGGGCTTACCCGAGGGTATGCATCTGAAACTATATAGCTNNAGTCAAGGAGAGGAGAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATC
AGGAAGAATACCAGTGGCGAAAGCGGCTCTCTGGACTTGAACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCC
ACGCCGTAAACGATGGATACTAGGTGTGGGTGAGGAATCATCCGTGCCGGAGTTAACGCAATAAGTATCCCGCCTGGGGAGTACGGCCGCAAGGTTGAA
ACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCACAGAATC
GAGTAGAAATACTTGAGTGCCTCGTAAGAGGAGCTGTGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCCTGTTGGTAGTTACCAGCGTAAAGACGGGGACTCTACCGAGACTGCCGTGGATAACACGGAGGAAGGTGGGGATGACGTCAAATCA
TCATGCCCTTTATGCCCTGGGCTACACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGAGCGAATCCCAGAAAACAGGTCCCAG
TTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACAC
CGCCCGTCACACCACGAGAGTTTACAACACCCGAAGTCAGTGACCTAACCGAAAGGGAGGAGCTGCCGAAGGTGGGGTAAATGATTGGGGNGAAG
>12/237
NTATTAGGGGTATNGGTGTTCNCGNNNTNTGTGNCGCCGCTATAGCNCTNANTATTNTCCNTGGGGGAGAGCGGTGTCAAGATTNAAANTCTCNNAAA
TAGAGGGGGGGGCCGCACAAGCACAGGAGCATATGTTGTNTAATGANNCAACACGAAAANCNTTTCACAGACTTGACATCTCTCGAGATATTCTGTAATGG
AGGAAAACACATCGGGGTCACGAAGACAGNTGGGGCACGGTGGTGTTCAGCTCGTGTNGTGAGAGGTTGGGGTAAGTCTCGCACCGCGAGCACCCCTT
ATTTTTATTTGTTATCACTTATTTGAGCACTCTAGCGCGAGTGCCCGGGTTTTCCGGGGGGAAGGGGGGGAGGAGGTCAAATCTTCACGTCCCCTATATG
TAGGGGTACACACGTGCTACAAAGGTGGGGACACAGAGATGCACCCTCGCGAGAGAGAGCAAAAATATAAAACCGATATCACTTTGGATAGTAGGGGGA
AAATCTCNTACACGAAAAGGGAGATGTTAGTAATATCGAATCAGAAAATGGCGGGGAATACGTTNCCGGGCCCTGTGCACACCGCCCCTCACACCACGA
GAGATGGCACTATCCAAAATTTTTGAGATATCGCNTCAGCGAGGCAGCGACCTAAGGTAGGGTCAGCGATTGGGGNGAAGTCGTAAC>6/137TAGAGA
TNNGANGTTTCTTTNGNNCNTAGGGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
CTATTGTTAGTTGCCATCATTAAGTTGGGCACTCTAGCGAGACTGCCAGTAATAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGA
CCTGGGCTACACACGTGCTACAATGGTTGGTACAAAGAGTCGNGAGTNGGTGACGGCAAGCAAATCTCANAAAGCCAATCTCAGTTCGGANNNTAGGCT
GCAACTCGCCTACATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGTGAATACGTTCCCGGGTCNNNGTACACACCGCCCGTCNCACC
ACGAGAGTTTGTAACACCCGAAGTCGGTGAGGTAACCTNTAGGAGCCANCCGCNTAAGNTGGGATAGA
>1/137
GCCGCGGTAATACGTAGGGGGCAANCGTNNNCCGGATTNACTGGNTGTAAAGGGAGCGTAGACGGNACTGCAAGTCTGGAGTGAAAGCCCGGGGCT
CAACCCCGGGACTGCTTTGGAAACTGTGGTGCTAGAGTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACA
CCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACGTTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGAATACTAGGTGTTGGGGAGCAAAGCTCTTCGGTGCCGCCGCTAACGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGG
AATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAAGTCTTGACATCGGAATGACCGGGATGTAA
CGATCCCTTCCCTACGGGGCATTCCAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
TATCCTTAGTAGCCAGCAAGTCAAGTTGGGCACTCTGGGGAGACTGCCAGGGATAACCTGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTA
TGATTTGGGCTACACACGTGCTACAATGGCGTAAACAAAGGGNNNCAAAGGAGCGATCTGGAGCAAACCCCAAAAATAACGTCTCAGTTCGGATTGTAGT
CTGCAACTCGACTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGAATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCA
TGGGAGTTGGTAACGCCCGAAGTCAGTGACCCA
>8/137
AAGCCCGGGGCTCAACCCCGGGACTGCTTTGGAAACTGTGGTGCTAGAGTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATA
TTAGGAGGAANACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACGTTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT
CCACGCCGTAAACGATGAATACTAGGTGTTGGGGAGCAAAGCTCTTCGGTGCCGCCGCTAACGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGA
ATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAAGTCTTGACATCGGAAT
GACCGGGATGTAACGATCCCTTCCCTACGGGGCATTCCAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAA
CGAGCGCAACCCTTATCCTTAGTAGCCAGCAAGTCAAGTTGGGCACTCTGGGGAGACTGCCAGGGATAACCTGGAGGAAGGTGGGGATGACGTCAAATC
ATCATGCCCCTTATGATTTGGGCTACACACGTGCTACAATGGCGTAAACAAAGGGAGGCAAAGGAGCGATCTGGAGCAAACCCCAAAAATAACGTCTCAG
TTCGGATTGTAGTCTGCAACTCGACTACATNNAGCTGGAATCGCTAGTAATCGCGGATCAGAATGCCGCGGTGAATACGTTCCCGGGTCTNGTACACACC
GCCCGTCACACCATGGGAGTTG
>9/137
GTAGTNCATGCNGTAAACGANGAANTNCTAGGTGTTGGAGGGTNTCCGTCNNTTCAGTGCCGCAGCTAACGCATTAAGCATTCCGCCTGGGGAGTACG
NCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCNCGCACAAGCGGTGGAGCATGTGGTTTAATTNGAAGCAACGCGAAGAACCTTACCAGGTCTT
GACATCTTTTGATCACNTGAGAGATCAGGTTTCCCCTTCGGGGGCAAAATGACAGGTGGTGCATGGTTGTCGTCAGCTCGNGTCGTGAGATGTTGGGTT
AAGTCCCGCAACGAGCGCAACCCTTATGACTAGTTGCCAGCATTTAGTTGGGCACTCTAGTAAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGA
CGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGGTACAACGAGTTGCGAGACCGCGAGGTCAAGCTAATCTCTTAAAGC
CATTNTCAGTTCGGACTGTAGGCTGCAACTCGCCTACACGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACCCCGCGGTGAATACGTTCCCGGGCCT

TGTACACCCCGCCCGTCACACCATGAGAGTTTGTAACACCCGAAGCCGGTGGCGTAACCCTTTTAGGGAGCGAGCCGTCTAAGGTGGGACAAATGATTA
GGGNGAANTCGTAA>8/237GGGAAGCAAAGGCTCTTTCGGGTGCCCGCCCGCTAACGCCAATAAAGTATTCCCACCTGGGGGAGTACGTTNCCAAAG
AATGANACTCAAAAGGAATTGACGGGGACCCCGCACAAGCNGGTGGAGCCATGTGGTTTAATNNGAAGCAACGCGAAGAACCTTACCAAGTCTTGACAT
CGGAATGACCGGGATGTAACGATCCCTTCCCTACGGGGCATTCCAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTC
CCGCAACGAGCGCAACCCTTATCCTTAGTAGCCAGCAAGTCAAGTTGGGCACTCTGGGGAGACTGCCAGGGATAACCTGGAGGAAGGTGGGGATGACG
TCAAATCATCATGCCCCTTATGATTTGGGCTACACACGTGCTACAATGGCGTAAACAAAGGGAGGCAAAGGAGCGATCTGGAGCAAACCCCAAAAATAAC
GTCTCAGTTCGGATTGTAGTCTGNNNCTCGACTACATNAAGCTGGAATNGCTAGTAATCGCGGATCAGAATGCCGCGGTGAATACGTTCCCGGGTCTNG
TACACACCGCCCGTCACACCATGGGAGTTGGNNACGCCNCGAAGTCAGTGACCCCNACCGTAAGGAGGGAGCTGNNGAAGGCGGGACTGATAACTGG
>37hö/1
TGCAGTCGAACGAAGCACAAGGNAGAGCTTGCTCTTCCTTGTGACTGAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCATACACAGGGGGAT
AACAGCTGGAAACGGCTGCTAATACCGCATAACCCGATAGTACCGCATGGTACGGACGGAAAAGATTTATCGGTGTATGATGGACCCGCGTCTGATTAGC
TAGTTGGAGGGGTAACGGCCCACCAAGGCGACGATCAGTAGCCGACTTGAGAGAGTGATCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAAGCTCTATCAGCAG
GGAAGATAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGT
GTAAAGGGAGCGTAGGTGGTGCGGTAAGTCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCATTTGAAACTATCGCACTGGAGTGCAGGAGAGG
TAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCGA
AAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGCCCACAGGGCTTCGGTGCCGCAGCTA
ACGCAATAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGA
AGCAACGCGAAGAACCTTACCAAGCCTTGACATCCCGATGACCGTTCCGTAATGGGAACTTCTCTTCGGAGCATCGGTGACAGGTGGTGCATGGTTGTCG
TCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCTTCAGTAGCCAGCAATTTGGA
>37hö/2
TGCAGTCGAGCGGACCGACGGGAGCTTGCTCCCTTAGGTCAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTGTAAGACTGGGATAACTCCG
GGAAACCGGGGCTAATACCGGATGCTTGATTGAACCGCATGGTTCAATCATAAAAGGTGGCTTTTAGCTACCACTTGCAGATGGACCCGCGGCGCATTAG
CTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAACTCTGTTGTTA
GGGAAGAACAAGTACCGTTCGAATAGGGCGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGT
GGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCCCGGCTCAACCCGGGGAGGGTCATTGG
AAACTGGGGAACTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGACTC
TCTGGTCTGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAG
AGGGTTTCCGCCCTTTAGTGCTGCAGCAAACGCATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAACCCTAGAGATAGGGCTTCCCCTTCGGG
GGCAGAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCAGCA
TTCAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTG
CTACAATGGGCAGAACAAAGGGCAGCGAAGCCGCGAGGCTAAGCCAATCCCACAAATCTGTTCTCAGTTCGGATCGCAGTCTGCAACTCGACTGCGTGA
AGCTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCC
GAAGTCGGTGAGGTAACCTTTTGGAGCCAGCCGCCGAAGGTGGGACAGATGATTGGGGNGAAGTCGT
>37hö/3
GATCATGGGTCANANTTGGATCTTGGNGNGAGGAGAGGCGGACGGGTGNTTTGAGCGTGGGTAACCTGCCGTACACNGGGGGATAACAGCTGGAAAC
GGCTGCTAATACCGCATAACCCGATAGTACCGCATGGTACGGACGGAAAAGATTTATCGGTGTATGATGGACCCGCGTCTGATTAGCTAGTTGGAGGGGT
AACGGCCCACCAAGGCGACGATCAGTAGCCGACTTGAGAGAGTGATCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTG
GGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGATAATGAC
GGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCG
TAGGTGGTGCGGTAAGTCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCATTTGAAACTATCGCACTGGAGTGCAGGAGAGGTAAGTGGAATTCC
TAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCGAAAGCGTGGGGAG
CAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTNNTGTCGGGGCCCACAGGGCTTCGGTGCCGCAGCTAACGCAATAAGTATT
CCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGA
ACCTTACCAAGCCTTGACATCCCGATAGACCGTTCCGTAATGGGAACTTCTCTTCGGAGCATCGGTTGTCGTCAGCTCGTG
TGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCTTCAGTAGCCAGCAATTCGGATGGGCACTCTGGAGAGACTGCCAGGGATAACCTGGA
GGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGGCTTGGGCTACACACGTGCTACAATGACAGTTACAAAGTGAAGCTACCTCGCGAGAGCAAG
CAAATCACAGAAAGGCTGTCTCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCTGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATA
CGTTCCCGNNNTTTGTNCACNTTCNGGATCACCTCCATTGGGGTTCACATCCTTTGAGGATCACCTCCTTTCTGGATCACCTCCTTTCTGGATCACCTCC
>37hö/4
AAGCACAAGGNAGAGCTTGCTCTTCCTTGTGACTGAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCATACACAGGGGGATAACAGCTGGAAA
CGGCTGCTAATACCGCATAACCCGATAGTACCGCATGGTACGGACGGAAAAGATTTATCGGTGTATGATGGACCCGCGTCTGATTAGCTAGTTGGAGGG
GTAACGGCCCACCAAGGCGACGATCAGTAGCCGACTTGAGAGAGTGATCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAG
TGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTATGTAAAGCTCTATCAGCAGGGAAGATAATG
ACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAG
CGTAGGTGGTGCGGTAAGTCAGATGTGAAAGCCCGGGGCTCAACCCCGGGACTGCATTTGAAACTATCGCACTGGAGTGCAGGAGAGGTAAGTGGAAT
TCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCGAAAGCGTGGG
GAGCNNCAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTANGTGTCGGGGCCCACAGGGCTTCGGTGCCGCAGCTAACGCAATAAGT
ATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGA
AGAACCTTACCAAGCCTTGACATCCCGATGACCGTTCCGTAATGGGAACTTCTCTTCGGAGCATCGGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTG
TCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATCTTCAGTAGCCAGCAATTTGGCCGGGCACTCTGGAGAGACTGCCAGGGATAACCT
GGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGGCTTGGGCTACACACGTGCTACAATGACAGTTACAAAGTGAAGCTACCTCGCGAGAG
CAAGCAAATCACAGAAAGGCTGTCTCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCTGGAATCGCTAGTAATCGCGAATCAGAATGTCGCGGT
GAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGTCAGTGACCCAACCGCAAGGAGGGAGCTGCCGAAG
GTGGAGCCGGTAACTGGGG
>37hö/5
CGAAGCTAAGATTGCTGAAGTTTTCGGATGGACGCAGACTTAGCTTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCTTCAGTGGGGGACAA
CAGTTGGAAACGACTGCTAATACCGCATAAGACCACAGTGCCACATGGCACGGTGATCAAAGGATTAATTTTCGCTGAAGGATGGACTCGCGCCTGATTA
GCTAGATGGTGAGGTAACGGCCCACCATGGCAACGATCAGTAGCCGGACTGAGAGGTTGATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTA
CGGGAGGCAGCAGTGGGGGATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTTTTCGGATTGTAAACCTCTGTCTT
CGGGGACGAAACAAATGACGGTACTCGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATT
ACTGGGTGTAAAGGGAGCGCAGGCGGAACTGCAAGTTGGGTGTTTAATCTACGGGCTCAACCCGTATCAGCATTCAAAACTGTGGTTCTTGAGTAATGCA
GAGGTAGGCGGAATTCCCGGTGTAGCGGTGGAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCNTACTGGGCATCAACTGACGCTGAG
GCTCGAAAGCATGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGATGATTACTAGGTGTGGGGAGATTGACCCTCTCCGTGCCGCA
GTTAACACAATAAGTAATCCACCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAAT
TCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCAACGACCGCCTAAGAGATTAGGCTTTCCTTCGGGACGTTGAGACAGGTGGTGCATGGTTGT

CGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTACTAGTTACTACGCAAGAGGACTCTAGTGAGACTGCCGTTGAC
AAAACGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTTATGACCTGGGCTACACACGTACTACAATGGCTATTAACAGAGAGAAGCAAGACCGC
AAGGTGGAGCAAACCTCAGAAAAATAGTCTCAGTTCGGATTGGGGGCTGCAACCCGCCCCCATGAAGCCGGAATTGCTAGTAATCGCGGATCAGCATGC
CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCCGGGGGGACCCGAAGTCGGTAGTCTAACCGCAAGGAGGACGCTC
GCCGAAGGTAAAACTTGGTGATCGGG

iv. 16S rRNA Gensequenzen thermophile Biogasanlage H1

>XV1
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAACGGAGATGGTGGGATAGCTTGCTATTTTATCATCTTAGTGG
CGGACGGGTGAGTAACACGTGGGCAACCTGCCCTAAAGAGGGGGATAACACCGAGAAATCGGTGCTAATACCGCATGAAATTAGAGTTTCGCATGAAGC
TTTAATCAAAGGAGCAATCCGCTTTAGGATGGGCCCGCGTCTGATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCAGTAGCCGATCTGA
GAGGATGACCGGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCA
ACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTCTTCTGGGACGAAATATGACGGTACCAGAGGAGGAAGCCACGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGCGCGTAGGCGGCCTGTTAAGTCAGATGTGAAAACCCCGGG
CTCAACCTGGGGAGTGCATTTGAAACTGATGGGCTAGAGTGTCGGAGAGGGAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAA
CACCAGTGGCGAAGGCGGCTTTCTGGACGATAACTGACGCTGANGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGNTAGTCCACGCTGTA
AACGATGGATACTAGGTGTGGGAGGTATCGACCCCTTCCGTGCCGCAGTAAACGCAATAAGTATCCCGCCTGGGGAGTACGNCCGCAAGGTTGAAACTC
AAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGAGCTTGACATCTCCTGCATAGTCC
GTAATGGGATGAAGTCCTTCGGGACAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAA
CCCTTATACTTAGTTGCCATCATTAAGTTGGGCACTCTAAGTAGACTGCCGGAGACAATTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTA
TGTTCTGGGCTACACACGTGCTACAATGGGTATAACAACGGGAAGCAAGCCAGTGATGGTAAGCAAATCCCTAAAAAGTACTCCCAGTTCAGATTGTTCTC
TGAAACTCGAGAACATGAAGTCGGAGTTGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCAT
GGGAGTTGGAAGCACCCGAAGCCGGTGACTTAACCGCAAGGAGAGAGCCGTCGAAGGTGAAGCCAATGACTGGGGTGAAGTCGTAACAAGGTAGCCGT
ATCGGAAGGTGCGGCTGGATCACCTCCTTTCTG
>XV2
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGGATTTATGACGGATTCCTTCGGGATGAAGACATAA
ATCACCTAGCGGCGGACGGGTGAGTAACGCGTGAGGAACCTACCTCCTACAAGGGGATAGCCTCGGGAAACTGGGATTAATACCCTATGACACTTAAGG
GTCGCATGATCTTTAGGTTAAAGCGATATAGCGGTAGGAGATGGCCTCGCGTCCCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGG
TAGCCGGCCTGAGAGGGTGAACGGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGGAA
GCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAGCTCTGTCCTATGGGACGATAATGACGGTACCATAGGAGGAAGCCCCGG
CTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGTACGTAGGCGGCTAGGTAAGTCAGATG
TGAAATCCATGGGCTCAACCCATGTAAGCATTTGAAACTGCTTAGCTTGAGTTAAGGAGAGGAAAGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATA
TCAGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGTCTTAAACTGACGCTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGNTAGTN
CACGCNGTAAACGATGAGTGCTAGGTGTCGGGGGTTTAACCTTCGGTGCCGCAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGNNCGCAAGAAT
GAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTG
CCGGGACTAGAGATAGTCCCTTCCCTTCGGGGACGGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCTTGCCTTTAGTTGCCAGCACTTTAGGTGGGCACTCTAGAGGGACTGCCGATGACAAATCGGAGGAAGGTGGGGATGACGTCAAAT
CATCATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGGTATAACAACGGGAAGCAAGCCAGTGATGGTAAGCAAATCCCTAAAAAGTACTCCCAG
TTCAGATTGTTCTCTGAAACTCGAGAACATGAAGTCGGAGTTGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACC
GCCCGTCACACCATGGGAGTTGGAAGCACCCGAAGCCGGTGACTTAACCGCAAGGAGAGAGCCGTCGAAGGTGAAGCCAATGACTGGGGTGAAGTCGT
AACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTG
>MV1 *Clostridium tertium*
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGTGATTTCCTTCGGGAATGAACCTAGCGGCGGACG
GGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGGGAGATTAATACCGCATAATATTGAAGTTTCGCATGAAACTTCAATCA
AAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTG
ATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCGC
GTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCG
CGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCCG
GGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGG
CGAANGCGGCTCTCTGGACTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTANCGATGAAT
ACTAGGTGTAGGGGTTGTCATGACCTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGAC
GGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTAGACTTGACATCTCCTGCATTACCCTTAACCGGGGAA
GTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTT
AGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGG
CTACACACGTGCTACAATGGCAAGTACAACGAGATGCAATACCGCGAGGTGGAGCTAACCTATAAAACTTGTCTCAGTTCGGATTGTAGGCTGAAACTCG
CCTACATGAAGCTGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTGG
CAATACCCAAAGTCCGTGAGCTAACCCGTAAGGGAGGCAGCGGCCTAAGGTAGGGTCAGCGATTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGAA
CCTGCGGCTGGATCACCTCCTTTCTG
>CV1 *Clostridium tertium*
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGTGATTTCCTTTGGGAATGAACCTAGCGGCGGACG
GGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTTCCGAAAGGGAGATTAATACCGCATAATATTGAAGCTTCGCATGAAGCTTCAATC
AAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGT
GATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCG
CGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCC
GGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTG
GCGAAGGCGGCTCTCTGGACTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGA
ATACTAGGTGTAGGGGTTGTCATGACCTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGNNTACGGTCGCAAGATTAA
>CV6 *Clostridium tertium*
AGAGTTTGATTCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGTGATTTCCTTCGGGAATGAACCTAGCGGCGGACG
GGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGGGAGATTAATACCGCATAATATTGAAGCTTCGCATGAAGCTTCAATC
AAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGT
GATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCG

CGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCC
GGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTG
GCGAAGGCGGCTCTCTGGACTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCAAACAGGANTAGATACCCTGGTAGTCCACGCCGTAAACGATGA
ATACTAGGTGTAGGGGTTGTCATGACNTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTG
ACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTAGACTTGACATCTCCTGCATTACCCTTAACCGGGG
AAGTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATTG
TTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGG
GCTACACACGTGCTACAATGGCAAGTACAACGAGATGCAATACCGCGAGGTGGAGCTAACCTATAAAACTTGTCTCAGTTCGGATTGTAGGCTGAAACTC
GCCTACATGAAGCTGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTG
GCAATACCCAAAGTCCGTGAGCTAACCCGTAAGGGAGGCAGCGGCCTAAGGTAGGGTCAGCGATTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGA
ACCTGCGGCTGGATCACCTCCTTTCTG
>MV7*Clostridium tertium*
AGAGTTTGATTCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGTGATTTCCTTCGGGAATGAACCTAGCGGCGGACG
GGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGGGAGATTAATACCGCATAATATTGAAGTTTCGCATGAAGCTTCAATC
AAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGT
GATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCG
CGTGAGTGATGAAGGTTTTCGGATTGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGACGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCC
GGGGGCTGCATTTCNNCTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTG
GCGAANGCGGCTCTCTGGACTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGA
ATACTAGGTGTAGGGGTTGTCATGACCTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTG
ACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTAGACTTGACATCTCCTGCATTACCCTTAACCGGGG
AAGTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTG
TTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGG
GCTACACACGTGCTACAATGGCAAGTACAACGAGATGCAATACCGCGAGGTGGAGCTAACCTATAAAACTTGTCTCAGTTCGGATTGTAGGCTGAAACTC
GCCTACATGAAGCTGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTG
GCAATACCCAAAGTCCGTGAGCTAACCCGTAAGGGAGGCAGCGGCCTAAGGTAGGGTCAGCGATTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGAGA
ACCTGCGGCTGGATCACCTCCTTTATG
>MV6 *Clostridium tertium*
AGAGTTTGATTATGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAGGTGATTTCCTTCGGGAATGAACCTAGCGGCGGACG
GGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGGGAGATTAATACCGCATAATATTGAAGCTTCGCATGAAGCTTCAATC
AAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGT
GATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCG
CGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCC
GCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCC
GGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTG
GCGAANGCGGCTCTCTGGACTGTAACTGACGCTGANGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGA
ATACTAGGTGTNNGGGTTGTCATGACCTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATT
>GRX1 *Clostridium tertium*
TCCTTCGGGAATGAACCTANCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGGGAGATTAATACCGCAT
AATATTGAAGCTTCGCATGAAGTTTCAATCAAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGG
CGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAA
TGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGATGAAGGTCTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGACGGTACCTGAGGAGGA
AGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGCGTAGGCGGACTTTTAA
GTGAGATGTGAAATACTCGGGCTCAACCCGGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTCCTAGTGTAGCGGTGAA
ATGCGTAGAGATTAGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGACTGTAACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGAT
ACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTAGGGGTCACCATGACCTCTGTGCCGCCGCTAACGCATTAAGTATTCCGCCTGGGGAGTAC
GGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCTAGACTT
GACATCTCCTGCATTACCCTTAATCGGGGAAGTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTT
AAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGAGGAAGGTGGGGATGA
CGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCAAGTACAGCGAGATGCAATACCGCGAGGTGGAGCTAACCTATAAAACTT
GTCTCAGTTCGGATTGTAGGCTGAAACTCGCNTACANGAANCTGGAGTT
>GRX2
TGCAGTCGAGCGAAGTGCTTACAACGGAATCCTTCGGGATGAAGATGTAAGTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGGAACCTACCTTCCA
CAAGGGGATAGCCTCGGGAAACTGGGATTAATACCCTATGAGACTTAGGTATCGCATGATACTTAGGTTAAAGCGTAAAGCGGTGGAAGATGGCCTCGCG
TCCCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGGACGGCCACACTGGAACTGAGACACGGTCCA
GACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAG
CTCTGTCCTATGGGACGATAATGACGGTACCATAGGAGGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCC
GGAATGACTGGGCGTAAAGGGTACGTAGGCGGCCGTGTAAGTCAGGTGTGAAAGGCTAGGGCTCAACCCTAGTGAGCACTTGAAACTGCATGGCTTGA
GTCGAGGAGAGGAAAGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGCCTTAGACTGAC
GCTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTCGGGACATAAGTTTCGGTGC
CGAAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGT
TTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGTCCTAGAGATAGGGCCTTCCCTTCGGGGACAGAGGTGACAGGTGGT
GCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGCCTTTAGTTGCTAACAGTTAGGCTGAGCACTCTAG
AGGGACTGCCGGCAATAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAA
AGGGCAGCGACCNAGCGATAGGAAGCGAATCCCATAAAGCCGATCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAGTTGCTAGTAA
TCGCGGATCAGAATGCCGCGGTGAATGCGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCAACAATGCCCGAAGCCAGTGAGCTAACC
GTAAGGGGGCAGCTGTCGAAGGCAGG
>GRM1
AGAGTTTGATTCTGGCTCNGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAAGTACTTACAACGGAATCCTTCGGGAAGAAGATGTAA
GTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGGAACCTACCTTCCACAAGGGGATAGCCTCGGGAAACTGGGATTAATACCCTATGAGACTTAGGT
ATCGCATGATACTTAGGTTAAAGCGTAAAGCGGTGGAAGATGGCCTCGCGTCCCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGGT
AGCCGGCCTGAGAGGGTGGACGGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAG
CCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAGCTCTGTCCTATGGGACGATAATGACGGTACCATAGGAGGAAGCCCCGGC
TAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAATGACTGGGCGTAAAGGGTACGTAGGCGGCCGTGTAAGTCAGGTGT
GAAAGGCTAGGGCTCAACCCTAGTGAGCACTTGAAACTGCATGGCTTGAGTCGAGGAGAGGAAAGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAT
ATCAGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGCCTTAGACTGACGCTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT

CCACGCCGTAAACGATGAGTGCTAGGTGTCGGGACATAAGTTTCGGTGCCGAAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTG
AAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGAC
AGTCCTAGAGATAGGGCCTTCCCTTCGGGGACAGAGGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAAC
GAGCGCAACCCTTGCCTTTAGTTGCTAACAGTTAGGCTGAGCACTCTAGAGGGACTGCCGGCAATAAGTCGGAGGAAGGTGGGGATGACGTCAAATCAT
CATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGCAGCGAGCGAGCGATCGTGAGCGAATCCCATAAAGCCGATCCCAGT
TCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAGTTGCTAGTAATCGCGGATCAGAATGCCGCGGTGAATGCGTTCCCGGGTCTTGTACACACC
GCCCGTCACACCATGGGAGTCAACAATGCCCGAAGCCAGTGAGCTAACCGTAAGGGGGCAGCTGTCGAAGGCAGGGTTGATGACTGGGGTGAAGTCGT
AACAAGGTAGCCGTATCGGAAGGTGCGGCTGGNTCACCTCCTTTCTGAT
>GRC1
AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAAGTGCTTTAAGTGGAATCCTTCGGGAAGAAACTTGAA
GCACTTAGCGGCGGACGGGTGAGTAATACGTGAGGAACCTACCTCCCACAAGGGGATAGCCTCGGGAAACCGGGATTAATACCCTATGACACCTAGGTA
TCGCATGATACTTAGGTTAAAGCGCTTTAGCGGTGGGAGATGGCCTCGCGTCCCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGGT
AGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAG
CCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAGCTCTGTCCTATGGGAAGATAATGACGGTACCATAGGAGGAAGCCCCGGC
TAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAATGACTGGGCGTAAAGGGTACGTAGGCGGCTGAGTAAGTCAGGTGT
GAAAGTCTAGGGCTCAACCCTAGTGAGCACTTGAAACTGCTTGGCTTGAGTCGAGGAGAGGAAGGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAT
ATCAGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGCCTTAAACTGACGCTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT
CCACGCCGTAAACGATGAGTGCTAGGTGTCGGGGGTTTAACCTTCGGTGCCGCAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGTTCGCAAGAAT
GAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTG
CCGGGACTAGAGATAGTCCTTTCCCTTCGGGGACGGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCA
ACGAGCGCAACCCTTGCCTTTAGTTGCCAGCACTTTAGGTGGGCACTCTAGAGGGACTGCCGATGACAAATCGGAGGAAGGTGGGGATGACGTCAAATC
ATCATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAACGGGTAGCAAGCGAGCGATCGTGAGCCAATCCCTAAAAGCCGATCTCAG
TTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATGCGTTCCCGGGTCTTGTACACAC
CGCCCGTCACACCATGGGAGTTAGCAATACCCGAAGCCAGTGAGCTAACCTTAAAGGGGGTAGCTGTCGAAGGTAGGGCTAATGACTGGGGTGAAGTC
GTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTATG
>RX1
GTCGAGCGAGGTNNTTCCTTCGGGAATGAACCTANCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTTATAGAGGGGAATAGCCTCCCGAAAGG
GAGATTAATACCGCATAATATTGAAGCTTCGCATGAAGTTTCAATCAAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGT
AACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGT
GGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGATGAAGGTCTTCGGATCGTAAAGCTCTGTCTTCAGGGACGATAATGA
CGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGGAGC
GTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCCGGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTGGAATTC
CTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGGCGAAGGCGGCTCTCTGGACTGTAACTGACGCTGAGGCTCGAAAGCGTGGGGA
GCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTAGGGGTCACCATGACCTCTGTGCCGCCGCTAACGCATTAAGTAT
TCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAA
GAACCTTACCTAGACTTGACATCTCCTGCATTACCCTTAATCGGGGAAGTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGT
CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACCGGGA
GGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCAAGTACAGCGAGATGCAATACCGCGAGGTGGA
GCTAACCTATAAAACTTGTCTCAGTTCGGATTGTAGGCTGAAACTCGCNTACANGAANCTGGAGTT
>GRX3
AGAGTTTGATCATGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGAGTAATTAAAGTAGCTTGCTACATTAATTACTTAGTGGC
GGACGGGTGAGTAACGCGTGGATAACCTACCTTACACAGGGGGATAACAACGAGAAATCGTTGCTAATACCGCATAAGTGTGATTTCCGCATGGGGATCA
TATGAAAGCTCCGGCGGTGTAAGATGGGTCCGCGTCTGATTAGCTAGTTGGTGAGATAACAGCCCACCAAGGCAACGATCAGTAGCCGGCTTGAGAGAG
TGGACGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGC
CGCGTGAAGGAAGAAGGTCTTCGGATCGTAAACTTCTATCAGCAGGGAAGAAGATAGTGACTGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAG
CAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGAATCACTGGGTGTAAAGGGTGCGTAGGCGGTAGGGCAAGTCAGATGTGAAAGCCCGGGGC
TCAACTCCGGGATTGCATTTGAAACTGTCTTACTAGAGTGCAGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAAGAACA
CCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGCTGAGGCACGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATGAGTGCTAGGTGTCGGGGTCTTATGATCTCGGTGCTGAAGTCAACGCATTAAGCACTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGG
AATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAAGGCTTGACATCCTCCTGACAGGCTCTTAAT
CGAGTCTTTCCTTCGGGACAGGAGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTT
ATCTTTAGTAGCCAGCGGGTAAAGCCGGGCACTCTAGAGAGACTGCCAGGGTGAACCTGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTAT
GTCTTGGGCTACACACGTGCTACAATGGCTGCGACAGAGGGAAGCGAAGGGGCGACCTGGAGCGAAACCCAAAAAAGCAGTCCCAGTTCGGATTGTAG
TCTGCAACTCGACTACATGAAGTCGGAATCGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACC
ATGGGAGTCGGGAGCACCCGAAGCCGGTGACCTAACCATATGGGAGGAGCCGTCGAAGGTGAAACCGATGACTGGGGTGAAGTCGTAACAAGGTAGC
CGTATCGGAAGGTGCGGCTGGATCACCTCCTTTATG
>GRC4
TGCAGTCGAGCGAGGGANTTATGACGGATTCCTTCGGGATGAAGACATAAATCACCTAGCGGCGGACGGGTGAGTAACGCGTGAGGAACCTACCTCCTA
CAAGGGGATAGCCTCGGGAAACTGGGATTAATACCCTATGACACTTAAGGATCGCATGATCTTTAGGTTAAAGCGATATAGCGGTAGGAGATGGCCTCGC
GTCCCATTAGCTAGTTGGTGAGGTAACGGCTTACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGAACGGCCACACTGGAACTGAGACACGGTCC
AGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAA
GCTCTGTCCTATGGGACGATAATGACGGTACCATAGGAGGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTC
CGGAATTACTGGGCGTAAAGGGTACGTAGGCGGCTAGGTAAGTCAGATGTGAAATCCATGGGCTCAACCCATGTAAGCATTTGAAACTGCTTAGCTTGAG
TTAAGGAGAGGAAAGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGCCTTAAACTGACG
CTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTCGGGGGTTTAACCTTCGGTGC
CGCAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGT
TTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGCCGGGACTAGAGATAGTCCTTTCCCTTCGGGGACGGAGGAGACAGGTGGT
GCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGCCTTTAGTTGCCAGCACTTTAGGTGGGCACTCTAG
AGGGACTGCCGATGACAAATCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAA
CGGGTAGCAAGCGAGCGATCGTGAGCCAATCCCTAAAAGCCGATCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAGTTGCTAGTAA
TCGCGGATCAGCATGCCGCGGTGAATGCGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTTAGCAATACCCGAAGCCAGTGAGCTAACC
TTAAAGGGGGCAGCTGTCGAAGGTAGGGCTAATGACTGGGG
>GRC3
GTCGAGCGAAGTGCTTTAAGTGGAATCCTTCGGGATAAAACTTGAAGCACTTAGCGGCGGACGGGTGAGTAATACGTGAGGAACCTACCTTCCACAAGG
GGATAGCCTCGGGAAACCGGGATTAATACCCTATGACACCTAAGAATCGCATGATACTTAGGTTAAAGCGATTTAGCGGTGGGAGATGGCCTCACGTCCC
ATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAGCTCT
GTCCTATGGGAAGATAATGACGGTACCATAGGAGGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCGAGCGTTGTCCGGAA

TGACTGGGCGTAAAGGGTACGTAGGCGGCTGAGTAAGTCAGGTGTGAAAGTCTAGGGCTCAACCCTAGTGAGCACTTGAAACTGCTTGGCTTGAGTCGA
GGAGAGGAAGGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAATACCAGTGGCGAAGGCGACCTTCTGGCCTTGGACTGACGCTG
ANGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTCGGGGGAGATATCCTTCGGTGCC
NAAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTT
TAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACCGGTGTAGAGATACGCCTTTCTTCCTTCGGGAAGACAGAGGAGACAGGTG
GTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGCCTCTAGTTGCCAGCACTTCGGGTGGGCACTC
TAGAGGGACTGCCGGCAAAAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTTATGCCCTGGGCTACACACGTGCTACAATGGTCGGTA
CAAAGGGCAGCGACCTAGCGATAGGAAGCGAATCCCAAAAAGCCGATCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAGTTGCTAG
TAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCAGCAATGCCCGAAGCCAGTGAGCTA
ACCCGTATGGGAGGCAGCTGTCGAAGGCAGGGCTGATGACTGGGG
>GRC5
AGTTTGATTCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAGCGAAGTGCTTACAGCGGAATCCTTCGGGGTGAAGCAGTAAGT
ATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGGAACCTGCCTTCCACAAGGGGATAGCCTCGGGAAACTGGGATTAATACCCTATGACACTTAAGTAT
CGCATGATACTTAGGTTAAAGCGTAAAGCGGTGGAAGATGGCCTCGCGTCCCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGGGTAG
CCGGCCTGAGAGGGTGGACGGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCC
TGATGCAGCGACGCCGCGTGAGCGAAGAAGGCCTTCGGGTCGTAAAGCTCTGTCCTATGGGACGATAATGACGGTACCATAGGAGGAAGCCCCGGCTA
ACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTGTCCGGAATGACTGGGCGTAAAGGGTACGTAGGCGGTTATGTAAGTCAGGTGTGA
AAGGCTAGGGCTCAACCCTAGTAAGCACTTGAAACTGCGTAACTTGAGTCGAGGAGAGGAAAGTGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATC
AGGAGGAATACCAGTGGCGAAGGCGACTTTCTGGCCTTGGACTGACGCTGAGGTACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCC
ACGCCGTAAACGATGAGTGCTAGGTGTCGGGGAGTAATCTTCGGTGCCGAAGTTAACACATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAA
ACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACA
GTCCTAGAGATAGGGCCTTCCCTTCGGGGACAGAGGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
AGCGCAACCCTTGCCTTTAGTTGCTAACAGTTAGGCTGAGCACTCTAGAGGGACTGCCGGCAATAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATC
ATGCCCTTTATGTCCTGGGCTACACACGTGCTACAATGGTCGGTACAAAGGGCAGCGAGCGAGCGATCGTGAGCGAATCCCATAAAGCCGATCCCAGTT
CGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAGTTGCTAGTAATCGTGGATCAGAATGCCACGGTGAATGCGTTCCCGGGTCTTGTACACACCG
CCCGTCACACCATGGGAGTCAACAATGCCCGAAGCCAGTGAGCTAACCGTTAGGGGGCAGCTGTCGAAGGCAGGGTTGATGACTGGGGTGAAGTCGTA
ACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTGATCTTGC
>RX3
TGCAGTCGAGCGAGGTGATTTCCTTCGGGAATGAACCTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTGCCTTATAGAGGGGAATAGCCTCCCG
AAAGGGAGATTAATACCGCATAATATTGAAGTTTCGCATGAAGCTTCAATCAAAGGAGCAATCCGCTATAAGATGGGCCCGCGGCGCATTAGCTAGTTGGT
GAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCA
GCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCAACGCCGCGTGAGTGATGAAGGTTTTCGGATCGTAAAGCTCTGTCTTCAGGGACGAT
AATGACGGTACCTGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAG
GGAGCGTAGGCGGACTTTTAAGTGAGATGTGAAATACTCGGGCTCAACCCGGGGGCTGCATTTCAAACTGAAAGTCTAGAGTGCAGGAGAGGAGAGTG
GAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACNCCAGTGGCGAAGGCGGCTCTCTGGACTGTAACTGACGCTGAGGCTCGAAAGCGT
GGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTAGGGGTTGTCATGACCTCTGTGCCGCCGCTAACGCATTA
AGTATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACG
CGAAGAACCTTACCTAGACTTGACATCTCCTGCATTACCCTTAACCGGGGAAGTCCCTTCGGGGACAGGATGACAGGTGGTGCATGGTTGTCGTCAGCTC
GTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTTGCTACCATTAAGTTGAGCACTCTAGCGAGACTGCCCGGGTTAACC
GGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCAAGTACAACGAGATGCAATACCGCGAGG
TGGAGCTAACCTATAAAACTTGTCTCAGTTCGGATTGTAGGCTGAAACTCGCCTACATGAAGCTGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTG
AATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTGGCAATACCCAAAGTCCGTGAGCTAACCCGTAAGGGAGGCAGCGGCCTAAG
GTAGGGTCAGCGATNGGGGNGAAGTC
>KV2 *Clostridium thermosuccinogenes*
GGCTTTCTGGACAGTAACTGNCGNNGAGGCNNGAAAGCGTGGGGAGCGAACAGGATTAGATACCNTGGTAGTCCACGCCGTAAACGATGGATANTAG
GTGTAGGAGGTATCGACCCCTTTCTGTGCCGGAGTAAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGG
GGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAG
CTCTTCGGAGCCTCTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTT
GGTAACACGTAGAGGTGACCACTCTAACGAGACTGCCGGTGACAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCT
ACACACGTACTACAATGGCTACAACAAAGGGAAGCGAAGCCGTAAGGCGGAGCAAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCTGCAACCCGC
CCACATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCAGC
AACACCCGAAGTCAGTAGCTTAACCGCAAGGAGGGCGCTGCCGAAGGTGGGGCCGATGATNGGGGGAAG
>KV4 *Clostridium thermosuccinogenes*
GTAAGTCCAGATGTGAAAATNCCATGGCTCAACCGTGGGGGCTGCATNTGAAACTNTGTTTCTTGAGTGCTGGAGAGGAAAGCGGAATTCNTAGTGTAG
CGGTGAAAATGCGTAGATATTAGGAGGAANCNCCCAGTGGCGAAGGCGGCTTTNTGGACAGTAACTGACGNTGAGGCGCGAAAGCGTGGGGAGCGAA
CAGGATTAGATACCCTGGTAGTCCNCGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTAAACACAATAAGTATCCCA
CCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAAC
CTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAGCTCTTCGGAGCCTCTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGA
GATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGGTAACACGTAGAGGTGACCACTCTAACGAGACTGCCGGTGACAAATCGGAGG
AAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTACAACAAAGGGAAGCGAAGCCGTAAGGCGGAGC
AAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCTGCAACCCGCCCACATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATAC
GTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCAGCAACACCCGAAGTCAGTAGCTTAACCGCAAGGAGGGCGCTGCCGAAGGTGGGG
CCGATGATTG >KV5 *Clostridium thermosuccinogenes*
GGGTGTAAAGGGCGCGTAGGCNGGGGATGTAAGTCAGATNNGAAATACCATGGCTCAACCGTGGGGCTGCATCTGAAACTATGNTCTTGAGTGCTGGA
GAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGG
CGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCNTGGTAGNNACGCNGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGG
AGTAAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTA
ATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAGCTCTTCGGAGCCTCTGTACAGGTGGTGCATGG
TTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGGTAACACGTAAAGGTGACCACTCTAACGAGA
CTGCCGGTGACAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTACAACAAAGGGA
AGCGAAGCCGTAAGGCGGAGCAAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCTGCAACCCGCCCACATGAAGTCGGAATTGCTAGTAATGGCAG
GTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCAGCAACACCCGAAGTCAGTAGCTTAACCGCAAGG
AGGGCGCTGCCGAAGGTGGGGCCGATGATTGGG
>GV1 *Clostridium thermosuccinogenes*
TGTTTCTTGAGTGCTGGAGAGGAAAGCCGGAATTCCTAGTGNAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGA
CAGTAACTGACGNTGAGGCGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCNTGGTAGTNCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTA
TCGACCCCTTCTGTGCCGGAGTAAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAG
CAGTGGAGTATGTGGTTTAATTNGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAGCTCTTCGGAGCCTC

TGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGGTAACACGTAGAG
GTGACCACTCTAACGAGACTGCCGGTGACAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAA
TGGCTACAACAAAGGGAAGCGAAGCCGTAAGGCGGAGCAAATCCCCAAAAGTAGTCCCAGTTTGGATGGTGGGCTGCAACCCGCCCACATGAAGTCGG
AATTGCTAGTAATGGCAGGTCAGCATATCNCCGCCAATACGTTCCCGGACCCANNTTNATGCCCGAGCCACTCCATTNGAGTCCGCAACCTCA
>GV2 *Clostridium thermosuccinogenes*
TCNTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTTGGACAGTAACTGACGNTGAGGCGCGAAAGCGTGGGG
AGCGAACAGGATTAGATANCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTAAACACAATAAGT
ATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGA
AGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAGCTCTTCGGAGCCTCTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGT
CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGGTAACACGTAAAGGTGACCACTCTAACGAGACTGCCGGTGACAAATCG
GAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTACAACAAAGGGAAGCGAAGCCGTAAGGCG
GAGCAAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCTGCAACCCCCCCACATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGA
ATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCAGCAACACCTGAAGTCAGTAGCTTANCTGCAAGGAGGGC
>GV3 *Clostridium thermosuccinogenes*
CGTGGGGNNTGCATNTGAAACTATGTTTCTTGAGTGCTGGAGAGGAAAGCGGAATTCNTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCA
GTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTAAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGG
AATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAG
ATGGCATAGCTCTTCGGAGCCTCTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTT
GTTGTTAGTTGGTAACACGTAAAGGTGACCACTCTAACGAGACTGCCGGTGACAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATG
TCCTGGGCTACACACGTACTACAATGGCTACAACAAAGGGAAGCGAAGCCGTAAGGCGGAGCAAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCT
GCAACCCGCCCACATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCAT
GAGAGTCAGCAACACCCGAAGTCAGTAGCTTAACCGCAAGGAGGGCGC
>GV7 *Clostridium thermosuccinogenes*
GCGTAGGCGGGGATGTAAGTCAGATGTGAAATACCATGGCTCACCNGTGGGGCTGCATNNNAAACTATGTTTCTTGAGTGCTGGAGAGGAAAGCGGAA
TTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGNTGAGGCGCGAAAGCGTGG
GGAGCGAACAGGATTAGATACCCTGGTAGNNACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTAAACACAATAA
GTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGC
GAAGAACCTTACCAGGGCTTGACATACAGAGGGATATGTCAGAGATGGCATAGCTCTTCGGAGCCTCTGTACAGGTGGTGCATGGTTGTCGTCAGCTCGT
GTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGGTAACACGTAAAGGTGACCACTCTAACGAGACTGCCGGTGACAAAT
CGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTACAACAAAGGGAAGCGAAGCCGTAAGG
CGGAGCAAATCCCCAAAAGTAGTCCCAGTTCGGATTGTGGGCTGCAACCCGCCCACATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCG
TGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCAGCAACACCCGAAGTCAGTAGCTTAACCGCAAGGAGGGCGCTGCCGAAG
GTGGGGCCGATGATTGGG
>JGU1 *Clostridium stercorarium*
NACCNAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGANCNACCTGCCCTATGNTGGGGGATAACACCGGGAAACCGGTGCTAATACC
GCATAANACCACNGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAGCTAGTTGGTAGGGTAACGGCCTACC
AAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGANACACGGCCCANACTCCTACGGGAGGCAGCAGTGGGGAATATTG
CGCANTGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATCGGGGACGAAGATGACGGTACCCG
AAGAACAAGCCACGGNTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGCGTGTAGGCGGG
GTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGANTGCCGGANAGGAAAGCGGAATTCCCNGTGTA
NCGGTGAAATGCGTANATATTGNNAGGAACACCAGNGGCGAAGGCGGCTTTCTGGACGGTAACTGACNCTGAGGCGCGAAAGCGTGGGGAGCAAACN
GGATTAGATACCCTGNTANTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATNNACCCCTTTCTGTGCCGTAGTTAACACAATAAGTATCCCACC
TGGGGAGTACGNCNGCAAGGCTGAAAANTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTNGAAGCAACGNGAAGAAC
CTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTC
GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAGACTGCCGGCGAGAAGTCGGA
GGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAATCCGCGAGGAGGA
GCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGNTGCAACTCGCNTGC
>*Neu29 C. stercorarium*
CGGGATCCGTGTTACGGAGGTCTTTGGACCGAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTGGGGGAT
AACACCGGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAG
CTAGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATC
GGGGACGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTG
GGTGTAAAGGGCGTGTAGGCGGGGTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGAGTGCCGGAG
AGGAAAGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGAGGC
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGTA
GTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAA
TTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCAT
GGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAG
ACTGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGG
GAAGCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGC
AGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGGCAACACCCGAAGCCGGTAGCCTAACCGAGA
GGGGGGCGCCGTCGAAGGTGGGGTCAGTGATTGGGG
>JGU2 *Clostridium stercorarium*
TGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTGGGGGATAACACCGGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATG
TCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAGCTAGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGAC
TGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGC
AGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATCGGGGACGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGCGTGTAGGCGGGGTGCCAAGTCNNGNGTGAAATAC
CGGGGCTTAACCTCGGGGGTGCATCNGANACTGGTGCTCTTGAGTGCCGGAGAGGAAAGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGG
GAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCAC
GCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGTAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGA
AACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGAC
GGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAAC
GAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAGACTGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCAT
CATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTT
CGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGCAGGTCAGCATA

>JGU3 *Clostridium stercorarium*
AGTCGAACGGGATCCGTGTTACGGAGGTCTTTGGACCGAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTG
GGGGATAACACCGGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTC
CGATTAGCTAGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAG
ACTCCTACGGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTC
CTTTGATCGGGGACGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGG
AATTACTGGGTGTAAAGGGCGTGTAGGCGGGGTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGAGT
GCCGGAGAGGAAAGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACG
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTG
TGCCGTAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGT
GGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTC
TAACGAGACTGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTA
CAGAGGGAAGCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAG
TAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGGCAACACCCGAAGCCGGTAGCCTA
ACCGAGAGGGGGGCGCCGTCGAAGGTGGGGTTCAGTGATTGGGG
>HAW1 *Clostridium stercorarium*
GGATCCGTGTTACGGAGGTCTTTGGACCGAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTGGGGGATAA
CACCGGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAGCT
AGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATCG
GGGACGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGG
GTGTAAAGGGCGTGTAGGCGGGGTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGAGTGCCGGAGA
GGAAAGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGANGCG
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGANGTATCGACCCCTTCTGTGCCGTAG
TTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAAT
TCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCATG
GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAGAC
TGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAA
GCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGCAG
GTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGGCAACACCCGAAGCCGGTAGCCTAACCGAGAGG
GGGGCGCCGT
>HAW2 *Clostridium stercorarium*
CGGGATCCGTGTTACGGAGGTCTTTGGACCGAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTGGGGGAT
AACACCGGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAG
CTAGTTGGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTAC
GGGAGGCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATC
GGGGACGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTG
GGTGTAAAGGGCGTGTAGGCGGGGTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGAGTGCCGGAG
AGGAAAGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGAGGC
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGTA
GTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAA
TTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCAT
GGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAG
ACTGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGG
GAAGCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGC
AGGTCAGCATACTGCCGNGAATANGTTCCCGGGCTT
>HAW3 *Clostridium stercorarium*
CCGTGTTACGGAGGTCTTTGGACCGAAGTGGCATGGTGAGAGTGGCGGACGGGCGAGTAACGCGTGAGCAACCTGCCCTATGCTGGGGGATAACACC
GGGAAACCGGTGCTAATACCGCATAAGACCACAGTGACGCATGTCACAGTGGTAAAAGCTGAGGCGGCATAGGATGGGCTCGCGTCCGATTAGCTAGTT
GGTAGGGTAACGGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTTGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTACGGGAG
GCAGCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGGAGGAAGAAGGCCTTTGGGTTGTAAACTCCTTTGATCGGGGA
CGAAGATGACGGTACCCGAAGAACAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGT
AAAGGGCGTGTAGGCGGGGTGCCAAGTCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATCTGAAACTGGTGCTCTTGAGTGCCGGAGAGGAA
AGCGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGAGGCGCGAA
AGCGTGGGGAGCAAACAGGANTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGTAGTTAA
CACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGA
AGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCATGGTTG
TCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAGACTGC
CGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGC
AAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGCAGGTC
AGCATACTGCCGTGAATACGTTCCCGGGCNTTGTACACACCGCCCGTCAC
>Neu26 *C. stercorarium*
TTCTGGACGGTAACTGACGNTGAGGCGCGAAAAGCGTGGGGAGCAAACAGGATTAGATACCNTGGTAGTCCNNCGCTGTAAACGATGGATACTAGGTG
TAGGAGGTATCGACCCNTTCTGTGCCGTAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGC
CCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCG
CAAGGAGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTT
GCCAGCAGTAAGATGGGCACTCTAACGAGACTGCCGGCGAGAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTA
CACACGTGCTACAATGGCGACTACAGAGGGAAGCAAATCCGCGAGGAGGAGCAAATCCCGAAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGC
CTGCATGAAGCCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCC
>Neu27 *C. stercorarium*
TCAGGTGTGAAATACCGGGGCTTAACCTCGGGGGTGCATNTGAAACTGGTGCTCTTGAGTGCCGGAGAGGAAAGCGGAATTCCCAGTGTAGCGGTGAA
ATGCGTAGATATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGAT
ACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGTAGTTAACACAATAAGTATCCCACCTGGGGAGTA
CGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGC
TTGACATCCCCCTGACGGATGTAGAGATACATCTTCTCCGCAAGGAGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTG
GGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCAGTAAGATGGGCACTCTAACGAGACTGCCGGCGAGAAGTCGGAGGAAGGTGGG
GATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAATCCGCGAGGAGGAGCAAATCCCG

AAAGGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGGCAACACCCGAAGCCGGTAGCCTAACCGAGAGGGGGGCGCCGTCGAAGGTGGGGTCAGTG
A
>Neu5 *C. cellulosi*
AGTACTTCNGTTCTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGCAACCTGCCCTTGAGCGGGGGATAGCGTCTGGAAACGGACGGTAATACCGC
ATAATGTACGTTGGAGGCATCTCCGATGTACCAAAGGANNAATCCACTCAAGGATGGGCTCGCGTCCGATTAGGTAGTTGGTGAGGTAACGGCCCACCA
AGCCTGCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGGATATTGC
ACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCTGTCTTTCGGGACGAAGGAAGTGACGGTACCGA
AAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGTGCGTAGGCGG
GTTGTCAAGTTGGATGTGAAATCTCTGGGCTTAACTCAGAGGTTGCATTCAAAACTGGCGATCTTGAGTGAGGTAGAGGCAGGCGGAATTCCCGGTGTAG
CGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCCTTAACTGACGCTGAGGCACGAAAGCATGGGGAGCAAACAG
GATTAGATACCCTGGTAGNCNATGCTGTAAACGATGATTGCTAGGTGTGGGTGGACTGACCCCATCCGTGCCGGAGTTAACACAATAAGCAATCCACCTG
GGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTAC
CAGGTCTTGACATCCACCGAATCCGGAAGAGATTCTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGA
GATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAAGACTGCCGTTGATAAAACGGAGGAAGGTGGG
GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAGGGAAGCAATACCGCGAGGTGGAGCGAATCCC
CAAAAGCGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCC
GGGCNTTGTACAC
>Neu7 *C. cellulosi*
CTTCNGTTCTATCTTANCGGCGGACGGGTGAGTAACGCGTGAGCAACCTGCCCTTGAGCGGGGGATAGCGTCTGGAAACGGACGGTAATACCGCATAAT
GTACGTTGGAGGCATCTCCGATGTACCAAAGGAGAAATCCACTCAAGGATGGGCTCGCGTCCGATTAGGTAGTTGGTGAGGTAACGGCCCACCAAGCCT
GCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGGATATTGCACAAT
GGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCTGTCTTTCGGGACGAAGGAAGTGACGGTACCGAAAGA
GGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGTGCGTAGGCGGGTTGT
CAAGTTGGATGTGAAATCTCTGGGCTTAACTCAGAGGTTGCATTCAAAACTGGCGATCTTGAGTGAGGTAGAGGCAGGCGGAATTCCCGGTGTAGCGGT
GAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCCTTAACTGACGCTGANGCACNAAAGCATGGGGAGCAAACAGGATTA
GATACCCTGGTAGTCCATGCTGTNNACGATGATTGCTAGGTGTGGGTGGACTGACCCCNTCCGTGCCGGAGTTAACACAATAAGCAATCCACCTGGGGA
GTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTNGAAGCAACGCGAAGAACCTTACCAGG
TCTTGACATCCACCGAATCCGGAAGAGATTCTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGT
TGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAAGACTGCCGTTGATAAAACGGAGGAAGGTGGGGATGA
CGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAGGGAAGCAATACCGCGAGGTGGAGCGAATCCCCAAAAG
CGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCNT
TGTACAC
>Neu22 *C. cellulosi*
TGCAGTCGAGCGGAGANAGTACTTCGNTTCTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGCAACCTGCCCTTGAGCGGGGGATAGCGTCTGGAA
ACGGACGGTAATACCGCATAATGTACGTTGGAGGCATCTCCGATGTACCAAAGGANNAATCCACTCAAGGATGGGCTCGCGTCCGATTAGGTAGTTGGT
GAGGTAACGGCCCACCAAGCCTGCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCA
GCAGTGGGGGATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCTGTCTTTCGGGACGAA
GGAAGTGACGGTACCGAAAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTA
AAGGGTGCGTAGGCGGGTTGTCAAGTTGGATGTGAAATCTCTGGGCTTAACTCAGAGGTTGCATTCAAAACTGGCGATCTTGAGTGAGGTAGAGGCAGG
CGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCCTTAACTGACGCTGAGGCACGAAAG
CATGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCATGCTGTAAACGATGATTGCTAGGTGTGGGTGGACTGACCCCATCCGTGCCGGAGTTAACAC
AATAAGCAATCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGC
AACGCGAAGAACCTTACCAGGTCTTGACATCCACCGAATCCGGAAGAGATTCTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGTGGTGCATGGTTGTC
GTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAAGACTGCCGTTGATAA
AACGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAGGGAAGCAATACCGCG
AGGTGGAGCGAATCCCCAAAAGCGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCGCGGATCAGCATGCCG
CGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCCGGAAACACCCGAAGTCGTTTGCGTAACCGAAAGGAGCGCGGCGC
>Neu32 *C. cellulosi*
GATAGCGTCTGGNAACGGACGGNNATACCNCATAATGTACGTTGGAGGCATCTCCGATGTACCAAAGGANNAATCCACTCNNGGATGGGCTCGCGTCC
GATTAGGTAGTTGGTGAGGTAACGGCCCACCAAGCCTGCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGGATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCT
GTCTTTCGGGACGAAGGAAGTGACGGTACCGAAAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCG
GAATTACTGGGTGTAAAGGGTGCGTAGGCGGGTTGTCAAGTTGGATGTGAAATCTCTGGGCTTAACTCANAGGTTGCATTCAAAACTGGCGATCTTGAGT
GAGGTAGAGGCAGGCGGAATTCCCGGTGTANCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCCTTAACTGAC
GCTGANGCACGAAAGCATGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCATGCTGTAAACGATGATTGCTAGGTGTGGGTGGACTGACCCCATCCG
TGCCGGAGTTAACACAATAAGCAATCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGT
GGTTTAATTNGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCACNGAATCCGGAAGAGATTNTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAA
GACTGCCGTTGATAAAACGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAG
GGAAGCAATACCGCGAGGTGGAGCGAATCCCCAAAAGCGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCG
CGGATCAGCATGCCGCGGTGAATA
>DG5 *C. cellulosi*
GTCGAGCGGAGATAGTACTTCGGTTCTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGCAACCTGCCCTTGAGCGGGGGATAGCGTCTGGAAACGG
ACGGTAATACCGCATAATGTACGTTGGAGGCATCTCCGATGTACCAAAGGAGAAATCCACTCAAGGATGGGCTCGCGTCCGATTAGGTAGTTGGTGAGGT
AATGGCCCACCAAGCCTGCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTG
GGGGATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCTGTCTTTCGGGACGAAGGAAGT
GACGGTACCGAAAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGGGT
GCGTAGGCGGGTTGTCAAGTTGGATGTGAAATCTCTGGGCTTAACTCAGAGGTTGCATTCAAAACTGGCGATCTTGAGTGAGGTAGAGGCAGGCGGAAT
TCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAAGGCGGCCTGCTGGGCCTTAACTGACGCTGAGGCACGAAAGCATGG
GGAGCAAACAGGATTAGATACCCTGGTAGTCCATGCTGTAAACGATGATTGCTAGGTGTGGGTGGACTGACCCCATCCGTGCCGGAGTTAACACAATAAG
CAATCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCG
AAGAACCTTACCAGGTCTTGACATCCACCGAATCCGGAAGAGATTCTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGTGGTGCATGGTTGTCGTCAGCT
CGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAAGACTGCCGTTGATAAAACGGA
GGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAGGGAAGCAATACCGCGAGGTGG
AGCGAATCCCCAAAAGCGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTGA
ATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCCGGAAACACCCGAAGTCGTTTGCGTAACCGAAAGGAGCGCGGCGCCGAAGGT
GGGATCGGTGATTGGGGNGAAGT

>DR3 *C.cellulosi*
AGTCGAGCGGANANAGTACTTCGGTTCTATCTTAGCGGCGGACGGGTGAGTAACGCGTGAGCAACCTGCCCTTGAGCGGGGGATAGCGTCTGGAAACG
GACGGTAATACCGCATAATGTACGTTGGAGGCATCTCCGATGTACCAAAGGAGAAATCCACTCAAGGATGGGCTCGCGTCCGATTAGGTAGTTGGTGAG
GTAATGGCCCACCAAGCCTGCGATCGGTAGCCGGACTGAGAGGTTGTACGGCCACATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAG
TGGGGGATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAGGGAAGAAGGTCTTCGGATTGTAAACCTCTGTCTTTCGGGACGAAGGAA
GTGACGGTACCGAAAGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGAATTACTGGGTGTAAAGG
GTGCGTAGGCGGGTTGTCAAGTTGGATGTGAAATCTCTGGGCTTAACTCAGAGGTTGCATTCAAAACTGGCGATCTTGAGTGAGGTAGAGGCAGGCGGA
ATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCAGTGGCGAANGCGGCCTGCTGGGCCTTAACTGACGCTGAGGCACGAAAGCATG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCATGCTGTAAACGATGATTGCTAGGTGTGGGNGGACTGACCCCATCCGTGCCGGAGTTAACACAATA
AGCAATCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACG
CGAAGAACCTTACCAGGTCTTGACATCCACCGAATCCGGAAGAGATTCTGGAGTGCCCTTCGGGGAGCGGTGAGACAGGTGGTGCATGGTTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTAATAGTTGCTACGCAAGAGCACTCTATTAAGACTGCCGTTGATAAAACGG
AGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTACTACAATGGCCGCCAACAAAGGGAAGCAATACCGCGAGGTG
GAGCGAATCCCCAAAAGCGGTCCCAGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAATTGCTAGTAATCGCGGATCAGCATGCCGCGGTG
AATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCCGGAAACACCCGAAGTCGTTTGCGTAACCGAAAGGAGCGCGGCGCCGAAGG
TGGGATCGGNGATTNGGG>DG3TGCAGTCGAGCGAAGTGATTTTGAATGAAGCCTTCGGGCAGAATTCAAAATTATCTTAGCGGCGGACGGGTGAGTAA
CGCGTGAGGAACCTGCCTTTCACAAAGGGATAGCCTCGGGAAACCGGGATTAATACCTTATGATACTAATTTACCGCATGGTGAATTAGTTAAAGCGCAAA
GCGGTGAAAGATGGCCTCGCGTCCCATTAGTTAGTTGGTGAGGTAACGGCTCACCAAGACCGCGATGGGTAGCCGGCCTGAGAGGGTGAACGGCCACA
CTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGAGGAAACTCTGATGCAGCGACGCCGCGTGAACGAAG
AAGGTCTTCGGATCGTAAAGTTCTGTCCTAGGGGACGATAATGACGGTACCCTAGGAGGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACG
TAGGGGGCGAGCGTTGTCCGGAATTATTGGGCGTAAAGGGTTCGTAGGCGGTTATTTAAGTCAGGTGTGAAAGGCATGGGCTTAACCTGTGTAAGCACT
TGAAACTGGATAACTTGAGTTAAGGAGAGGAAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAATACCAGTGGCGAAGGCGACTT
TCTGGACTTATACTGACGCTGAGGAACGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAGGTGTTGG
AGGGAAACCTTCGGTGCCGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGTACGCAAGTATGAAACTCAAAGGAATTGACGGGGACCCGCACAA
GCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATGCCCCTGCCCGCTCTAGAGATAGGGCTTTACCCTTCGGGG
TACAGGGGACACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGCCTTTAGTTGCCATC
ATTAAGTTGGGCACTCTAGAGGGACTGCCGATGACAAATCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTTATGTCCTGGGCTACACACGTG
CTACAATGGTCGGTACAACGAGCAGCAACACAGTGATGTGAAGCAAATCTCTAAAAGCCGATCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAA
GTCGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATGCGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTTGGCAATACCCG
AAGCCAGTAAGCTAACCGTAAGGAGGCAGCTGTCGAAGGTAGGGTCAATGACTGGGGNGAAGTCG
>T3/55
GAGTTTNATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAACACATGCAAGTCGAACGGGGTTATCGTGATGAAGTCTTCGGATGGAATTACGATAAC
CTAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCCCATACAGGGGGATAACAGCCGGAAACGGCTGCTAAAACCGCATAAGCGCACACCGGC
GCATGCCGGAGTGTGAAAATGTATGCCGAAAGGTATGCAGGTATGGGATGGGCCCGCGTCTGATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCG
ACGATCAGTAGCCGGCTTGAGAGAGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATG
GGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTACTTCGGTACGTAAAGCTCTATCAGCAGGGAAGAAGATGACGGTACCTGACTAAGAAG
CCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGTGTAAAGGGAGCGTAGGTGGCGAAGCAAGT
CAGATGTGAAAGCCCGGGGCTCAACTCCGGGACTGCATTTGAAACTGTTTTGCTAGAGTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAAT
GCGTAGATATTAGGAGGAACACCAGTGGCGAANGCGGCTTACTGGACTGTAACTGACACTGANGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATAC
CCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGGTCAAAAGACNTTCGGTGCCGTCGCAAACGCATTAAGTATTCCACCTGGGGAGTAC
GTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGCCTT
GACATCCCATTGACCGGTCCTTAACCGGACCTTTCCTTCGGGACAATGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTA
AGTCCCGCAACGAGCGCAACCCTTATCCTTAGTAGCCAGCACTTCGGGTGGGGACTCTAGGGAGACTGCCAGGGACAACCTGGAGGAAGGCGGGGATG
ACGTCAAATCATCATGCCCCTTATGGACTGGGCTACACACGTGCTACAATGGCGGTGACAAAGGGAAGCGACCCTGCGAAGGTGAGCGAATCCCAAAAA
AGCCGTCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTTCCCGGGTC
TTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGTCAGTGACCCAACCCGAAAGGGAGGGAGCTGCCGAAGGTGGAGCCGATGACT
GGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCT
>SD1I
ACGGAGTTTATATTATGAAGTCTTCGGATGAATTGATATAAACTTAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAACA
GCCGGAAACGGTTGCTAAAACCGCATAAGCGCACAGCAGCGCATGCTGCAGTGTGAAAATATTTTATAGGTATGAGATGGGCCCGCGTCTGATTAGCTAG
TTGGTGAGGTAACTGCTCACCAAGGCGACGATCAGTAGCCGGCTTGAGAGAGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGA
GGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTACGTAAAGCTCTATCAGCAGGG
AAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGT
AAAGGGAGCGTAGGTGGCAAGTCAAGTCAGATGTGAAAGCCCGGGGCTCAACTCCGGGATTGCATTTGAAACTGGTTTGCTAGAGTGCAGGAGAGGTAA
GTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCGAAAG
CGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGAGCAAAAGCTCTTCGGTGCCGTCGCAAACG
CATTAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTNGAAGC
AACGCGAAGAACCTTACCAGGCCTTGACATCCCATTGACCGGTCCTTAACCGGACCTTTCCTTCGGGACAATGGAGACAGGTGGTGCATGGTTGTCGTCA
GCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTTAGTAGCCAGCAAGTAAAGTTGGGCACTCTAGGGAGACTGCCAGG
GATAACCTGGAGGAAGGCGGGGATGACGTCAAATCATCATGCCCCTTATGGACTGGGCTACACACGTGCTACAATGGTGGCTACAAAGGGAGGCGACCC
TGTGAAGGTGAGCAAATCCCAAAAAGGCCATCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGCGAATCAGCAT
GTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGTCAGTGACCCAACCGTAAGGAGGGAGC
T
>SD1G
TCGAACGGAGTTTATATTATGAAGTCTTCGGATGAATTGATATAAACTTAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGAT
AACAGCCGGAAACGGTTGCTAAAACCGCATAAGCGCACAGCAGCGCATGCTGCAGTGTGAAAATATTTTATAGGTATGAGATGGGCCCGCGTCTGATTAG
CTAGTTGGTGAGGTAACTGCTCACCAAGGCGACGATCAGTAGCCGGCTTGAGAGAGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTACGTAAAGCTCTATCAGCA
GGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGG
GTGTAAAGGGAGCGTAGGTGGCAAGTCAAGTCAGATGTGAAAGCCCGGGGCTCAACTCCGGGATTGCATTTGAAACTGGTTTGCTAGAGTGCAGGAGAG
GTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCG
AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGAGCAAAAGCTCTTCGGTGCCGTCGCA
AACGCATTAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTCG
AAGCAACGCGAAGAACCTTACCAGGCCTTGACATCCCATTGACCGGTCCTTAACCGGACCTTTCCTTCGGGACAATGGAGACAGGTGGTGCATGGTTGTC
GTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTTAGTAGCCAGCAAGTAAAGTTGGGCACTCTAGGGAGACTGC
CAGGGATAACCTGGAGGAAGGCGGGGATGACGTCAAATCATCATGCCCCTTATGGACTGGGCTACACACGTGCTACAATGGTGGCTACAAAGGGAGGC
GACCCTGTGAAGGTGAGCAAATCCCAAAAAGGCCATCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGCGAATC

AGCATGTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGTCAGTGACCCAACCGTAAGGAGG
GAGCTGCCGAAGGTGGAGCCGATGACTGGGGTGAAGT
>SD1D
TCTTCNGATGAATTGATNTAAACTTANNGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAGGGGGATAACAGCCGGAAACGGTTGCTAAA
ACCGCATAAGCGCACAGCAGCGCATGCTGCAGTGTGAAAATATTTTATAGGTATGAGATGGGCCCGCGTCTGATTAGCTAGTTGGTGAGGTAACTGCTCA
CCAAGGCGACGATCAGTAGCCGGCTTGAGAGAGTGACCGGCCACATTGGGACTGAGACACGGCCCAAACTCCTACGGGAGGCAGCAGTGGGGAATATT
GCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTACGTAAAGCTCTATCAGCAGGGAAGAAAATGACGGTACCTGA
CTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGTAGGTGGCA
AGTCAAGTCAGATGTGAAAGCCCGGGGCTCAACTCCGGGATTGCATTTGAAACTGGTTTGCTAGAGTGCAGGAGAGGTAAGTGGAATTCCTAGTGTAGC
GGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAGGCTCGAAAGCGTGNGGAGCAAACAGGA
TTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGAGCAAAAGCTCTTCGGTGCCGTCGCAAACGCATTAAGTATTCCACCTGG
GGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAATTNGAAGCAACGCGAAGAACCTTACC
AGGCCTTGACATCCCATTGACCGGTCCTTAACCGGACCTTTCCTTCGGGACAATGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGT
TGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTTAGTAGCCAGCAAGTAAAGTTGGGCACTCTAGGGAGACTGCCAGGGATAACCTGGAGGAAGG
CGGGGATGACGTCAAATCATCATGCCCCTTATGGACTGGGCTACACACGTGCTACAATGGTGGCTACAAAGGGAGGCGACCCTGTGAAGGTGAGCAAAT
CCCAAAAAGGCCATCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATACGTT
CCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCNGAAGTCAGTGACCCAAC
>RK1P
TGCAAGTCGAACGGAGTTTATATTATGAAGTCTTCGGATGAATTGATATAAACTTAGTGGCGGACGGGTGAGTAACGCGTGGGTAACCTGCCTCATACAG
GGGGATAACAGCCGGAAACGGTTGCTAAAACCGCATAAGCGCACAGCAGCGCATGCTGCAGTGTGAAAATATTTTATAGGTATGAGATGGGCCCGCGTC
TGATTAGCTAGTTGGTGAGGTAACTGCTCACCAAGGCGACGATCAGTAGCCGGCTTGAGAGAGTGACCGGCCACATTGGGACTGAGACACGGCCCAAAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAGCGACGCCGCGTGAGTGAAGAAGTATTTCGGTACGTAAAGCTCTA
TCAGCAGGGAAGAAAATGACGGTACCTGACTAAGAAGCCCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGGGCAAGCGTTATCCGGATTT
ACTGGGTGTAAAGGGAGCGTAGGTGGCAAGTCAAGTCAGATGTGAAAGCCCGGGGCTCAACTCCGGGATTGCATTTGAAACTGGTTTGCTAGAGTGCAG
GAGAGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACTGTAACTGACACTGAG
GCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGAGCAAAAGCTCTTCGGTGCCG
TCGCAAACGCATTAAGTATTCCACCTGGGGAGTACGTTCGCAAGAATGAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTT
AATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATCCCATTGACCGGTCCTTAACCGGACCTTTCCTTCGGGACAATGGAGACAGGTGGTGCATG
GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTCCTTAGTAGCCAGCAAGTAAAGTTGGGCACTCTAGGGA
GACTGCCAGGGATAACCTGGAGGAAGGCGGGGATGACGTCAAATCATCATGCCCCTTATGGACTGGGCTACACACGTGCTACAATGGTGGCTACAAAGG
GAGGCGACCCTGTGAAGGTGAGCAAATCCCAAAAAGGCCATCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGCCGGAATCGCTAGTAATCGC
GAATCAGCATGTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGGGAGTCGGAAATGCCCGAAGTCAGTGACCCAACCGTAA
GGAGGGAGCTGCCGAAGGTGGA
>Neu4
TGCAGTCGAGCGAGAAGAGCTCCTTCGGGAGTAATTCTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTGCCTTAGTGAGGGGGATAGCCTCCCG
AAAGGGAGATTAATACCGCATAACATTATTCTATCGCATGATAGAATAATCAAAGGAGCAATCCGCACTAAGATGGGCCCGCGGCGCATTAGCTAGTTGGT
GAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCA
GCAGTGGGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGAGCGATGAAGGTCTTCGGATTGTAAAGCTCTGTCTTTAGGGACGAT
AATGACGGTACCTAANGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGATTTACTGGGCGTAAAGA
GTATGTAGGCGGATATTTAAGTCAGATGTGAAATTCCCGGGCTTAACCTGGGCGCTGCATTTGATACTGGATATCTAGAGTGTGGGAGAGGAAAGCGGAA
TTCCTAGTGTAGCGGNGAAATGCGTAGAGATTAGGAAGAACACCAGTGGCGAANGCGGCTTTCTGGACCATAACTGACGCTGANATACGAAAGCGTGGG
GAGCAAACANGANTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTCGGGGGTTACCACCCTCGGTGCCGCAGCAAACGCAATAAGT
ATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGACCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGA
AGAACCTTACCTAGACTTGACATCTCCTGAATTACTCGTAATGGAGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTC
GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCGTTAGTTGCTACCATTAAGTTGAGCACTCTAACGAGACTGCCGCGGTTAACGTGGAG
GAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCCGGTACAACGAGATGCAAACCCGTGAGGGGGAG
CCAAACTTCAAAGCCGGTCCCAGTTCGGATTGTAGGCTGAAACTCGCCTACATGAAGTCGGAGTTGCTAGTAATCGCGAATCAGCATGTCGCGGTGAATA
CGTTCCCGGGTCTTGTACACACCGCCCGTCACACCATGAGAGCCGGTAACACCCGAAGTCCGTGAGGTAACCTTTATGGAGCCAGCGGCCGAAGG
>Neu18
GGCGCTGCATCTGAAACTGTAACTCTTGAGTGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGG
CGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGA
TACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTG
ACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACCGTTCTAGAGATAGN
NCTTCCCTTCGGGGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTGTT
GTTAGTTGCCATCATTAAGTTGGGCACTCTAGCGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTG
GGCTACACACGTACTACAATGGCTATTACAGAGGGAAGCGAAGCCGCGAGGTGGAGCAAATCCCCCAAAGATAGTCCCAGTTCAGATTGCAGGCTGAAA
CTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGA
GTNTGCAACACCNGAAGTCAGTAGTCTAACCGCAAGGAG
>GS2Neu4/2
GGGGACGATAATGACGGTACCCAAGGANCAANCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTNNGGAATTACTG
GGTGTAAAGGGCGCGTAGGCGGGGATGCAAGTCAGATGTGAAATTCCGGGGCTCAACCCCGGCGCTGCATCTGAAACTGTATTTCTTGAGTGCTGGAGA
GGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCG
CGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAG
TTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATT
CGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTT
GTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGC
CGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCG
AGACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTGAAATTCGCNTGCATGAAGTTGGAATTCACATCAATTGCAGATCAC
CTCCTTTCCGGATCACCTCCTTTCNNGNTCACCTCCNTTCTGGATCACCTCCTTTAGGGTTCACCTCCTTTCTGGATCACCTCCTTTCTGGATCACCTCCTT
TCTGGATCACCTCCTTTCTGG
>6/455 *Thermoanaerobacterium thermosaccharolyticum*
AAAACCTGGGCTTACCCGAGGGTATGCATCTGAAACTATATAGCTNNAGTCAAGGAGAGGAGAGCGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGA
TCAGGAAGAATACCAGTGGCGAAAGCGGCTCTCTGGACTTGAACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTC
CACGCCGTAAACGATGGATACTAGGTGTGGGTGAGGAATCATCCGTGCCGGAGTTAACGCAATAAGTATCCCGCCTGGGGAGTACGGCCGCAAGGTTG
AAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCACAGAA
TCGAGTAGAAATACTTGAGTGCCTCGTAAGAGGAGCTGTGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGC
AACGAGCGCAACCCCTGTTGGTAGTTACCAGCGTAAAGACGGGGACTCTACCGAGACTGCCGTGGATAACACGGAGGAAGGTGGGGATGACGTCAAAT
CATCATGCCCTTTATGCCCTGGGCTACACACGTGCTACAATGGCCTGAACAGAGGGCAGCGAAGGAGCGATCCGGAGCGAATCCCAGAAAACAGGTCCC

AGTTCAGATTGCAGGCTGCAACCCGCCTGCATGAAGACGGAGTTGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACAC
ACCGCCCGTCACACCACGAGAGTTTACAACACCCGAAGTCAGTGACCTAACCGAAAGGGAGGAGCTGCCGAAGGTGGGGTAAATGA
>Neu23
AGTCGAGCGAGGAACTTCGGTTCCTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTCATAGAGGGGGATAGCCTCCCGAAAGGGAGATTAA
TACCGCATAACACTGCGGTTTCGCATGAAACTGCAGTTAAAGGAGCAATCCGCTATGAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACGGC
TCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT
ATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCGCGTGGAGGATGAAGGCCTTCGGGTTGTAAACTCTTGTCTTCTGGGACGATAATGACGGTACC
AGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGATGCGTAGGCG
GATATTTAAGTGGGATGTGAAATACCCGGGCTTAACCTGGGTGCTGCATTCCAAACTGGATATCTAGAGTGCAGGAGAGGAAAGCGGAATTCCTAGTGTA
GCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGCTGAGGCATGAAAGCGTGGGGAGCAAACAG
GATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTGGGAGGTACCAAACCTTCCGTGCCGCCGTTAACACATTAAGTATTCCGCCTG
GGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTAC
CTAGACTTGACATCTCCTGAATTACCGGTAATGCGGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGT
TGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTGTTAGTTGCTACCATTCAGTTGAGCACTCTAGCGAGACTGCCGGGGTTAACTCGGAGGAAGGTGG
GGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCGAGTACAAAGAGATGCGATACCGCGAGGTGGAGCCAAACTCA
AAAACTCGTCCCAGTTCGGATTGTAGGCTGAAACTCGCCTACATGAAGCCGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCCCGG
GCCTTGTACACACCGCCCGTCACACCATGAGAGTTGGCAATACCCGAAGTCCGTAGTCTAACCGCAAGGAGGACGCGGCCGAAGGTAGGGTCAGCGAT
TGGGG
>Neu30
TGCAGTCGAGCGAGGACTTCGGTTCCTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTCATAGAGGGGGATAGCCTCCCGAAAGGGAGATT
AATACCGCATAACACTGCGGTTTCGCATGAAACTGCAGTTAAAGGAGCAATCCGCTATGAGATGGGCCCGCGGCGCATTAGCTAGTTGGTGAGGTAACG
GCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACATTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGG
AATATTGCACAATGGGCGAAAGCCTGATGCAGCAACGCCGCGTGGAGGATGAAGGCCTTCGGGTTGTAAACTCTTGTCTTCTGGGACGATAATGACGGT
ACCAGAGGAGGAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTGGGCGTAAAGGATGCGTAG
GCGGATATTTAAGTGGGATGTGAAATACCCGGGCTTAACCTGGGTGCTGCATTCCAAACTGGATATCTAGAGTGCAGGAGAGGAAAGCGGAATTCCTAGT
GTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGCTGAGGCATGAAAGCGTGGGGAGCAAA
CAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATACTAGGTGTGGGAGGTACCAAACCTTCCGTGCCGCCGTTAACACATTAAGTATTCCGC
CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGCGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCT
TACCTAGACTTGACATCTCCTGAATTACCGGTAATGCGGGAAGCCCTTCGGGGCAGGAAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGA
TGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATTGTTAGTTGCTACCATTCAGTTGAGCACTCTAGCGAGACTGCCGGGGTTAACTCGGAGGAAGGT
GGGGATGACGTCAAATCATCATGCCCCTTATGTCTAGGGCTACACACGTGCTACAATGGCGAGTACAAAGAGATGCGATACCGCGAGGTGGAGCCAAAC
TCAAAAACTCGTCCCAGTTCGGATTGTAGGCTGAAACTCGCCTACATGAAGCCGGAGTTGCTAGTAATCGCGAATCAGAATGTCGCGGTGAATACGTTCC
CGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTTGGCAATACCCGAAGTCCGTAGTCTAACCGCAAGGAGGACGCGGCCGAAGGTAGGGTCAGC
GATTGGGG
>Neu21
TGAAGTTTTCGGATGGATTGGTGTGNNGANAGTGGCGAACGGGCGAGTAACGCGTGAGCAACCTACCTCGTACTGGGGGATAACGTCNGGAAACCGG
CGCTAATACCGCATANNACTACNATGTCNCATGGCGTTGTANTCAAAGCTCCGGCGGTACGAGACGGGCTCGCGTCCGATTANCTAGTTGGTAGGGTAA
CGGCCTACCAAGGCGACGATCGGTAGCCGGACTGANAGGTCGGCCGGCCGCATTGGGACTGAGACACGGCCCANACTCCTACGGGAGGCAGCAGTG
GGGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACNCCGCGTGAAGGAAGAAGGCCTTCGGGTTGTAAACTTCTTTGGCAGGGGACNAANAANG
ACGGNACCCTGANAACAAGCCACGGCTAACTACGTGCCANCAGCCGCGGTAATACGTANGTGGCGAGCGTTGTCCGGAATTACTGGNNGTAANGGGC
GTGTANGCGGGGGGACAAGTCAGATGTGAAATACCGCAGCTTAANTGCGGGGCTGCNTNTGAAACTGTTCTTNTTGAGTGTNGGAGAGGAAAGTGGAA
TTCCCAGNGTAGCGGTGAAATGCGTAGAGATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGACAACTGACGNTGAGGCGCGAAAGCGTG
GGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAANCGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATA
AGTATCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAAC
GCGAAGAACCTTACCAGGGCTTGACATCCCCCTGACCGGTGTAGAGATACACCTTCCCTTCGGGGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACTTTGGGTGGGCACTCTAACGAGACTGCCGGCGA
CAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAAGCCG
CGAGGTGGAGCAAATCCCCAAAAGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCACAC
TGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGACAACACCCGAAGTCGGTAGCCTAACCGCGAGGNGGGCGCCG
CCGAAGGTGGGGTTGGTGATTGGGG
>Neu24
TGAAGTTTTCGGATGGATTGGTGTGGTGAGAGTGGCGAACGGGCGAGTAACGCGTGAGCAACCTACCTCGTACTGGGGGATAACGTCGGGAAACCGGC
GCTAATACCGCATAAGACTACNATGTCNCATGGCGTTGTANTCAAAGCTCCGGCGGTACGAGACGGGCTCGCGTCCGATTANCTAGTTGGTAGGGTAAC
GGCCTACCAAGGCGACGATCGGTAGCCGGACTGAGAGGTCGGCCGGCCGCATTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGG
GGAATATTGCGCAATGGGGGAAACCCTGACGCAGCGACGCCGCGTGAAGGAAGAAGGCCTTCGGGTTGTAAACTTCTTTGGCAGGGGACGAAGAAAGA
CGGTACCCTGANAACAAGCCACGGCTAACTACGTGCCANCAGCCGCGGTAATACGTANGTGGCGAGCGTTGTCCGGAATTACTGGGCGTAAAGGGCGT
GTANGCGGGGGGACAAGTCAGATGTGAAATACCGCANCTTAACTGCGGGGCTGCATCTGAAACTGTTCTTCTTGAGTGTCGGAGAGGAAAGTGGAATTC
CCAGTGTAGCGGTGAAATGCGTAGAGATTGGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACGACAACTGACGCTGAGGCGCGAAAGCGTGGGG
AGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGT
ATCCCGCCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAACGCG
AAGAACCTTACCAGGGCTTGACATCCCCCTGACCGGTGTAGAGATACACCTTCCCTTCGGGGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCG
TGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACTTTGGGTGGGCACTCTAACGAGACTGCCGGCGACAAG
TCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAAGCCGCGAG
GTGGAGCAAATCCCCAAAAGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCACACTGCC
GTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGACAACACCCGAAGTCGGTAGCCTAACCGCGAGGGGGGCGCCGCCGA
AGGTGGGGTTGG
>5VG1/PASC
AATGCGTAGAGATTGGGAGGANCACCAGTGGCGAAGGCGGCTTTNTGGACGACAACTGACGNTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAG
ATACCCTGGTAGTCCACGCTGTAAACGATGGATANTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCGCCTGGGGAG
TACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGG
CTTGACATCCCTCTGACCGGTGTAGAGATACACCTTCCCTTCGGGGCAGGGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGG
GTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACTTTGGGTGGGCACTCTAACGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGG
GATGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTGCTACAATGGCGACTACAGAGGGAAGCAAAGCCGCGAGGTGGAGCAAATCCCC
AAAAGTCGTCCCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCACACTGCCGTGAATACGTTCCCG
GGCCTTGTACACACCGCCCGTCACACCATGAGAGCTGACAACACCCGAAGTCGGTAGCCTAACCGCGAGGGGGGCGCCGCCGACAGGTGGGGTTGGT
GA
>1a *Clostridium clariflavum*

AACCCNGGGCGCTGCATCTGAAANCTGTATTTCTTGAGTGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAAC
NCCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAA
ACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAA
AGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTA
GAGATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAA
CCCTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTT
ATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAG
GCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACC
ATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGAGAGGAGGGCGCTGCCGAAGGTGGGGCAGATGATTGGGG

>2a *Clostridium clariflavum*

TGGNGNAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTTAGATACCNTGGTAGTCCACGCCGTAAACG
ATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGG
AATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAG
ATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT
TGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGT
CCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTG
AAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGA
GAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGAGAGGAGGGCG

>3a *Clostridium clariflavum*

GAGAGGAAAGCGGNAATTCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACACCAGNGNCGAAGGCGGCTTTNNGGACAGTAACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGNNACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCNTTCTGTGCC
GGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTT
TAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCAT
GGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGA
CTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGA
AGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAG
GTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGAGAGG
AGGGCGCT

>4a *Clostridium clariflavum*

CCCGGCGCTGCATCNGAAACTGTATTTCTNGAGTGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACNCCA
GTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGNTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGA
TGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGA
ATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGAT
AGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
TTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCC
TGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGAGACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTGAA
ACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGA
GTCTG

>5a *Clostridium clariflavum*

GAAACTGTATTTCTGAGTGCTGGAGAGGAAAGCGGAATTCNTAGTGGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACACCANTGGCGAAGGCGNCTT
TCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCNTGGTAGTCCACGCCGTAAACGATGGATNNTAGGTGTAG
GAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCTCTTCGG
AGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGCCATC
ATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTA
CTACAATGGCTGCTACAAAGGGAAGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTGAAACTCGCCTGCATGAA
GTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGA
AGTCAGTAGTCTAACCGAGAGGAGGGCGCTGCCGAAGGTGGGGCAGATGATTGGGGTGA

>6a *Clostridium clariflavum*

ACTGGGTGTAAAGGGCGCGTAGGCGGGGATGCAAGTCAGATGTGAAATTCCGGGGGCTCAACCCNGGCGCTGCATCTGAAACTGTATTTCTTGAGTGCTG
GAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTNCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGC
CGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGT
TTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCA
TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAG
ACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGG
AAGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCA
GGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGAGAG
GAGGGCGCTGCCGAAGGTGGGGCAGATGATTGG

>7a *Clostridium clariflavum*

CAACCCCGGCGCTGCATCTGAAACTGTATTTCTGAGTGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAGGAACNC
CAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAAC
GATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAG
GAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGA
GATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACC
CTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTAT
GTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGAGACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAGGC
TGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCAT
GAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGAGAGGAGGGCGCTGCCGAAGGTGGGGCAGATGATTGGGGNG

>8a *Clostridium clariflavum*

AACCCNGGCGCTGCATNTGAAACTGTATTTCTNGAGTGCTGGAGAGGAAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGAGANTAGGAGGAACA
CCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGNTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTNCACGCCGTAAA
CGATNGATACTAGGTGTAGGAGGTATCGACCCNTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAA
AGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTA
GAGATAGGGCTTCTCTTCGGAGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAA
CCCTTGTTGTTAGTTGCCATCATTAAGTTGGGCACTCTAACGAGACTGCCGGTGATAAATCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTT
ATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGAGACCGTGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCAGATTGCAG

GCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACC
ATGAGAGTCTGCAACACCCG
>Neu19
TCGAGCGAACCAATAAGAAGCTTGCTTTTTGTTGGTTAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCTGTAAGACCGGGATAACTCCGGGAA
ACCGGTGCTAATACCGGATAGATTATCTTTCCGCCTGGAGAGATAAGGAAAGATGGCTTTTGCCATCACTTACAGATGGGCCCGCGGCGCATTAGCTAGT
TGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGA
GGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGCGAAGAAGGTCTTCGGATCGTAAAGCTCTGTTGTTAGGG
AAGAACAAGTATCGGAGGAAATGCCGGTACCTTGACGGTACCTGACGAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCA
AGCGTTGTCCGGATTTATTGGGCGTAAAGCGCGCGCAGGCGGTCCTTTAAGTCTGATGTGAAATCTTGCGGCTCAACCGCAAGCGGTCATTGGAAACTG
GGGGACTTGAGTGCAGAAGAGGAAAGCGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGT
CTGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTGGAGGGTT
TCCGCCCTTCAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGC
GGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTCNNGNCCGCCCTGGAGACAGGGTCTTCCCTTCGGGGACAG
GATGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGGTTCTAGTTGCCAGCATTCAGT
TGGGCACTCTAGAGCGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAA
TGGATGGTACAAAGGGCAGCGAAGCGGCGACGCATGAGCGAATCCCAGAAAACCATTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCC
GGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAG
TCGGTGAGGTAACCGTAAGGAGCCAGCCGCCGAAGG
>K/GRsteril6c
TTNTNGGGCGTAAAGCGCGCGCAGGCGGTCCNTTTAAGTCTGATGTGAAATCTTGCGGCTCAACCCGCAAGCGGTCATTGGAAACTGGGGGACTTGAG
TGCAGAAGAGGAAAGCGGAATTNCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGTCTGTAACTGA
CGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTNCACGCCGTAAACGATGAGTGCTAAGTGTTGGAGGGTTTCCGCCCTT
CAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGGTCGCAAGACTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCTCCTGACCGCCCTGGAGACAGGGTCTTCCCTTCGGGGACAGGATGACAG
GTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGGTTCTAGTTGCCAGCATTCAGTTGGGCACT
CTAGAGCGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATGGT
ACAAAGGGCAGCGAAGCGGCGACGCATAAGCGAATCCCAGAAAACCATTCTCAGTTCGGATTGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCT
AGTAATCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGAGG
TAACCGCAAGGAGCCAGCCGCCGAAGGTGGGACAGATGATTGGGGNGAAGTC
>1 *Clostridium thermocellum*
TAACCCNGGNNCTGCNTNTGAAACTGTATCTCTTGAGTGCTGGAGAGGAAAGCGGANTTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACA
CCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAA
GGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAG
AGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
CCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCTAGCGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCC
CTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGC
AGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAC
NCCATGAGAGTNTGCAACACCNGAAGTCAGTAGTNTANC
>2 *Clostridium thermocellum*
AAACTGTATNTCTTGAGTGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTT
CTGGACAGTAACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGG
AGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGC
ACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGG
GCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCA
CGTTAAGGTGGGCACTCTAGCGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACAC
GTACTACAATGGCTGCTACAAAGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCAT
GAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACC
CGAAGTCAGTAGTCTAACCGCAAGGAGGGCGCTGCCGAAGGTGGGGC
>3 Clostridium thermocellum
AAAGGGCGCGTAGGCGGGGATGCAAGTCAGATGTGAAATTCNGGGGCTTAACCCCGGNGCTGCATNTGAANCTGTATNTNTTGAGTGCTGGAGAGGAA
AGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGNNACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGAGGCGCGAA
AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCCGGAGTTAA
CACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTTTAATTCGA
AGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGGTGCATGGTTGTC
GTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCTAGCGAGACTGC
CGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAAGGGAAGCG
ATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATGGCAGGTCA
GCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACC
>4 *Clostridium thermocellum*
4CTGGGTGTAAAGGGCGCGTAGGCGGGGATGCAAGTCAGATGTGAAATTCNGGGGCTTAACCCNGGCGCTGCATNTGAAACTGTATCTCTTGAGTGCT
GGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCNTTCTGTGC
CGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGT
TTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGGTGC
ATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCTAG
CGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAA
AGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAAT
GGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCG
CAAGGAGGGCGC
>5 *Clostridium thermocellum*
ACTGGGTGTAAAGGGCGNGTAGGCGGGGATGCAAGTCAGATGTGAAATTCCGGGGCTTAACCCCGGCGCTGCNTNTGAAACTGTATCTCTTGAGTGCT
GGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCTGA
GGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTGCC
GGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTGGTT
TAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGGTGCA
TGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCTAGC
GAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTACAAA

GGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTAATG
GCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACCGC
AAGGAGGNCNCTNCCGAAGGTGGGGCAGAN

>6*Clostridium thermocellum*

GAAACTGTATNTCTTGAGTGCTGGAGAGGAAAGCGNAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTT
TNTGGACAGTAANNGACGNTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAG
GAGGTATCGACCCCTTCTGTGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCG
CACAAGCAGTGGAGTATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGG
GGCAGAGGAGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGC
ACGTTAAGGTGGGCACTCTAGCGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACA
CGTACTACAATGGCTGCTACAAAGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAANCTNNCCTGCA
NGAAGTCGGAATTGCTAGTAATGGCAGGTCAGCATACTGCCG

>7*Clostridium thermocellum*

ATTACTGGGTGTAAAGGGCGCGTAGGCGGGGATGCAANTCAGATGTGAAATTCCGGGGCTTAACCCCGGNGCTGCATNTGAAACTGTATNTCTTGAGTG
CTGGAGAGGAAAGCGGAATTCNTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACGCT
GAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTGTG
CCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGTG
GTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGGT
GCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCT
AGCGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTAC
AAAGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTA
ATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACC
GCAAGGAGGGCGCTGCCGAAGGTGGGGCAGATGATTGG

>8 *Clostridium thermocellum*

AATTACTGGGTGTAAAGGGNGCGTAGNCGGGGATGCAANTCAGATGTGAAATTCCGGGGCTTAACCCCGGNGCTGCATNTGAAACTGTNTNTCTTGAG
TGCTGGAGAGGAAAGCGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTTCTGGACAGTAACTGACG
CTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGGATACTAGGTGTAGGAGGTATCGACCCCTTCTG
TGCCGGAGTTAACACAATAAGTATCCCACCTGGGGAGTACGGCCGCAAGGTTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCAGTGGAGTATGT
GGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGGCTTGACATCCCTCTGACAGCTCTAGAGATAGGGCTTCCCTTCGGGGCAGAGGAGACAGGTGG
TGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCGTTAGTTGCCAGCACGTTAAGGTGGGCACTCT
AGCGAGACTGCCGGCGACAAGTCGGAGGAAGGTGGGGACGACGTCAAATCATCATGCCCCTTATGTCCTGGGCTACACACGTACTACAATGGCTGCTAC
AAAGGGAAGCGATACCGCGAGGTGGAGCAAATCCCCAAAAGCAGTCCCAGTTCGGATTGCAGGCTGAAACTCGCCTGCATGAAGTCGGAATTGCTAGTA
ATGGCAGGTCAGCATACTGCCGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAGAGTCTGCAACACCCGAAGTCAGTAGTCTAACC
GCAAGGAGGGCGCTGCCGAAGGTGGGGCAGATGA

v. Amplikonsequenzierung des BM1T Reaktors

>1/11a

CCAATCATCTGCCCCACCTTCGGCGGCTCCCTCCTAAAAGGTTGGGTCACCGACTTCGGGTGTGACAAACTTTCGTGGTGTGACGGGCGGTGTGTACAA
GGCCCGGGAACGTATTCACCTCGGTATGCTGACCCGAGATTACTAGCGATTCCGACTTCATGCAGTCGAGTTGCAGACTGCAATCCGAACTGAGAGCAG
CTTTTTGGGATTTGCTCCAGGTCACCCCTTCGCTCCCCTCTGTACTGCCCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCATGATGACTTGACGT
CAACCTCACCTTCCTCCGACTTGTCGCCGGCAGTCCCTCTAGAGTGCCCAACTTAATGATGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAA
CCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTCTCCTCTGCCCAAAGGGAAGGCCTATCTCTAGACCCGTCAGAGGGATGTCAAG
CCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACT
CCCCAGGCGGAACACTTACTGCGTTAACTGCGGCACAGGAGGGGTCGATACCCCCTACACCTAGTGTTCATCGTTTACGGCTGGGACTACCAGGGTATC
TAATCCTGTTTGCTCCCCCAGCTTTCGCGCCTCAGCGTCAGTTACAGTCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCAGATATCTACGCATTTCACC
GCTACACCTGGAATTCCACTCTCCTCTCCTGCACTCTAGCAACGTAGTATCAGGCTATGTGCGTCGGTTGAGCCGATGCCTTTCACGCCTGACTTACGTCA
CCGCCTACGCGCCCTTTACGCCCAATGATTC

>1/12a

CCCCAATCATCCGCCCCACCTTCGACGGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCGGGTGTTGCGAACTTTCGTGGTGTGACGGGCGGTGTGTA
CAAGGCCCGGGAACGTATTCACCGCGGTATGCTGACCCGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAG
TGGCTTTCTGGGATTCGCTCCGCCTCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAGCACGTGTGTAGCCCAGATCATAAGGGGCATGATGATTTGAC
GTCATCCCCACCTTCCTCCGACTTGTCGCCGGAAGTCCCTTTAGAGTCCCCAACTTAATGATGGCAACTAAAGGCAGGGGTTGCGCTCGTTGCGGGACTT
AACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGtcaccTTGCTCCGAAGAGAAACCACATCTCTGTGGCTGTCAAGGGATGTCAAGC
CCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTC
CCCAGGCGGGGTACTTATTGCGTTAACTCCGGCACAGAAAGTATCGAAACTCCCTACACCTAGTACCCATCGTTTACGGCTAGGACTACCANGGTATCTAA
TCCTGTTTGCTCCCCACGCTTTCNTACCTCANCGTCAGTCCAAGG

>1/13a

CCCCAATCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACA
AGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTG
CTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGT
CGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACT
TAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAA
GCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTAC
TCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTTGCTCCCCACGCTCTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCG
CTACACTAGGAATTCCACTTTCCTCTCCAGCACTCAAGAGACACAGTTTCAGATGCAGTGCCGAGGTTGAGCCCCGGAATTTCACATCTGACTTGCATCCC
CGCCTACACGCCCTTTACACCCAGTAATTCCGG

>1/14a

CCCCAATCATCAACTCCACCTTCGACGGCTGCCTCCTTGCGGTTAGCGCACNGGNTTCGGGTGTAGTTGACTTTCGTGGTGTGACGGGCGGTGTGTACA
AGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCA
GCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGT
CATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTA
ACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAA

GTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTAC
TCCCCAGGCGGGGTACTTATTGCGTTAACTGCGGCACAGAGGTATCGAAACCCCTACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTA
ATCCTGTTTGCTCCCCTAGCTTTCGTGCCTCAGCGTCAGTTATCGTCCAGGCAGTCGCCTTCGCCACTGGTGTTCCTCCCGATATCTACGCATTTCACCGC
TACACCGGGAATTCCACTGCCCTCTCCGACACTCAAGCCTCCCAGTTTCAAATGCAACCCCACGGTTAAGCCGTGGTCTTTCACATCTGACTTAAAAGGCC
GCCTACGCACCCTTTACGCCCAGTAATTCCGGACAACGCTCGCCCCCTACGTTTTACCGCGGCTGCTGG
>1/12
TTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTC
CCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCC
TTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATG
CACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCTTCGAATTAAAC
CACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGGTACTTAATGCGTTAACTACGGCA
CGGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCTAGCTTTCGTGCCTCAGCG
TCAGTTATCGTCCAGGTAGTCGCCTTCGCCACTGGTGTTCCTCCCAATATCTACGCATTTCACCGCTACACTGGGAATTCCACTACCCTCTCCGACACTCAA
GCCCTCCAGTTTCGAATGCCATTCCGCGGTTAAGCCGCGGGCTTTCACATCCGACTTAAAGGGCCGCCTACGCACCCTTTACGCCCAGTAATTCCGGACA
ACGCTCGCCCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGGGCTTTCTTCTTGGGTACCGTCAAGCTTTC
>1/14
GAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAA
AGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAG
GAGGTGATCCNGAAAGGAGGTGATCCTGAACGGAGGTGATCCCTAAtCGATGTGAGCCCcAAAGtAGGTGATGtGtaTCCCTGGTCGTAAGGACCATGAT
GACTTGACATCATCCCCACCTTCCTCCAGTTTATCACTGGAAGTCTCctTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTG
CGGGACTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAgGCACATTCTCATCTCTGAAAACTTCCGTG
GATGTCAAGACCAGGTAAGGTTCTTCGCGTTGCATCNAATTAAACCACATGCTCCaCCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGC
GGCCGTACTCCCCAGGCGGTCGACTTAaCGCGTTAGCTCCgGAAGCCACGcCTCAAGGGCACAACCTCCAAGTCGACATCGTTtACgGTGNgacTACCAG
GGTATCTAATCCTGTttGCTCTCCACGCTTTCGCAcCTGA
>1/15
GGTAAGCGCCCTCCCGAAGGTTAAGCTACCTACTTCTTTTGCAACCCACTCCCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCG
TGGCATTCTGATCCACGATTACTAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACTACGACGCACTTTATGAGGTCCGCTTGCTCTC
GCGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCAGTTTAT
CACTGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATTTCACAACACGAGC
TGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGATGTCAAGACCAGGTAAGGTTCTTCGCGTT
GCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGGCGGTCGACTTAACG
CGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCT
TTCGCACCTGAGCGTCAGTCTTCGTCCAGGGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATTTCACCGCTACACCTGGAATTCTACCCC
CCTCTACGAGACTCAAGCTTGCCAGTATCAGATGCAGTTCCCAGGTTGAGCCCGGGGATTTCACATCTGACTTAACAAACCGCCTGCGTGCGCTTTACGC
CCAGTAATTCCGATTAACGCTTG
>1/16
CTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACC
TGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTT
CCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTC
CTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCAT
GCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCNAATTAAA
CCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGC
ACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGC
GTCAGTTTCaNgCCAGAGAGCCGCCTTCGCCACTGNtGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCA
AGATATACAGTTTCAAATGCAgCTCACC
>1/19
AACCTTCGACGGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCGGGTGTTGCGAACTTTCGTGGGGTGACGGGCGGGGGTGTACAAGGCCCGGGAAC
GTATTCACCGCGGTATGCTGACCCGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTCTGGGATT
CGCTCCGCCTCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAGCACGTGTGTAGCCCAGATCATAAGGGGCATGATGATTTGACGTCGTCCCCACCTT
CCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACGGGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATC
TCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAG
GTTCTTCGCGTTGCTTCNAATTAAACCACATACTCCaCTGcTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGG
GATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGNACTACCAGGGTATCTAATCCTGTTTG
CTCCCCACGCTCTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTC
>1/110
TTGGACGGCTCCCTCCCTTACGGGTTACTaAaCCGTTCtTTTGtGTTGCCAACTTTGGTGGGGTGACGGGCGGGGGTGTACGAGGCCCGGGAACGTATTCA
CCGCGGTATGCTGACCCACTATTACTAaTTATTACtACTTGATGCAgGCGAGTTGACTCCTGCCATCCGAACTGAGAGTGGCTTTCTGGGATTCTGTCCGCC
TCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAGCACGTGTGTACCCCAGATCATAGGGGGCATGAAGATTTGACGTCCTCCCCATTTTCCTCCGACTT
GTCGCCGGCAGTCCCTTTAGAGTCCCGAACTTAATGCTGGCAACTAAGAGAAAGGGTGGCGCTCGTTGCGGAATTTAACCCAACATCTCACGACACAAGT
TGACAACAACCATGcaCCACCGGTCACCCaNTCCCGAAGGAgGgtTATGTTTCCATGGNTTGTTAGGGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGC
TTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGgTACTTAATGCGT
TAACTACGGAACGgATGGTATC
>1/113
GGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCtGGTGTTGCGAACTTTCGTGGGGTGACGGGCGGGGTGTACAAGGCCCGGGAACGTATTCACCGC
GGTATGCTGACCCGCGATTACTAaCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCG
CGGCTTCNCTGCCCTCTGTANCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGT
CACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAG
CTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAgCATTCAAGGGATGTCAAGcCCTGGTAAGGTTCTTCGCGTTGC
TTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGTGT
TAACTACGGaACACAagGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACgGCNTG
>1/114
TTCGGCGGCTCCCTCCTAAAAGGTTGGGTCACCGACTTCGGGTGTGACAAACTTTCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCA
CCTCGGTATGCTGACCCGAGATTACTAGCGATTCCGACTTCATGCAGTCGAGTTGCAGACTGCAATCCGAACTGAGAGCAGCTTTTTGGGATTTGCTCCA
GGTCACCCCTTCGCTCCCCTCTGTACTGCCCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGA
TTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCAGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGC
GTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAA

TGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAG
CTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTC
TCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCC
CAGTGATTCCGGACAACGCTTgCCcTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCCCAGGTACCGTCA
>1/115
TTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCT
GCGATTACTAgCGATTCCGGCTTCATGCAGGCGAGTTGCAGCcTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCC
CTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCT
TAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGC
ACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACC
ACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGGTACTTAATGCGTTAACTACGGCAC
GGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCTAGCTTTCGTGCCTCAGCGT
CAGTTATCGTCCAGGTAGTCGCCTTCGCCACTGGTGTTCCTCCCAATATCTACGCATTTCACCGCTACACTGGGAATTCCACTACCCTCTCCGACACTCAA
GCCCTCCAGTTTCGAATGCCATTCCGCGGTTAAGCCGCGGGCTTTCACATCCGACTTAAAGGGCCGCCTACGCACCCTTTACGCCCAGTAATTCCGGACA
ACGCTCGCCCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGGGCTTTCTTCTTGG
>1/116
ACGGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCGGGTGTTGCGAACTTTCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACC
GCGGTATGCTGACCCGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTCTGGGATTCGCTCCGC
CTCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAGCACGTGTGTAGCCCAGATCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGACT
TGTCGCCGGCAGTCCCTTTAGAGTCCCCAACTTAATGCTGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAG
CTGACGACAACCATGCACCACCTGTCACCCAGTCCCGAAGGAGGGCCATGTTTCCATGGCTTGCTAGGTGATGTCAAGACCTGGTAAGGTTCTTCGCGTT
GCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGGTACTTAATG
CGTTAACTACGGCACGGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCTAGCT
TTCGTGCCTCAGCGTCAGTTATCGTCCAGGTAGTCGCCTTCGCCACTGGTGTTCCTCCCAATATCTACGCATTTCACCGCTACACTGGGAATTCCACTACC
CTCTCCGACACTCAAGCCCTCCAGTTTCGAATGCCATTCCGCGGTTAAGCCGCGGGCTTTCACATCCGACTTAAAGGGCCGCCTACGCACCCTTTACGCC
CAGTAATTCCGGACAACGCTCGCCCCCTACGTATTACCGCGGCTGCTGGCAc
>1/117
CTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCA
CGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCAC
CTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAT
TTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGC
GTTGCTTCgAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCcCAGGTGgGATACTTATT
GTGTTAACTACgGCACAGAAgggTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCaGGGTATCTAATCCTGTTTGCTCCCCACGCTC
TCGCGNCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGG
>1/118
CTTGCGACGGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCGGGTGTTGCAAACTTTCGTGGGGTGACGGGCGGGGTGTACAAGGCCCGGGAACGTA
TTCACCGCAtTATGCTGACCTGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGATCTGCTTTCTGGGATTCGCT
CCGTGTCGCCACTTCGCTGCCCTCTGTACAGACCATTGTAGCACGTGTGTAGCCCAGGACATAGGGGGCATGATGATTTGACGTCATCCCCACCTTCCTC
CGGTTTGTCACCGGCAGTCTCATTAGAGTGCCCAACTTAATGATGGCAACTAATGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACaACAACCATGCACCACCTGTCTCCTTGCTCCGAAgAgAAACCATGTTTCCATGGcGGTCAAGGGATGTCAAGCCCTGGTAAGGtTCTTCGC
GTTGCTTCgAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGGTACTTAT
TGCGTTAACTGCGGCACAGATGGTATCNAAaCCACcTACACCTAGTacCCATCGTTTACGGCTAGGACTACCA
>1/21a
CCCCAATCATCTGCCCCACCTTCGGCGGCTCCCTCCTAAAAGGTTGGGTCACCGACTTCGGGTGTGACAAACTTTCGTGGTGTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACCTCGGTATGCTGACCCGAGATTACTAGCGATTCCGACTTCATGCAGTCGAGTTGCAGACTGCAATCCGAACTGAGAGCA
GCTTTTTGGGATTTGCTCCAGGTCACCCCTTCGCTCCCCTCTGTACTGCCCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCATGATGACTTGACG
TCAACCTCACCTTCCTCCGACTTGTCGCCGGCAGTCCCTCTAGAGTGCCCAACTTAATGATGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTA
ACCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTCTCCTCTGCCCAAAGGGAAGGCCTATCTCTAGACCCGTCAGAGGGATGTCAA
GCCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTA
CTCCCCAGGCGGAACACTTACTGCGTTAACTGCGGCACAGGAGGGGTCGATACCCCCTACACCTAGTGTTCATCGTTTACGGCTGGGACTACCAGGGTA
TCTAATCCTGTTTGCTCCCCACGCTTTCGTGCCTCAGCGTCAGTCCAAGGCCAGGAAGCCGCCTTCGCCACTGGTGTTCCTCCTGATATCTACGCATTTCA
CCGCTACACCAGGAATTCCGCTTCCCTCTCCTTAACTCAAGCTAAGCAGTTTCAAGTGCTTACATGGGTTGAGCCCATGCCTTTCACACCTGACTTACTAAG
CCGCCTACGCACCCTTTACGCCCAGTAATTCCGGACAACGCTCGCCCCCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCGGGGCTTC
>1/22a
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCT
GCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACA
CTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCANATTTGACTTATATATCCGCCTACGTACCCTTTA
CGCCCAGTGATTCCGGACAACGCTT
>1/23a
CCACTTCCCCCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGG
GAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGA
GATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCAC
CTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCGACAT
CTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGT
AAGGTTCTGCGCGTTGCATCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGG
CGGGGTACTTAATGCGTTAACTACGGCACGGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCTGT
TTGCTCCCCTAGCTTTCGTGCCTCAGCGTCAGTTATCGTCCAGGTAGTCGCCTTCGCCACTGGTGTTCCTCCCAATATCTACGCATTTCACCGCTACACTG
GGAATTCCACTACCCTCTCCGACACTCAAGCCCTCCAGTTTCGAATGCCATTCCGCGGTTAAGCCGCGGGCTTTNACATCCGACTTAAAGGGCCGCCTAC
GCACCCTTTACGCCCAGTGATTCCGGACAACGCTTNCCCTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCCCAGGTACCG
>1/24a

TTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCG
CAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGT
CACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTT
CACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGC
TGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTT
GCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATG
CGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTT
TCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTC
CTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCG
>1/25a
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCT
GCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACA
CTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTAC
GCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCCCAG
>1/26a
ACAGACCGTGCTAAATCTGACCGTGAGGAACACCGCCGTTCTTAAAAATCGCTGACATTANAAACCGCTGTTCNTANAAATCGATGTTCTTANNcNTATGC
TGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTT
CGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCT
GTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAANGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACA
ACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCNNGTCTCGGTAAGGTTCTGCGCGTTGCATCGA
ATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACTCCCCAGGCGGGGTACTTAATGCGTTAACT
ACGGCACGGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACC
TGAGCGTCAGTTTCAGGCCAGA
>1/27a
CAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGT
CACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTG
TCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGA
GCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGC
GTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAA
TGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAG
CTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTC
TCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTT
>1/28a
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCT
GCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACA
CTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTAC
GCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCG
>1/29a
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGACCAGCTTTCTGGGATTTGCT
CCAGATCACTCCTTCGCTTCCCTCTGTACTGGCCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGGCATTACCCGCTGGCAACTAAGGATAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGTCACGGGATGTCAAGTCTCGGTAAGGTTCTT
CGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACAC
TTAATGCGTTTGCTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTCGCTCCCC
CAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCCGCTCACCGGTTGAGCCGATGCCTTTNACATTTGACTTATATATCCGCCTGCGTACCCTTTA
CGCCCAGTAATTCCGG>1/210aCCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGT
GTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACT
GGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGAT
TTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCG
GGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGG
GATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGC
GACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCG
GGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCA
TTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCG
>1/211a
TCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCC
GGGAACGTATTCACGGCAGTATGCTGACCTGCCATGgaaAATCgaTGTtcttacatGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGG
ATTTGCTCCACCTCGCGGCTTCGCTGCCCtNtgTaacNaccgTNGTTatACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCT
TCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATC
TCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAG
GTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGG
GATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACC
>1/212a

CAAATGACTTCTGGTGCATCTAANTCGGNTGGGgTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAG
CGATTCCGGCTTCATGCAGGCGAGTtgcAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGA
CCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCC
GACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTC
TCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCA
CCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTC
GAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCAGG
CCAGAGAGCCGCCTTCGCCACTGGGgTTCTTCCTGATATCTACGCATTC
>1/213a
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
ACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCT
GCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACA
CTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTAC
GCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGG
>1/214a
CCCAATCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTNGCNGACTCTCATGGTGTGACGGGCGGTGTGTACAA
GGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCT
TTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCG
TCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAA
CCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCC
CTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCC
CCAGGTGGGATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCAGGGTATCTAAT
CCTGTTTGCTCCCCACGCTCTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTA
CACTAGGAATTCCACTTTCCTCTCCAGCACTCAAGAGACACAGTTTCAGATGCAGTGC
>1/215a
CCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATT
CACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTC
CATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCG
GCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTG
CGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACAC
TTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCC
CAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCC
GCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTAC
GCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACGTGAGCATCGACAAATGACTTCTGGTGCATCTAACTCGGGTGGTGT
GACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAA
TCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGC
ATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGC
TCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAG
GCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAG
TCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGAC
TACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATC
TACGCATTCCACCGCTACACCAGGAATTCCACTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACAT
TTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGG
CTTCCTTCCCAGGa
>1/53
CTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAgTATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCATCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCC
CTCTGTACTGACCATTGTATCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCT
TAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGC
ACCACCTGTCTCCGTGTGCCCGAAgGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCgAATTAAACCA
CATGCTCCACCGCTTGTGCGGGCCCCCGTCNaTTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCNGCACT
GAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGgACTACCGgGGTATCTAATCCCGTTTGCTACCCCANCTTTCGTACCTGANCGTCA
GTTTCAgGACAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCAcCGCTACACCA
>1/57
TTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTC
CCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCC
TTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATG
CACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAAC
CACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCA
CTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCG
TCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCACTCTCCTCTCCTGAACTCA

AGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGACAA
CGCTTGCCcTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCCCAGGTACCGTCAGCTCATG
>1/58
TTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAaTCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTC
CCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATgAtGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCC
TTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATG
CACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAAC
CACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCA
CTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCG
TCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCA
AGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGACAA
CGCTTGCCCTCTACGTATTACCG
>1/59
CCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGAC
CTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCT
TCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCT
CCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
TGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAA
ACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGG
CACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAG
CGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACT
CAAGAGACACAGTTTCAGATGCAGTGCCGAGGTTGAGCCCCGGAATTTCACATCTGACTTGCATCCCCGCCTACACGCCCTTTACNCCNAGTAATTCCGG
ATAaCGCTCGCCACCTACGTATTACCGCGGCTGCTGG
>1/510
TTGGGAAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCAACTCTCAGGGGGGGACGGGCGGGGGTGTACAAGGCCCGGaAACGTATTCC
CGGGAGTATGCTGACCTGCCATTACTAGCAATTCCGAcTTCATGCaGGCGANTTTCaGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCAC
CTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCATGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAT
TTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACgTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACAC
GAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAACGGTCAAgGgATGTcAAGccNTGGTAAGGNTCTTCGCGT
TGCTTCNAATTATACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCaNGTGGGGATACTTATT
GTGTTAACTAC
>1/71
GGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGAC
GACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCAT
CGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTA
ACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTA
CCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCC
TGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATT
CCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCCCAGGTACCGTCAGCTCATGGTCATTCCCTACCA
TGAGGGTTCTTCCCTGAGGACAGAACTTTACGACCCGAAGGCCTTCTTCGTTCACGCGGCGTTGCTCCGTCAGGCTTTCGCCCATTGCGGAAGATTCCCC
ACTGCTGCCTCCCGTAGGAGTCTGGGCCgggTCTCAGTCCCAGTGTGGCTGATCATCCTCTCAGACCAGCTACGGATCGTTGCCCTGGTGAGCCGTTACC
CCACCAGCTAGCTAATCCGACG
>1/72
TTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGgNGGtGtGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTA
CTAGCAATTCCGACTTCATGCAGGCGAGTTTCaGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTA
GCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAG
TCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCAC
CTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATG
CTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAG
GATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTT
CAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATA
CAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTgCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCG
>1/74
TTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGtGTGTACAAGGCCCGGGAACGTATTCAC
CGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCAT
GTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCT
TTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGA
GCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGC
GTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAA
TGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGC
TCTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTT
CCTCTCCAGCACTCAAGAGACACAGTTTCAGATGCAGTGCCGAGGTTGAGCCCCGGAATTTCACATCTGACTTGCATCCCCGCCTAcccGCCCTTTANcCC
NAGTAATTCCGGaa
>1/75
TCATCTGTCCAACCTTCGACGGCTCCCTCCTTGCGGTTAGGCCACCGGCTTCGGGTGTTACAGACTCTCGGGGTGGGACGGGCGGGGTGTACAAGACC
CGGGAACGTATTCACCGCGACATGCTGATTCGCGATTACTAGCGATTCcGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAAAGTGGCTTTT
TGGGTTTCGCTCCgCCTCGCGGCTTCGCCTCCCTTTGTACCACCCATTGTAGCACGTGTGTAGCCCAGGCCATAAGGGGCATGATGATTTGACGTCATCC
CCACCTTCCTCCAGTTTGTCACTGGCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAA
CATCTCACGACACGAGCTGACgACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGAAACTCACTATCTCTAGCGATGGCGgAGGAATGTCAAGGC
CTGgTAAGGTTCTTCGCGTTGCTTCaAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCC
CANGCGGANTACTTATTGCGTTATCTTCANCACCGATCTTGCGACCGACACTTANTACTCATCGTTTACgGCGTGNACTACCANGGTATCTAATCCTGTTT
GCTCCCCACGCTTTCGTGcCTCAGCGT
>1/76
TCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTG
CGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCC
CTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCT
TAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGC

ACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACC
ACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCAC
TGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGT
CAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAA
GATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGACAAC
GCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACGTAGt
>1/77
CTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGGGGTGTACAAGGCCCGGGAACGTATTCA
CGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCAC
CTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAT
TTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGC
GTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGgGATACTTAT
TGTGTTAACTACGGCACAGAANGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACG
CTCTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGGAATTCCACT
T
>1/79
CCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGAC
CTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCT
TCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCT
CCTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATG
CACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCAC
ATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTG
AAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCA
GTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGA
TATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATTGGGCCGCCTACGCGCCCTTTACGCCCAGTAATTCCGGACAACG
CTCGCCCCCTACGTGTTACCGCGGCTGCTGG
>1/710
TTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTC
CCTCTGTACTGACCATTNNAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTC
CTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCAT
GCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAA
CCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGC
ACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGC
GTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTC
AAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGACA
ACGCTT
>1/711
CTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACC
TGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTT
CCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTC
CTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCAT
GCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCgAATTAAA
CCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGC
ACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGC
GTCAGTTTCagGCCAGAGAGCCGCCTTCGCCACTGGtGTTCTTCCTGATATCTACGCATTCCACCGCTACACCANGGAATTCCGCTCTCCTCTCCTGA
>1/713
TTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGGGGGGACGGGCGGGGGTGTACAAGGCCCGGGAACGTATTCA
CGGGAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCAC
CTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGAT
TTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACgTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACAC
GAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAACGGTCAAGGgATGTCAAGCCCTGGTAAGGTTCTTCGCG
TTGCTTCNAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCagGCgggACACTTAATG
CGTTAACTTCNGCACTGAAgGATTCNAACCCCCAACACCTANTGTCCATCGtTTACNGNTGGGACTACCGGGGTATCTAATCCCGTgTTGCTACCCCAGCT
TTCGTACCTGA
>1/714
CTTCGGCATTCGCAAATGAATTCTGGTGCATCTAACTCGGGtGGTGTGACGGGCGGTGTGGAGAAGGCCCGGGAACGTATTCaCCGGGAATGCTGACCT
GCGATTACTAGCGATTCCGGCTTCATGCAGGCNAgTTGCAGCCTGCAATCCGAACTGGGGCTGCTTTTANGGATTTGCTCCACCTCGCGGCTTCGCTGCC
CTCTGTGGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCAACTTCCTCCGATTTGTCACCGGCAGTCTC
GCTAGAGTGCCCACCTCTACTTGCTGGCAACTAAGAACGGGGGTTGCGCTCGTTGCGGGACTTANCCCATCATCTCaCGACaCGAGCTGACNACAACCA
TGcaNCACcTGTCTCCTTGcCCCGAAGGGAAACTGTATTTCAACGGCGGTCaAGGGATGTCAAGCCCTGgtAAGGTTCTTCGCGTTGCTTCgaATTATCNC
NCATACTCcACTGCTTGTGCGGGCCCCCGTCaaTNCCTTTGAGTTTCAGC
>1/718
CTTCGACGGCTGCCTCCTTGCGGTTAGCGCACCGGCTTCGGGTGTAGTTGACTTTCGTGGGGGGACGGGCGGGGGGTACAAGGCCCGGGAACGTATT
CACCTCGGTATGCTGACCCGAAATTACTAGCGATTCCTCCTTCAGGGAGTCAAGTTGCAAACTCCGATCCGAACTGGGACCGGCTTTTTGCGATTCGCTC
CCCCTCGCGGGTTCGCTGCGCTCTGTACCGGCCATTGTAGCACGTGTGTAGCCCAGGTCATAAGGGGCATGATGACTTGACGTCTTCCCCACCTTCCTC
CGATTTGTCACCGGCAGTCTCCTTAAAGTCCCCGGCTTTACCCGCTGGCAACTAAGGATAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACG
ACACGAGCTGACGACAGCCATGCACCACCTGTCTCCTCTGTCCCGAAgGAAAGATACATCTCTGCATCCGTCagAGGGATGTCAAGACCTGGTAAGGTTC
TTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGCGGGAT
ACTTAATGCGTTGGCTGCGGcACGGAAGGGGTCGATACCTCCCACACCTAGtATCCATCGtTTACGG
>1/719
TTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGtGTGACGGGCGGtGTGtACAAGGCCCGGGAACGTATTCACGGCagtATGCTGACCTG
CcATTACTAGCAATTCcgaCTTCatGCAGGCGAGTTTCagCcTGCAATCCGAACTGGGGCTGCTTTTAGGgATTTGCTCcaCCTCGCGGCTTCGCTGCCCTCT
GTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGaTTtGACGTCGTCCCCACCTTCCTCCGATTTGTCACcGGCAGTCTCGCTAG
AGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCAC
CACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACaGCGGtcaaGGgatGtcAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTaaaCCACATACT

CCACTGCTTGTGCGGGCCCCCGTCAaTTCCTTTGAGTTTCAGCCTTGCGGccGTACTCCCCaGGTGGGATActtATTGTGTTAAcTACGGCacaGAAGGGGT
CGATACCTCCTACACcTAGT
>1/720
TTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCAC
GGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCT
CGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTT
GTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACG
AGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGaTGTCAAGCCCTGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTG
TGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCT
CTCGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTC
CTCTCCAGCACTCAAGAGACACAGTTTCAGATGCAGTGCcaNGGTTGAGCCCCGGAATTTCACATCTGACTTATTGGGCCGCCTACGCGCCCTTTACGCC
CAGTAATTCCGG
>1/721
GTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACT
GGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGAT
TTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCG
GGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGG
GATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGC
GACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCG
GGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCA
TTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTT
ATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCCGGA
>1/723
ACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATG
CTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACT
TCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGC
TGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGAC
AACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGA
ATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACT
TCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCT
GAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTG
AACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCCCAGTGATTCC
GGACAACGCTTgCccTCTACGTATTACCGCGGCTGCTGGCACGTAGTTAGCCAGGGCTTCCTTCC
>1/724
GCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTA
TGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGC
TTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCG
GCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGAC
GACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCG
AATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGTGTTAAC
TACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTCTCGCG
CCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTCCTCTCCA
GCACTCAAGAGACACAGTTTCAGATGCAGTGCCGAGGTTGAGCCCCGGAATTTCACATCTGACTTGCATCCCCGCCTACACGCCCTTTACACC
>1/725
GGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTG
AGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCA
CCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACA
TCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGG
TAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAG
GCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCG
TTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACC
AGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCACCGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACG
TACCCTTTACGCCCAGTGATTCCGGACAACGCTTGCCCTCTACGTATTACCGCGGCTGCTGGCACG
>1/727
GAGGTGATCCACAAAGGAGGTGATCCAGAAAGGAGGTGAGCCAGAAAGGAAGTGATCCAgAAAGGAGGTGAACCCTAATGGAGGGGAtCCNGAaGTGA
GGTGAGCCCGGGAACgTATTCACGGCagTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCANGCGAGTTTCAGCCTGCAATCCGAACTGGGGC
TGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGAC
GTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGA
CTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCAGTCAAGGGATGTCA
AGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTA
CTCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGNACTACCagGGTATC
TAATCCTGTTTGCTCCCCACGCTCTCGCGCCTCA

vi. Amplikonsequenzierung des BM2T Reaktor

>4.1
TGCATGTATGTNANTTGCGAtCTTCcTAATGACGTTGAGATGATTCTCGTGTATTAGGTCCGGGAANgCtTCcCGCANAAAGCtCTNTGCgATTaCAttcATTC
CGGCTTCGTGCCaGTGATTcGTTGCCTGcATCaGAACTGTAATCAAATTCCCGATATTGGCTCCTTGTCACCACTTCGcTTCCCTCTGNAgTGACcATTGCA
GCCCGTGTGTCTTCCCTCCCATTAcCGGGCATGATGATTTGATTTCTTCCCCACCTTCCTCAAGTTTGTCTCTGGCAGTCTCGTTAGAGTCCCCAACTTAAT
GCTGGCAACTAACGACgAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCaTGCACCACCTGTCTCCTCCGTCC
CCAAAAGGAAACTCACTATCTCTAGcGTTGGCGGAGGGATGTCAaGGCCTGGTAAGGTTCTTCgCGTTGCTTCGAATTaaACCACATGCTCCACCGCTTGT
GCGGATCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAgGCGgGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCTAANCCCCA
ACACCTAGTGTCCATCGTTTACGGCTGGGACTAcCCGGGGTATCTAATTCCCGNTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTTCAGGCCA

>4.2
GGAAATGACTTCTGGTGCATCTAACTCGGGTGGAGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTA
GCGATTCCGGCTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCA
ACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGC
CCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCT
GTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCA
CTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGCGTTAACTTCGGCACTGAAGGATTC
GAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAgCGTCAGTTTCagCC
AGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAGATATACAGTTTCA
AATGCAGCTCAcNGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCCTTTACGCC
>4.3
ACGACGGATCTCTCCCTTACGGGTTAAGTCACCGGCTTCGGGTGTTGTTGACTTTCGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCA
CCGCAGCATGCTGATCTGCGATTACTAGCGATTCCTACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGCTGCTTTCTGGGATTGGCTCCG
GATCGCTCCTTCGCTACCCTCTGTACAGCCCATTGTAGCACGTGTGTAGCCCAGATCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGAC
TTGTCGCCGGCAGTCCCTTTAGAGTCCCCAACTTAATGCTGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGA
GCTGACGACAACCATGCACCACCTGTCACCCAGTCCCGAAGGAGGGCCATGTTTCCATGGCTTGCTAGGTGATGTCAAGACCTGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATG
CGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGaCTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTT
TCGTACCTGAGCGTCAGTTTCAGGCCAGANAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCT
>4.4
GGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCATAAAGGAGGTGATCCATAAAGGAGGTGATCCACAAAGGAGGTGATCCAGAAAGG
AGGTGATCCAGAAAGGAGGTGATCCGNGAAGGACGGGATCCCGATTAGAAGCGATTCCGACTTCATGGAGTCGAGTTGCAGACTCCAATCCGGACTACG
ACGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTT
GACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGG
GACTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGAT
GTCAAGACCAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGC
CGTACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGGACTACCaNG
GTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCGTCCAGgGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATT
TCANCGCTACNCCTGGGAATTCTACC
>4.5
TGGAGGGAAAGGCGAAGtGAACAAAGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCA
GCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACA
TAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGG
GTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTC
ACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTT
GAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAANGATTCGAACCCCCAACACCTAGTGTCCATCGTTTAC
GGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCagGCCAGAgAGCCGCCTTCGCCACTGGTGTTCTT
CCTGATATCTACGCATTCCACCGCTACACCagGAATTCCGCTCTNCTCTCCTGAACTCANATATACAGTTTCAAATGCAGCTCAC
>4.7
AAAGGGTGACAAACTTTCGTGGTGTGACGGGCGGTGTGTACTAGGCCCGGGAACGTATTCACCTCGGTATGCTGACCCGAGATTACTAGCGATTCCGGC
TTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGC
ACGTGTGTAGCCCGAGACATAAAGGGCATGATNGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGNCCCCGACTTTACT
CGCTGGTAACTAAGGACAGGGGTTGNGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACNACAACCATGCACCACCTGTCTCCGTGTG
NTCTAATGCAGACTCTGCTTTCACACATAGGAACGGGAACGTCAAGCCTCGTTAAGGNTCTGCGCGTTGCCATCG
>4.8
TAGTGACTTCAGGGGATGCCAACTCTCATGGAGGGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGC
AATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAAC
CATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCC
ACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGT
CTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACT
GCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGA
TACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGaCTACCaNGtATCTAATCCTGTTTGCTCCCCACGCTCTCGCGCCTCAGCGTCAGTTGCTGTCCA
GAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTCCTCTCCA
>4.11
GAAAGGAGGTGATCCAGAAAGGAGGTGATCCATAAAGGAGGTGATCCATAAAGGAGGTGATCCACAAAGGAGGTGATCCANAAAGGAGGTGATCCAGAA
AGGAGGTGATCCATAAAGGAGGTGATCCAGAAAGGAGGTGATCCCGAAAGGAGGGGATCGAGTTGGANACTCCAATCCGGAGTACGACGCCCTTTATG
AGGTCCGCTTGCTCTCACGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTTGACGTCATCCCC
ACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGGGACTTAACCCAAC
ATTTCACAACACGAGCTGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGATGTCAAGACCAGGT
AAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGG
CGGTCGACTTAACGCGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGGacTACCaGGTATCTAATCCTGTT
TGCTCCCCACGCTTTCGCACCTGANCGTCAGTCTTCGTCCagGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATTTCACC
>4.12
AAAGGAGGTGATCCATAAAGGAGGTGATCCATAAAGGAGGTGATCCACAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCATAAA
GGAGGTGATCCAGAAAGGAGGTGATCCCGAAAGGAGGTGATCNAGAAAGgAgatGATCCtGAAAAGTACtACGCACTTTATGAGGTCCGCTTGCTCTCACG
AGGTCCCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTTGACGTCATCCCCACCTTCCTCCAGTTTATCAC
TGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATTTCACAACACGAGCTGA
CGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGATGTCAAGACCAGGTAAGGTTCTTCGCGTTGCA
TCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCGTACTCCCCAGGCGGTCGACTTAACGCGTT
AGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACNGCGTGgACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCG
CACCTGAGCGTCAGTCTTCGTCCagGGGCCGCCTTCGCCACCGGTATTCCTCCAGATCTCTACGCATTTCACCGCTACAcCTGGAATTCTACCCCCCTCTA
CGAaACTCAAGCTTGCCAGTATCAaATGCAGTTCCCAGGTTGAGCCCGGGGATTTCACATCTGACTT
>4.13
GAAGTAAGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGG
GATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTT
GACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGG
ACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGAT
GTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGAC

CGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGG
GTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCagcCAGANAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCC
ACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTCA
>4.14
TACAAGACTCTCGtGGTGtGACGGGCGGTGtGTACAAGACCCGGGAACGTATTCCCGCGACATGCTGATTCGCGATTACTAGCGATTTCCGGCTTCTGCA
GGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTTTGGGTTTCGCTCCGCCTCGCGGCTTCGCCTCCCTTTGTACCACCCATTGTAGCACGTGTG
TAGCCCAGGCCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCAGTTTGTCACTGGCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAAC
TAACGACAGGGGTTGCGCTCGtTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAA
CTGTATTTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGtGCGGGCCCCCGTC
AATTCCTTTGAGTTCAGCCTTGCGGCCGTACTCCCCAgGtGGGATACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACNTcctACACCTAGTATtCAT
CGTTAACGCCAGGACTACcGGGNTATCTAATCCCGTCGCTCCCTGGNTTTCGCGCCTCAtCGTCAGG
>4.15
GTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGcCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATG
CAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGT
GTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGG
CAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACgAGCTGACGACAACCATGCACCACCTGTCTCCTTTGcCCCgAAG
GGAAACTGTATTTCCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCCTTCGCGTTTGCTTCNAATTAAACCACATACTCCACTGCTTGTGCGGGC
CCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGGCCGTACTCCCCAGGTGGGATACTTATTTGTGTTAACTACgGNACANAAAGGGGtCGATACCTCCTA
CACCTAGTATCCATTCGTTACcGCGTGGACTAaCCAGGGTATCTAATCCTGTTTGCTCCCCACGCTCTCGCGCCTCAaCGTCAgTGGCTGTC
>4.16
GGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCCCAAAGGAGGTGATCCAGAAAGG
AGGTGATCCAGAAAGGAGGTGATCCGGGAAGGACGGGAtCCCGATTGGAAGCGATTCCGACTTCATGGAGTCGAGTTGCAgACTCCAATCCGGACTACG
ACGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTT
GACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGG
GACTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGAT
GTCAAGACCAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGC
CGTACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACgGCGTGGACTACCAGG
GTATCTAATCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCGTCCAGGGGGCCGCCTTCNCCAcCGGTATTCCTCCAGATCTCTACGCATT
TCACCGCTACACCTGGAATTCTACCCCCCTCTACGANACTCAGGCTTGCCAGTATCAGATGCAGTTCCCAGGTTGAGCC
>4.17
AAAGGATGTGACTCGGGTGGAGGGAAGGGCGANGTGAACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCT
TCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCA
CGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGC
TGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCC
GAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCNAATTAAACCACATGCTCCACCGCTTGTGCG
GGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAAC
ACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCagGCCAGANAGCCG
CCTTCGCCACTGGTGTtCTTCCTGATATCTACGCATTCCACCGCTACACCA
>4.18
TGACTTCTGGTGCATCTAACTCGGGTGGAGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGAT
TCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCAT
TGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTT
TACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGT
GTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCT
TGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACC
CCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGANCGTCAGTTTCAGGCCAG
ANAGCCGCCTTCGCCACTGGTGTTCCTCCCGATATCTACGCATTTCACCGCTACACCGGGAATTCCGCTTCCCTCTCNTGCCCTCAGCCATGCAGTTTCG
ATCGGCATCCTAGAGTTNAGCCCTAGACTTTCACNACCGACTTGCACGACCGCCTACCCGCCCTTTACGCCCAGTAATTC
>4.19
GGAGGTGATCCAGAAAGGAGGTGATCCAGAAAGGAGGTGATCCATAAAGGAGGTGATCCATAAAGGAGGTGATCCAcAAAGGAGGTGATCCAgAAAGGA
GGTGATCCAGAAAGGAGGTGATCCgTAAAGGAGGGGATCCCGAATAGAGGCGATCCCGACTTCANGGGATCGAGTTGCgaACTCCAATCCGGACTACgA
CGCACTTTATGAGGTCCGCTTGCTCTCGCGAGGTCGCTTCTCTTTGTATGCGCCATTGTAGCACGTGTGTAGCCCTGGTCGTAAGGGCCATGATGACTTG
ACGTCATCCCCACCTTCCTCCAGTTTATCACTGGCAGTCTCCTTTGAGTTCCCGGCCGGACCGCTGGCAACAAAGGATAAGGGTTGCGCTCGTTGCGGGA
CTTAACCCAACATTTCACAACACGAGCTGACGACAGCCATGCAGCACCTGTCTCACGGTTCCCGAAGGCACATTCTCATCTCTGAAAACTTCCGTGGATGT
CAAGACCAGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCATTTGAGTTTTAACCTTGCGGCCG
TACTCCCCAGGCGGTCGACTTAACGCGTTAGCTCCGGAAGCCACGCCTCAAGGGCACAACCTCCAAGTCGACATCGTTTACGGCGTGgaCTACCagGTAT
CTAATCCTGTTTGCTCCCCACGCTTTCGCACCTGAGCGTCAGTCTTCGTCCagGGGCCGCCTTCGCCACCGGTATTCCTCCAaATCTCTACGCATTTCA
>4.20
GGAGGGAAGGGCGAAGtGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAG
CCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACAT
AAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGG
GTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTC
ACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTT
GAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTAC
GGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCNGGCCAGAgAGCCGCCTTCGCCACTGGTGTTCTT
CCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATATACAGTTTCAAATGCAGCTcAcCGGGTTAAGCCGGTGC
CTTTCACATTTGACTTATA
>4.21
ACGGATCCCTCCTTGCGGTTAGGCCACCGGCTTCGGGTGTTACAGACTCTCGTGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGC
GACATGCTGATTCGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTTTGGGTTTCGCTCCGCCTC
GCGGCTTCGCCTCCCTTTGTACCACCCATTGTAGCACGTGTGTAGCCCAGGCCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCAGTTTGT
CACTGGCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTG
ACGACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGGAACTCACTATCTCTAGCGATGGCGGAGGGATGTCAAGGCCTGGTAAGGTTCTTCGCGT
TGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCCCAGGCGGAGTACTTATTG
CGTTAACTTCAGCACCGATCTTGCGACCGACACTTAGTACTCATCGTTTACGGCGTGGaCTACCANGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTG
CCTCAGCGTCAGGGGTAAGCCAGAANGCCGCCTTCGCCACTGGNGTTCCTCCACATCTCTACGCATTTCACCGCTACACGTGG
>4.22

CACCGATCTTCAGGTGATACAGACTCTCGTGGTGGGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCGACATGCTGATTCGCGATTACTAG
CGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTTTGGGTTTCGCTCCGCCTCGCGGCTTCGCCTCCCTTTGTACCA
CCCATTGTAGCACGTGTGTAGCCCAGGCCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCAGTTTGTCACTGGCAGTCTCGTTAGAGTCC
CCAACTTAATGCCGGCAACTAACGACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTC
TCCTCCGTCCCCAAAAGGAAACTCACTATCTCTAGCGATGGCGGAGGAATGTCAAGGCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTC
CACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCCCAGGCGGAGTACTTATTGCGTTAACTACGGCACGGATGGTA
TCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCTAGCTTTCGTGCCTCAGCGTCAGTTATC
GTCCAGGTAGTCGCCTTCGCCACTGGTGTTCCTCCCAATATCTACGCATTTCACCGCTACACTGGGAATTCCACTACCCTCTCCGAcACTCAAGCCCTCCA
GTTTCGAA
>4.23
ACTCTCATGGTGGGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCG
AGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCC
CAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTA
ACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAAC
TGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAA
TTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACAGAANGGGTCGATACCTCCTACACCTAGTATCCA
TCGTTTACgGCGTGGaCTACCaGGGTATCTAATCCTGTTTGCTCCCCACGCTCTCGCGCCTCANCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTG
GTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTCCTCTCCAGCACTCAAGAGA
>4.24
TGGCTTCGGGTATTGCCAACTCCCATGGGGGGACGGGCGGTGTGTACAAGACCCGGGAACGCATTCACCGCGGCATGCTGATCCGCGATTACTAGCAA
CTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCGGCTTTTAGGGATTAGCTCCCCATCGCTGGGTGGCAACCCGTTGTACCGAC
CATTGTAGCACGTGTGTAGCCCAGGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGATTTGTCATCGGCAGTCCCTCTAGAGTGCCCA
CCCGAAGTGCTGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCT
CCTCCGTCCCCGAAGGGAAAGACCTATCTCTAGGCCCGGCAGAGGGATGTCAAGCCCTGGTAAGGTTCTTCgCgtTGCTTCgAATTAAACCACATGCTCCG
CTGCTTGTGCGGGTCCCCGTCAAtTCCTTTGAGTTTCATGCNTGCGCACgtTACTCCCCCAgGCGGAGTGCNTTNNNGCGTTAGCTGCGGCACTGAAGGG
TAATCCCCTCC
>4.25
ACAACTGACTTCAGGTGATGCCAACTCTCATGGAGGGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTA
GCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTAGCGGCTTCGCTGCCCTCTGTAGCA
ACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGC
CCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCT
GTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCA
CTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTAATGCGTTAACTTCGGCACTGAAGGATTC
GAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAGCGTCAGTTTCagcC
AGAGAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCACTCTCCTCTCCTGAACTCAAGATATACAGTTTC
AAATGCAGCTccACCGGGTTAAGCCGGTGCCTTTCACATTTGACTTATATATCCGCCTACGTACCC
>4.26
TGCGGTTAGGCCACCGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCT
GCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCC
CTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTC
GCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCA
TGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACC
ACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACA
GAANGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGaCTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTCTCGCGCCTCAgCGTC
AGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTCCTCTCCAGCACTCAAGA
>4.30
CATATCGCAAATGACTTCTGAGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCG
ATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTC
TGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAG
AGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACC
ACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACA
TGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTTAACTTCGGCACTGA
AGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCGTACCTGAgCGTCAGT
TTCAGGCCAGAgAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAATTCCGCTCTCCTCTCCTGAACTCAAGATAT
ACaGTTTCAAATGCAGCTcA
>4.31
TGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCGACATGCTGATTCGCGATTACTAGCGATTCCGGcTTCATGCAGGCGAGTTGCAGCC
TGCAATCCGAACTGAGAGTGGCTTTTTGGGTTTCGCTCCGCCTCGCGGCTTCGCCTCCCTTTGTACCACCCATTGtagCACGTGTGTAGCCCAAGCCATAA
GGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCAGTTTGTCACTGNCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACGAGGGTT
GCGCTCGtTGCGGGACTTAACCCAACATCTCACGACACGaGCTGACGACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGAaACTCACTATCTCTA
gcgATGGCGGAGGGATGTCAAGGCCtGgTAAGGTTCTTCGCGTTGCTTCgAATTAAACCACATGCTCCACCGCTTGtGCGGtCCCcgTCNATTCCTTTGAgTT
TCAGTCTTGCaACCGTACTTCCCAGGCGGGAGTACTTATTGGCGTTAACTTCAAGACCCGATCTTGCGACCGACACTTATtACTCATCGTTTACGGCGGGG
ACTACCAGGGTATNCTAATCCGGTT
>4.32
CCCCCAATCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCAGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTG
CTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGTGCATGATGATTTGACGTC
GTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTT
AACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAG
CCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCaATTCCTTTGAGTTTCAgGCCTTGCGGCCGTAC
TCCCCAGGTGGGATACTTATTGTGTTAACTACGGCACANAAGGGGTCGATACCTCCTACACCTAGTATCCATC
>4.33
GATGACTTCTGAGTGCATCTAACTCGGGTGGAGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGC
GATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGA
CCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCC
GACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTC
TCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCA
CCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCCCAGGCGGAGTACTTATTGCGTTAACTTCAGCACCGATCTTGCGA

CCGACACTTAGTACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGTGCCTCAGCGTCAGGGGTAAGCCAGAA
NGCCGCCTTCGCCACTGGTGTTCCTCCACATCTCTACGCATTTCACCGCTACACGTGGAATTCCACCTTCCTCTCTTACCCTCAAGTCCGCCAGTTTCTAAT
CCCC
>4.35
GACTTTCaCCCcAaTCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGT
GTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTG
GGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATT
TGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGC
GGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGNCCCCGAAGGGAAACTGTATTTTCTACAGCGGTCAAGG
GATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCgAAATTAAACCACATACTCCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTG
CGGCCGTACTCCCCAGGTGGGATACTTATTGTGTTNAACTACGGGCACAGAAgGGGTCGATAcCTCCTACACCTA
>4.36
TGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCA
GCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACA
TAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGG
GTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGNCCGAAGGCAGACTCTGCTTTC
ACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGNTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTT
TGAGTtTCAGTCTTGCGACCGTACTCCCCAgGCGGGGACACTTAATGCGTTAACTTCGGCACTGNAAGGATTCGAACCC
>4.37
GGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTT
CATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTAC
GTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTG
CTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCC
GAAGGGAAACTGtATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGtTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGG
CCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGNGGATACTTATTGTGTTAACTACGGCACAgAAGGGGTCGATACCTCCTACACC
TAGTATCCATCGTTTACNGCGTGGACTACCAGGGTATCTAATCCTGNTTGCTCCCACGCTCTCGCGCCTCANCGTCAGTTGCTGTCCAaAAGTCGCcTTCN
CCACTGGTGTTCCTCC
>4.38
AGCCCCAGTCACCNGCCCCACCTTCGGCAGCTCCCTCCCCTACGGGGTTGGGTCACTGACTTCGGGTGTTGCCAGCTCCCATGGCTTGACGGGCGGTG
TGTACAAGGCCCGGGAACGTATTCACCGCGACGTAGCTGATTCGCGATTACTAGCAATTCCGTGTTCACGCAGGCGAGTTGCAGCCTGCGATCCCAACT
GGGGCCGGCTTTAAGGGATTCGCTCCGGATCGCTCCTTCGCAGCCCGTTGTACCGGCCATTGTAGCGCGTGTGTAGCCCAGGGCATAAGGGCCATGAT
GATTTGACGTCATCCCCACCTTCCTCCGGATCGTCTCCGGCAGTCTCCAGTGAGTCCCCGGCATTACCCGCTGGCAACACTGGATAGGGGTTGCGCTCG
TTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCAACACCTGTCATGGCGTCCGTATTGCTACGGAAGANCGTGTTTCCACGGC
TGGCACCATGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCGTCGAATTAAACCACACGCTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTT
TAGCCTTGCGGCCGTaCTCCCCAgGCGGAGTGCTTANTTGTGTTAACTCCGGCACGGAGGAtTGtGCCCCCCACACCTAGCACTCATCGTTTACGGNTGG
GACTACCANGGTATCTAATCCTGTTTGCTCCCCCAGCTTTCGCGCCTCAGCGTCAGGTACAGTCCAGGAAGCCGCCTTC
>4.40
CCCCTATCaTCCGCCCCACCTTCGACGGgTCCCTCCCTTACGGGTTAGGGCACCNGCTTCGGGTGTTGCGAACTTTCTGGGTGTGACGGGCGGTGTGTA
CAAGGCCCGGGAACGTATTCACCGCGGGATGCTGACCCGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAG
TGGCTTTCTGGGATTCGCTCCGCCTCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAtCACGTGTGTATCCCACATCATAAGGGGCATGATGATTTGAC
GTCATCCCCACCTTCCTCCGACTTGTCGCCGGCGGTCCCTTTANAGTCCCCAACTTAATGCTGGCAACTAAAGACAAGGGTTGCGCTCGTTGCGGGACTT
AACCCAACATCTCACNACaCGAGCTGACgACAACCATGGACCACCTGTCATCCAgTCcCNaaGGAGGGCcATGttTCCATGGCTTGCTAGGTGATGTCAAG
ACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTACT
CCCCAGGCGGGGTACTTAATGCGTTAACTTTCGGCACTGANGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACNGcTGGGACTACGGGGTATCTAA
TCCCgTTGCTACCCAGCTTTCGTACCTGAGCGTCAGTTTCAGGCCAGAaAGCCGCCTT
>7.1
CCGGGAaCGTATTCaCCGCAgTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCaTGCaGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTT
CTGAGATTGGCTCCATGTCACCaCtTCGCTTCCCTCTGTACTGACCaTTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCC
CCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTagAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCA
ACATCTCACGACACGANCTGACGACaACCATGCACCACCTGTCTcCGTGTGCCCgAAGGCAgACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTTC
GGTAAGGTTCTGCGCCGTTGCATCgaATTAAACCAc
>7.2
aCTTCaCCCCAATCATCTGTCCCACCTTCGACGGCTCCCTCCTTGCGGTTAGGCCACCGGCTTCGGGTGTTACAGACTCTCGTGGTGTGACGGGCGGTG
TGTACAAGACCCGGGAACGTATTCACCGCGACATGCTGATTCGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGA
GAGTGGCTTTTTGGGTTTCGCTCCGCCTCGCGGCTTCGCCTCCCTTTGTACCACCCATTGTAGCACGTGTGTAGCCCAGGCCATAAGGGGCATGATGATT
TGACGTCATCCCCACCTTCCTCCAGTTTGTCACTGGCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACGAGGGTTGCGCTCGTTGCGGG
ACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGAAACTCACTATCTCTAGCGATGGCGGAGG
AATGTCAAGGCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGC
GACCGTACTTCCCAGGCGGAGTtACTTATTGCGTTAACTTTCAGCACCGATCTTGCGACCGACACTTAGTACTCATCGTTTACGGCGTGGACTACCAGGGT
ATCTAATCCTGTTTTGCTCCCCACGCTTTCGTGCCTCagCGTC
>7.3
CATACGCAAATGACTTCTGGTGCATCTAACTCGGgTGGTGTGACAGGCGGTGGGTACAAGGCCCGGGAACGTATTCACCGCagtAGCTGACCTGGCATTA
CGAGCGATTCCGGCTTCATGCAGGCCAGTTGCAGCCTGCGATCCGAACTGGGATCAGCTTTCTGAGATTGGGTCCTTGTCACCACTTCTCTTCCCTCTGT
ACTGACCATTGTAGCACGTGTGTATCCCGAGACATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTCTCCTTACAGT
CCCCGACTTTACTCaCTGGTAACTAANGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACTACACCaAGCTGACtACNACCATGcaCCACCT
GTcTCCGTGTGCCCNAAGGCacAcTCTGCTttAACAgATAGGCACgGGATGTCAAGTCTCGGTGAAGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCC
ACCGCTTGTGCGGGCCCCCGtCaaGTNCTTTGAcTTTCAGTCTTGCNACAGTACTCCCCA

>7.4
CCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCAC
CGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCAT
GTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCT
TTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGA
GCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGgtaagGTTCTTCGCGTTG
CTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTNcCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTTATTG
TGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGNACTACCAGGGTATCTAATCCTGtTTGCTCCCCACGCT
CTCGCGCCTCaaCGNTCAGTTGCTGTCCAGAAAAGTCGCCTTC
>7.6

CCCCAATCATCAACCCCACCTTCGGCAGACGTCCCACCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTAC
AAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTG
CTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGT
CGTCCCCACCTTCCTCCGATTTGTCACCGGCGGTCTCGCTAGAGTGCCCACCTCTACCTGCTGGCAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACT
TAACCCAACATCTCACGACACGAGCTGACAACAACCATGCACCACCTGTCTCCTTGNCCCGAAGGGAAACTGTATTTCTACAggAgTCAAGGGATGTCAAG
CCCTGGTAAGGgTCTTCGCGTTGCTTCcAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGNCGTACT
CCCCANGTGGGATACTTATTGTGTTAACTACGGcACAgAAAGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTGGACTACC
>7.7
GTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCA
TGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGT
GTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCT
GGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTaACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGA
AGGGAAACTGTATtTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAAGGTTCTTCGCGTTGCTTTCGAATTAAACCACATACTCCACTGCTTNGTGCGG
GCCCCCCGTCAATTNCCTTTGAGTTTCAGCCTTTGCGGCCGTACTNCCCCAGGTGGGNATACTTATTGTGTTAACTACNGcACAGAAGGGGTCGATACCtT
CCTACACCTAGTATCCATCGTT
>7.8
ACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCT
GACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTC
GCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCTG
TCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAA
CCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAgACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAAT
TAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGGACACTTAATGCGTTAACTT
CGGCACTGAAAGGATTCGAACCCCCAACACCtTAtTGTCCATCGTTTACGGCTGGGActTAaCCGGGGGTATCTAATCCCGTTTGCTACCCCAGCTTTCNTA
CCTGA
>7.9
GACTGCACCCCGTTACGGTTAGCTCACTGAGCTTCGGGCATTGCTGACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGC
GGCATGCTGATCCGCGATTACTAGCAACTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGATCGGCTTTCGGGATTCGCTCCCCATC
GCTGGGTGGCTGCCCGCTGTACCGACCATTGTAGCACGTGTGTAGCCCAGGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGACTTT
TTGCCGGCAGTCCCTCTAGAGTGCTCAGCTTTACCTGTTAGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAG
CTGACGACAACCATGCACCACCTGTCTCCTCTGTCCCCGAAGGGAAAGCCCTATCTCTAGAGCGGTCAGAGGGATGTCAAGCCCTGGTAAGGTTCTTCGC
GTTGCTTCGAATTAAACCACATGCTCCGCTGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCATTCTTGCGAACGTACTCCCCAGGCGGAGTGCTTAA
TGTGTTAACTCCGGCACCGAAGATAAATCCCCGACACCTAGCACTCATCGTTTACGGCGTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCACGCTCT
CGCGCCTCAGCGTCAGTTGCTGTCCAGAAAGTCGCCTTCGCCACTGGTGTTCCTCCTAATCTCTACGCATTTCACCGCTACACTAGGAATTCCACTTTCCT
CTCCAGCACTCAAGA
>7.11
ACACACCTTCGGCATTCGCAAATGACTTCTGGTGCaTCTaaCTTtgGTGGtGTGACGGGCGGTGTGTTCAAGGCCCGGGAACGTATTCaCCGCAGTATGCT
GACCTGCGATTACTAGCGATTCCGGCTTCATGCAgGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGtCACCACTTCg
CTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCcTcCGgCTTtCACCGGcTGTC
TCCtTAGAGTCCCCGACTTTACTCGCTGGTAACTAAgGACAGGgGTTGCGct
>7.12
CCCAtCATCAGCCCCACCTTCGTGCAGCGTCcCCCTTGCGGTTAGACTACTCANTTCgGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAT
GCCCGGGAACGTATTCACGGCAgTaTGCTGACcTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTT
TTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTANTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGT
ccCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAgAGTGcCCACCTCTACGTGCTGGCaACTAACGACaGGgGTTGCGCTCGTTGCGGGACTTAACC
CAACATCTCACGACTcGAGCTGACNACAACCATGCACCACCTGTCTccTTGCCCCGAAGGGaAACTGtATTTCTACAgCGGCCAAGGGATGTCAAGCCCTG
gTtAGGTtctTCGCGTTGCTNCNAatTAACCACATACTCCACTGCTTGTGCGGGCCCCCTTCAATTCCTTTGAGTTTTCA
>7.13
CCCCAATCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACcGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACA
AGGCCCGGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTG
CTTTTAGGGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGT
CGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACT
TAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCcCCgAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAG
CCCTGGTAAGGtTCTTCGCGTTGCTTCgAATTAAACCACATACTCCCACTGCTTGTGCGGGGCCCCCGNCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTA
CTCCCC
>7.14
CATCCAGCTCCACCTTCGACGGCTCCCTCCTTACGGTTAGGCCACCGGCTTCGGGTGTAACTGACTTTCGTGGTGTGACGGGCGGTGTGTACAAGGCCC
GGGAACGTATTCACCGCGGTATGCTGACCCACGATTACTAGCGATTCCGACTTCACGGAGGCGAGTTGCAGCCTCCGATCCGAACTGGGACGCACTTTA
AGGGATTCGCTTCAGGTCACCCTGTCGCTGCCCGTTGTATGCGCCATTGTAGCACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCATCC
CCACCTTCCTCCAGTTTATCACTGGCAGTCCCTTTAGAGTGCCCACCTTGCGTGATGGCAACTAAAGGTAAGGGTTGCGCTCGTTGCGGGACTTAACCCA
ACATCTCACGACACGAgcTGACGACAACCATGCACCACCTGTCAACCAGCTTCCCGAAGGAATCATCCATGTCTCCATGAATTACATCCGGTTGATGTCAA
GCCCTGGtAAGGtTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACT
CCCCAgGCGGGACACTTAATGCGtTAACTTCgGCACTGAAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGNtGGGACTACCGGGGTATCTA
AATCCCGTTTGCTACCCCAGCTTT
>7.15
TCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCC
GGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAG
GGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCC
ACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGacAGGGGTTGCGCTCGTTGCGGGACTTAACCCAA
CATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGT
AAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGG
tGGGATACTTATTGTGTTAACTACGGCACAGAANGGGtTCGATACCTCCTACACCTAGTATCCATCGTTTACGGCGTgGGACTACCAGGGTATCTAATCCTG
TTNGCTCCCCACGCTCTCGCGCCTCANCGTC
>7.16
aCAAGGCCCGGGAACGTATTcaCCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCNTGCANGCGAGTTGCANCCtGCAATCCGAACTGGGAT
CaGCTTTCTGAGAtTGGCTCCATGTCACCACTTCgCTTCCCTCTGTACTGACCaTTGTAGCACGTGTGTAGCCCGAGACATAaAGGGCATGATGATTTGACG
TCATCCCCACCTTCCTCCGGCTTTCACCGGCTGTcTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTT
AACCCAACATCTCACGACACNAGCTGACgACaACCATGcaCCaCCTGTCTCCGtGTGCCCGAAGGcagAcTCTGcTTtTcaCAGATAGGCTCgGGATGtTAGT
CTCGGtAAgGTTCTGC

>7.19
GCAAATGACTTCTGAGTGCATCTAACTCGGGTGGAGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACT
AGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGCTTCCCTCTGTAC
TGACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGT
GCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCAC
CTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATACTC
CACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGTGGGATACTTAATGCGTTTGCTTCGGCACTGAAGGAT
TCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTCGCTCCCCCAGCTTTCGTACCTGAgCGTCAGTTTCAG
GCCAGAgAGCCGCCTTCGCCACTGGTGTTCTTCCTGATATCTACGCATTCCACCGCTACACCAGGAAtTCCGCTCTCCTCTCCTGAACTCAAGATATACAGT
TTCAAATGCCGCTCACCGGGTTGAGCCGATGCCCTTTCACATTTGACTTATATATCCGCCTGCGTACC
>7.20
GGTATTGCTAACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCGGCATGCTGATCCGCGATTACTAGCAACTCCGGCTT
CATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGATCGGCTTTTAGGGATTGGCTCACGATCGCTCGCTTGCTACCCGTTGTACCGACCATTGTAGCA
CGTGTGTAGCCCAGGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGATTTGTCATCGGCAGTCCCTCTAGAGTGCCCACCTTAACGT
GCTGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAACCATGcACCACCTGTCACCTCTGTCC
CCGAAGGGAAGGCTCTATCTCTAgAGCGGGCAGAGGGATGTCAAGCCCTGGTAAGGTTCTTCGCGTTGCATCGAATTAAACCACATGCTCCGCTGCTTGT
GCGGGTCCCCGTCATTCCTTTTGAGTTTCATGCTTGCGCACGgTACTCCCCA
>7.21
CCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGacTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGT
ATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTGCT
CCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCCTC
CGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCAC
GACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCcGAAGGGAAACTGTATTtTCTACAGcGGTCAAGGGATGTCAAGCCCTGGTAAGGTTCT
TCGCGTTGCTTCGAATTTAAACCACATACTCCACTGCTTTGTGCGGGCCCCCGTCAATTCCTTTTGAGTTTTCAGCCTTGCGGGCGTACTCCCCAGGTGGG
GATACTTATTGTGTTAACTACGGCACAGAAgGGGGCCGATACCTCCTACACCTAGTAttCATCGTTACGGCGNGGAACTACCAgGGtATCTAATCCTG
>7.22
CCCCTTACCaGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCNAAcTCaGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCA
CCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCA
TGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGC
TTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACG
AGCTGACGACAACCATGCACCAcCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCG
CGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTT
AATGCGTTAACTTCgGCACTGAAGGATTCGAACCCCCNNACACCTAGTGTCCATCGTTTACGGCTGGGaCTACCGGGGtATCTAATCCCGGTTTGCTACCC
AGCTTTCGTACCTGA
>7.23
CCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATT
CACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTC
CATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCG
GCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGAAACTCACTATCTCTAGCGATGGCGGAGGGATGTCAAGGCCTGGTAAGGTT
CTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCCCAGGCGGAGT
ACTTATTGCGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGGNGTGGACTACCAGGGTATCTAATCCTGTTTGCT
CCCCACGCTCTCGCGCCTCAGCCGTCAGTTGCTGTCCAGAAAGTCGCCT
>7.27
ACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTcacTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCA
GTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAgCTTTCTGAGATTGGCTCCATGTCA
CCACTTCGCTTCcctCTGTACTGACCATTGTAgCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCAC
CGGTTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGtAACTAAGGacAGGGGTTGCGcTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGAC
GACAACCATGCACCACCTGTCTNCgTGTGCCCgAagGCaaACTCTGCTTTCACAgATAGGCAC
>7.28
CCCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGaCGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGAC
aCGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCT
GCGCGTTGCATCGAATTAAACCACATGCTCCCACCGCTTGTGCGGGCCCCCCGTCAATTCCTTTGAGTTTTCAGTCTTGCGACCGTACTCCCCCAggGCG
GGACACTTAATgcgTAACTTCGGcaCTGAAGGATTTC
>7.29
CCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAAC
GTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAGGGATTTG
CTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCCACCTTCC
TCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGCAACTAACGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCA
CGACACGAGCTGACGACAACCATGCACCACCTGTCTCCTTGCCCCGAAGGGAAACTGTATTTCTACAGCGGTCAAGGGATGTCAAGCCCTGGTAAGGTTC
TTCGCGTTGCTTCGAATTAAACCACATACTCCACTGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTCAGCCTTGCGGCCGTACTCCCCCAGGTGGGA
TACTTATTGTGTTAACTACGGCACAGAAGGGGTCGATACCTCCTACACCTAGTATCCATCGTTTACGGNGNGGACTACCAGGGTATCTAATCCTGTTTGCT
CCCCACGCTCTCGCGCCTCANCGTCAGTTGCTGTCCAGAAAGTCGCCTT
>7.30
CCCCCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATT
CACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTC
CATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCG
GCTTTCACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTG
CGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACA
CTTAATGCGTTAACTTCGGCACTGAAgGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGA
>7.31
CCTTACCAGACACACCTTCGGCATTCGCAAATGACTTCTGGTGCAtcTAACTCggGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACC
GcAGTATGCTGACCTGCGatTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGT

CACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTT
CACCGGCTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGC
TGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCttTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTT
GCATCGAATTAAACCACATGCTCCCACCGCTTGTGCGGGCCCCCGTCAATTCCCTTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTA
ATGCGTTAACTTCcGCACTGAAGGATTCGAANCCCCCAACACCCTAGTGTCCNTCGTTTACGGNTGGGAACTACCGGGGTATCTAATCCCGTTTGCTACC
CCAGCTTTCGTACCTGAGCG
>7.32
GACACACCTTCGGCATTCGCAAATGACTTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTAT
GCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCAC
TTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGG
CTGTCTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAAcATCTCACGACACGAGCTGACGA
CaACCaTGCACCACCCTGTCTCCGTGTGCCCGAAGGcAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTCGGTAAGGTTCTGCGCGTTGCATC
GAATTAAACCACATGCTCCACCGCTTGTGCGGGGCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGCGGGACACTTAATGCGTT
AACTTCNGCACTGAAGGATTCGAACCCCCAACACCTAGTGTCCATCGTTTACGGCTGGGaCTACCGGGGGtATCTAATCCCGTTtGCTACCCCAGCTTTCG
TACCTGAGCGTC
>7.33
CCcCCcTTACCAGACACACCTTCGGCATTCGCAAAATGACTTCTGGTGCATCtAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTAT
TCACCGCAgTATGCTGACCTGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCT
CCATGTCACCACTTCGCTTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCC
GGCTTTCACCGGCTGTCTCCTTAGAgtCCCCGACTTTACTCGCTGGTAACTAANGACAGGGGTTGCGCTCGTTGCgGGACTTAACCCAACATCTCACGACC
CGAGCTGacgaCAACCATGCACCCACCTGTCTCCGTGTGCCCNAAcGCAAACTCTGCTTTCACANATAGGCACGGGATGTCAAGTCTCG
>7.34
CTTCGGCATTTCGCAAATGACTNTCTGGTGCATCTAACTCGGGTGGTGTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGA
CCTGCGATTACTAGCGATTCCGGNNTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGGGATCAGCTTTCTGAGATTGGCTCCATGTCACCACTTCGC
TTCCCTCTGTACTGACCATTGTAGCACGTGTGTAGCCCGAGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCGGCTTTCACCGGCNGT
CTCCTTAGAGTCCCCGACTTTACTCGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCNNGACACGAGCTGACGACAA
CCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCNAGTCTCGGTAAGGTTCTGCGCGTTGCATCGAA
TTAAACCACATGCTCCACCGNNTTGTGCGGGGCCCCCCGTCAATTCCTTTGAGTTTC
>7.36
CGGATCCCCCCTTGCGGTTAGGCCACCGGCTTCGGGTGTTACAGACTCTCGTGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCG
ACATGCTGATTCGCGATTACTAGCGATTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAGTGGCTTTTTGGGTTTCGCTCCGCCTCG
CGGCTTCGCCTCCCTTTGTACCACCCATTGTAGCACGTGTGTAGCCCAGGCCATAAGGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCAGTTTGTC
ACTGGCAGTCTCGTTAGAGTCCCCAACTTAATGCTGGCAACTAACGACGAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACACGAGCTGA
CGACAACCATGCACCACCTGTCTCCTCCGTCCCCAAAAGGAAACTCACTATCTCTAGCGATGGCGGAGGGATGTCAAGGCCTGGTAAGGTTCTTCGCGTT
GCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTTCCCAGGCGGAGTACTTATTGC
GTTAACTTCAGCACCGATCTTGCGACCGACACTTAGTACTCATCGTTTACGGCGTGGaCTACCAGGGTATCTAATCCTGTTTGC
>7.38
ATCCACAGACTGCCCCCTTACGGTTAGCTCACTGAGCTTCGGGCATTGCTGACTCCCATGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTC
ACCGCGGCATGCTGATCCGCGATTACTAGCAACTCCGGCTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGATCGGCTTTCGGGATTCGCTCC
CCATCGCTGGGTGGCTGCCCGCTGTACCGACCATTGTAGCACGTGTGTAGCCCAGGACATAAAGGGCATGATGATTTGACGTCATCCCCACCTTCCTCCG
ACTTTTTGCCGGCAGTCCCTCTAGAGTGCTCAGCTTTACCTGTTAGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCACGACA
CGAGCTGACGACAACCATGCACCACCTGTCTCCTCTGTCCCCGAAGGGAAAGCCCTATCTCTAGAGCGGTCAGAGGGATGTCAAGCCCTGGTAAGGTTC
TTCGCGTTGCTTCGAATTAAACCACATGCTCCGCTGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGTTTCATTCTTGCGAACGTACTCCCCAGGCGGAGTG
CTTAATGTGTTAACTCCGGCACCGAAGATAAATCCCCGACACCTAGCACTCATCGTTTACGGCGTGGaCTACCaGGTATCTAATCCTGTTTGCTCCCCACG
CTTTCGTACCTCAgCGTCAGTCCAAGGCCAGAAGGCCGCCTTCNCCACTGGtGTTCCTCCTGATATCTACGCATTTCACCGCTACACCAGGAATTCCGCCT
TCCT
>7.39
TCATCAGCCCCACCTTCGGCAGCGTCCCCCTTGCGGTTAGACTACTGACTTCGGGTGTTGCCGACTCTCATGGTGTGACGGGCGGTGTGTACAAGGCCC
GGGAACGTATTCACGGCAGTATGCTGACCTGCCATTACTAGCAATTCCGACTTCATGCAGGCGAGTTTCAGCCTGCAATCCGAACTGGGGCTGCTTTTAG
GGATTTGCTCCACCTCGCGGCTTCGCTGCCCTCTGTAGCAACCATTGTAGTACGTGTGTAGCCCAGGACATAAGGGGCATGATGATTTGACGTCGTCCCC
ACCTTCCTCCGATTTGTCACCGGCAGTCTCGCTAGAGTGCCCACCTCTACGTGCTGGTAACTAAGGACAGGGGTTGCGCTCGTTGCGGGACTTAACCCAA
CATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCTCCGTGTGCCCGAAGGCAGACTCTGCTTTCACAGATAGGCACGGGATGTCAAGTCTC
GGTAAGGTTCTGCGCGTTGCATCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGAcCGTACTCCC
CAGGCGGGACACTTATTGCGTTAACTTCNGNACTGAAGGATTCGAACCCCCAACACCTAGTGtcCATCGTTTACGGCTGGGACTACCGGGGTATCTAATC
CCGTTtGCTACCCCAGCTTTCGTA
>7.40
CCCCAATCATCCGCCCCACCTTCGACGGCTCCCTCCCTTACGGGTTAGGCCACCGGCTTCGGGTGTTGCGAACTTTCGTGGTGTGACGGGCGGTGTGTA
CAAGGCCCGGGAACGTATTCACCGCGGTATGCTGACCCGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCAATCCGAACTGAGAG
TGGCTTTCTGGGATTCGCTCCGCCTCACAGCTTCGCTTCCCTCTGTACCACCCATTGTAGCACGTGTGTAGCCCAGATCATAAGGGGCATGATGATTTGAC
GTCATCCCCACCTTCCTCCGACTTGTCGCCGGCAGTCCCTTTAGAGTCCCCAACTTAATGCTGGCAACTAAAGGCAAGGGTTGCGCTCGTTGCGGGACTT
AACCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCACCCAGTCCCGAAGGAGGGCCATGTTTCCATGGCTTGCTAGGTGATGTCAA
GACCTGGTAAGGTTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGCCCCCGTCAATTCCTTTGAGTTTCATGCTTGCGCACGTAC
TCCCCAGGCGGGGTACTTAATGCGTTAACTACGGCACGGATGGTATCGAAACCACCCACACCTAGTACCCATCGTTTACGGCTAGGACTACCAGGGTATC
TAATCCTGTTTGCTCCCCTAGCTTTCGTGCCTCAGCGTCAGTTATCGTCCAGGTAGTCGCCTTCGCcACT

vii. LEBENSLAUF

Berufserfahrung und Praktika

09/2012 – 09/2015	**Lehrstuhl für Mikrobiologie,** TU München Wissenschaftliche Mitarbeiterin Methodenentwicklung, Isolierung und umfassende Charakterisierung hydrolytischer Bakterien aus Umweltproben, Klonierung, Proteinexpression und Aufreinigung, Genom- und Metagenomanalysen
05/2012 – 09/2012	**SchueTo Umwelttechnik GmbH,** Frohnleiten (Österreich) Projektmanagement Wasserkraft und Gewässerökologie, Erstellung von Fachgutachten und Stellungnahmen
09/2011 – 04/2012	**Lehrstuhl für aquat. Systembiologie,** TU München Wissenschaftliche Hilfskraft Bearbeitung verschiedener gewässerökologischer Fragestellungen
01/2010 – 02/2011	**Lehrstuhl für Mikrobiologie,** TU München **EXIST Stipendium**. Entwicklung eines rekombinanten Enzymkomplexes zum Abbau von cellulosehaltiger Biomasse, Klonierung, Proteinexpression und –Aufreinigung in *Bacillus subtilis* und *E. coli*
11/2007 – 01/2008	**Lehrstuhl für Mikrobiologie,** TU München, Laborpraktikum Charakterisierung verschiedener *Clostridium thermocellum* Stämme und ausgewählter Umweltisolate, sowie qualitative und quantitative Untersuchung der cellulolytischen Aktivität auf aufbereiteter Zuckerhirse
07/2008 – 09/2008	**Biomechanical and Environmental Engineering Department,** Jomo Kenyatta University (Kenia), Auslandspraktikum **DAAD Stipendium**. Untersuchung der antibiotischen Wirkung von *Mondia whytei* Extrakt, Vergleich verschiedener Extraktionsmethoden, Hemmhoftests
07/2004 – 09/2004	**Kläranlage Haibach,** Passau, Laborpraktikum. Abwasseranalytik, Kläranlagentechnik, Mikroskopie

Ausbildung

10/2009 – 03/2012	**Technische Universität München,** Studium in Umweltplanung und Ingenieurökologie **Schwerpunkte:** Umweltmikrobiologie, aquatische Ökologie, Abwasserreinigung und –Entsorgung Praktika in aquatischer Ökotoxikologie, Bioindikation der Gewässergüte und Abwasserreinigung **Beurteilung:** sehr gut bestanden **Masterarbeit** am Lehrstuhl für aquatische Systembiologie der TUM: „Invasive Grundeln in der oberen Donau“
10/2007 – 10/2009	**Technische Universität München,** Studium der Molekularen Biotechnologie **Schwerpunkte**: Mikrobiologie pathogener Organismen, molekulare Medizin, Biomedizinische Forschung Praktika in Mykopathologie, Histologie, mikrobielle Biotechnologie Master of Science in Molekularer Biotechnologie **Beurteilung:** mit Auszeichnung bestanden **Masterarbeit** am Lehrstuhl für Mikrobiologie der TUM: „Untersuchung cellulolytischer Umweltisolate für den Einsatz zur Optimierung der Hydrolysestufe in Biogasfermentern“ Identifizierung und Charakterisierung der cellulolytischen Vertreter mittels DNA-Sequenzanalyse und MALDI-TOF-TOF
10/2004 – 08/2007 08/2007	**Technische Universität München,** Studium der Biochemie Bachelor of Science in Biochemie **Beurteilung**: sehr gut bestanden **Bachelorarbeit** am Lehrstuhl für Mikrobiologie der TUM: “Mikrobielle Hydrolyse schwer abbaubarer cellulosehaltiger Substrate”
09/1995 – 06/2004	**Auersperg-Gymnasium**, Passau

Weitere Qualifikationen

Sprachen	**Deutsch** (Muttersprache) **Englisch** (fließend)
Computerkenntnisse	sehr gut: Windows, MS Office, Clone Manager, GenDB gut: R, Primer
Sonstiges	Ausbildereignung Qualitätsmanagementsysteme (DIN EN ISO 9001:2008) Versuchstierkunde Führerschein Klasse B

Soziales Engagement

1995 – heute	Aktive Mitgliedschaft im Kirchenchor
1997 – heute	Mitarbeit in der Jugendgruppe des Bayerischen Waldvereins