# Genom- und Transkriptionsanalyse von Bacillus licheniformis DSM13 - einem Organismus mit großem industriellem Potential

#### Dissertation

zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultäten der Georg-August-Universität zu Göttingen

vorgelegt von
Birgit Veith
aus Jena

Göttingen 2004

D7

Referent: Prof. Dr. G. Gottschalk

Korreferentin: Prof. Dr. R. Schmitz-Streit

Tag der mündlichen Prüfung: 25.01.2005

Abkürzungsverzeichnis	VI
1. Einleitung	1
2. Material und Methoden	5
2.1. Organismen und Plasmide	5
2.2. Nährmedien	6
2.3. Zellanzucht und Stammhaltung	7
2.3.1. Zellanzucht	7
2.3.2. Ermittlung der maximalen Wachstumsrate	8
2.3.3. Anzucht von B. licheniformis in kontinuierlicher Kultur (Chemostat)	9
2.3.4. Limitierung des Wachstums im Chemostat durch die Konzentration	12
der C- oder N-Quelle	12
2.3.5. Anaerobe Anzucht von Bacillus licheniformis	12
2.3.6. Reinheitskontrolle	12
2.3.7. Stammhaltung	12
2.4. Techniken für das Arbeiten mit DNA	13
2.4.1. Geräte und Lösungen	13
2.4.2. Isolierung chromosomaler DNA	13
2.4.3. DNA-Fällung mit Isopropanol	14
2.4.4. Isolierung von Plasmid DNA	14
2.4.5. Aufreinigung von PCR-Produkten	15
2.4.6. Isolierung von DNA-Fragmenten aus Agarosegelen	15
2.4.7. Scheren von DNA	15
2.4.8. DNA Konzentrationsbestimmung	16
2.4.9. Agarose Gelelektrophorese	16
2.5. Techniken für das Arbeiten mit RNA	18
2.5.1. Geräte und Lösungen	18
2.5.2. Zellaufschluss von B. licheniformis Zellen mit der Zellmühle	18
2.5.3. Isolierung von RNA	19
2.5.4. DNase Verdau und Qualitätskontrolle der RNA	19
2.5.5. Phenol/Chloroform Extraktion und Fällung von RNA	19
2.5.6. RNA Konzentrationsbestimmung	20
2.6. Modifikaton von Nukleinsäuren	20
2.6.1. Restriktion	20
2.6.2. Dephosphorylierung	21
2.6.3. Herstellung von blunt ends	21
2.6.4. Ligation	21
2.6.5. Transformation von E. coli DH5α-Zellen durch Elektroporation	22

2.6.6. Hitzetransformation von E. coli XL10-Gold Zellen (Inoue et al., 1990)	22
2.6.7. Blau-Weiß Screening (X-Gal-Test)	24
2.7. Polymerasekettenreaktion (PCR)	24
2.7.1. 'Standard' PCR	25
2.7.2. Long-range PCR	25
2.7.3. Reverse Transkription PCR	26
2.7.4. Quantitative Real-time PCR	27
2.8. Herstellung einer Plasmidgenbank	29
2.9. Sequenzierung	29
2.9.1. Sequenzierung mit MegaBace	30
2.9.2. Sequenzierung mit ABI377	31
2.10. Sequenzierungs-Software	32
2.11. Editierung der Sequenzen	33
2.11.1. "Plasmid walking"	33
2.12. Lückenschluss ( <i>Gap closure</i> )	35
2.12.1. Lückenschluß über den Vergleich der automatisch annotierten Sequenz	
von B. licheniformis mit dem Genom von B. subtilis	36
2.12.2. Lückenschluß mit Hilfe der Vectoretten Technik (Kilstrup and Kristiansen, 2000	) 37
2.12.3. Lückenschluß mittels kombinatorischer Multiplex PCR	39
2.12.3.1. Kombination der Primer nach Tettelin et al., 1999 (modifiziert)	39
2.12.3.2. Kombination der Primer mit Hilfe des Kombinationsdreiecks	40
2.13. Qualitätsverbesserung der Sequenz ( <i>Polishing</i> )	41
2.14. Annotation und Sequenzanalyse	42
2.14.1. ORF Vorhersage	42
2.14.2. Automatische Annotation - ERGO (Overbeek et al., 2003)	43
2.14.3. Funktionelle Annotation durch Vergleich lokaler Sequenzhomologien (BLAST)	45
2.14.4. Funktionelle Annotation über Homologie zu Bacillus subtilis	45
2.14.5. Komparative Genomanalyse von B. licheniformis, B. subtilis und B. halodurans	45
2.14.6. Sequenzanalyse basierend auf der codon usage – SIGI (Merkl, 2004)	46
2.15. Erstellen des Genomchips	46
2.15.1. Primergenerierung	46
2.15.2. Amplifizierung der "probes"	48
2.15.3. Spotten und UV-Immobilisierung	48
2.16. Hybridisierung und Auswertung von Microarrays	49
2.16.1. Markierung von RNA mit Cy3/Cy5 Fluoreszenzfarbstoff	50
2.16.2. Wavelength scan zur Überprüfung der Markierungsreaktion	52
2.16.3. Hybridisierung der <i>probes</i> mit markierter <i>target</i> -cDNA	52
2.16.4. Quantifizierung der Arraydaten mit ,GenePix Pro 4.0'	54
2.16.5. Analyse der Transkriptionsdaten	56

3.	Ergebnisse	58
	3.1. Das Sequenzierungsprojekt <i>Bacillus licheniformis</i>	58
	3.1.1. Sequenzierungsmethode	58
	3.1.2. Rohsequenzierung	59
	3.1.3. Editierung der Sequenz	60
	3.1.3.1. Sequenzverbesserung mittels Plasmid walking	60
	3.1.3.2. Verlängerung der Contigs mittels Plasmid walking	61
	3.1.3.3. Identifizierung und Überprüfung von rRNA-Clustern und repetitiven Sequenzen	61
	3.1.4. Lückenschluß (, <i>Gap closure</i> ')	62
	3.1.4.1. Lückenschluß über den Vergleich mit homologen "ORF-Clustern" in B. subtilis	62
	3.1.4.2. Lückenschluß mittels Vectoretten Technik	63
	3.1.4.3. Lückenschluß über kombinatorische Multiplex PCR	63
	3.1.5. Qualitätsverbesserung ( <i>Polishing</i> )	64
	3.1.6. Zusammenfassung	65
	3.2. Annotation des Genoms von Bacillus licheniformis	66
	3.3. Analyse des Genoms von Bacillus licheniformis	67
	3.3.1. Allgemeine Eigenschaften des B. licheniformis Genoms	67
	3.3.2. Ermittlung des Replikationsursprunges	70
	3.3.3. Identifizierung von Fremdgenen und hochexprimierten Genen	71
	3.4. Genomevergleich von B. licheniformis mit B. subtilis	73
	3.4.1. Genomvergleich auf Ebene der Aminosäuresequenz	73
	3.4.1.1. Orthologe Gene in B. licheniformis und B. subtilis	73
	3.4.1.2. Gene von B. licheniformis ohne entsprechendes Ortholog in B. subtilis	74
	3.4.1.3. Industriell relevante Exoenzyme mit und ohne Ortholog in B. subtilis	76
	3.4.2. Genomvergleich auf Ebene der Nukleotidsequenz	78
	3.4.2.1. Alignment der Genomsequenzen	79
	3.4.2.2. Konservierung funktioneller Nukleotidsequenzen	79
	3.5. Genomvergleich von Bacillus licheniformis mit anderen Bacilli der	
	B. subtilis-Gruppe II	82
	3.5.1. Bacilli der <i>Bacillus subtilis</i> Gruppe II	82
	3.5.2. Vergleich der Genomdaten	82
	3.5.3. Vergleich der Proteome von B. licheniformis, B. subtilis und B. halodurans	84
	3.6. Wachstum von Bacillus licheniformis in kontinuierlicher Kultur	85
	3.6.1. Optimierung des Wachstums von B. licheniformis und die Verwertung verschiedener	•
	C-Quellen	86
	3.6.2. Limitierung des Wachstums durch C- und N-Quelle im Chemostat	86
	3.6.3. Anzucht von B. licheniformis mit unterschiedlichen C-Quellen im Chemostat	88
	3.7. Transkriptionsanalyse von Bacillus licheniformis DSM13	89
	3.7.1. Erstellen des Genomchips	89

	3.7.2. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit Acetat	90
	3.7.2.1. Expression der Gene des C-Stoffwechsels	91
	3.7.2.2. Expressionsanalyse der Isocitrat-Lyase und der Malat-Synthase mittels	
	Real-time PCR	95
	3.7.2.3. Expressionsdaten anderer auffällig regulierter Enzyme	96
	3.7.3. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit	
	2,3-Butandiol	97
	3.7.3.1. Expression der Gene des C-Stoffwechsels	97
	3.7.3.2. Expressionsdaten anderer auffällig regulierter Enzyme	98
	3.7.4. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit	
	Gluconat	99
	3.7.4.1. Expression der Gene des C-Stoffwechsels	100
	3.7.4.2. Expressionsanalyse des Gluconat-Operons mittels Real-time PCR	102
	3.7.4.3. Expressionsdaten anderer auffällig regulierter Enzyme	103
	3.7.5. Genexpression des Wachstums unter Glucoselimitierung im Vergleich zum	
	Wachstum unter Stickstofflimitierung	104
	3.7.5.1. Expression der Gene des C-Stoffwechsels	105
	3.7.5.2. Expressionsdaten anderer auffällig regulierter Enzyme	105
	3.7.6. Expression hypothetischer und konserviert hypothetischer ORF's	107
	3.7.7. Zusammenfassung	110
4. D	Diskussion	112
	4.1. Allgemeine und komparative Analyse des Genoms von B. licheniformis	112
	4.1.1. Generelle Eigenschaften des Genoms von B. licheniformis	112
	4.2. Das genetische Potential von B. licheniformis und B. subtilis im Vergleich	117
	4.2.1. C-Stoffwechsel	118
	4.2.2. Anaerobes Wachstum	119
	4.2.3. N-Stoffwechsel	121
	4.2.4. Sekretionsapparat und Exoenzyme	122
	4.2.5. Antibiotika, Sekundärmetabolite und Siderophore	127
	4.2.6. Sporulation	128
	4.2.7. Restriktionssysteme	128
	4.3. Verifizierung des genetischen Potentials von B. licheniformis mittels	
	genomweiter Transkriptionsanalyse	129
	4.3.1. Aspekte des C-Stoffwechsels	130
	4.3.1.1. Genexpression unter Verwertung von C2-Substraten	130
	4.3.1.2. Vergleichende Transkriptionsanalyse bei C- und N-Limitierung	136
	4.3.1.3. Vergleichende Transkriptionsanalyse bei Wachstum mit Gluconat bzw.	
	Glucose als C-Quelle	138

	4.3.2. Bestätigung hypothetischer und konserviert hypothetischer ORF's	
	aufgrund transkriptionsanalytisch ermittelter Regulation	139
	4.4. Ausblick	140
5. 2	Zusammenfassung	141
	5.1. Das Genomprojekt <i>Bacillus licheniformis</i>	141
	5.2. Komparative Genomanalyse von B. licheniformis DSM13 mit B. subtilis	142
	5.3. Genomvergleich von B. licheniformis DSM13 mit anderen Bacilli der	
	B. subtilis Gruppe II	143
	5.4. Vergleich der Genome von B. licheniformis, B. subtilis und B. halodurans auf der	
	Ebene der Aminosäuresequenz	143
	5.5. Untersuchung des Wachstums von B. licheniformis	144
	5.6. Transkriptionsanalytische Untersuchung von B. licheniformis	144
Lit	eraturverzeichnis	146
Α.	Anhang	160
	A.1. Expressionsdaten ausgewählter B. licheniformis Gene der Transkriptionsanalyse	
	Acetat gegen Glucose	160
	A.2. Expressionsdaten ausgewählter B. licheniformis Gene der Transkriptionsanalyse	
	2,3-Butandiol gegen Glucose	163
	A.3. Expressionsdaten ausgewählter B. licheniformis Gene der Transkriptionsanalyse	
	N-unlimitiert/Glucose-limitiert gegen Glucose-unlimitiert/N-limitiert	168
	A.4. Hochexprimierte Gene in B. licheniformis	170
	A.5. Fremdgene (Aliengenes) in B. licheniformis	173

# Abkürzungsverzeichnis

A Adenin
Abb. Abbildung
Amp Ampicillin

aa Aminosäure(n), *amino acids*ATP Adenosin-5'-triphosphat

B. Bacillus

(bi)dest. (zweifach) destilliert

BLAST Basic Local Alignment Search Tool

BLASTN BLAST auf Nukleotidebene
BLASTP BLAST auf Proteinebene

bp Basenpaar(e)bzw. beziehungsweise°C Grad CelciusC Cytosin

ca. circa

cDNA komplementäre Desoxyribonukleinsäure (complementary)

Cm Chloramphenicol

cm Zentimeter

cm<sup>2</sup> Quadratzentimeter

CoA Coenzym A
d.h. das heißt
Da Dalton

dCTP Desoxyribocytosin-5'-triphosphat

dest. destilliert

DH Dehydrogenase
DMSO Dimethylsulfoxid

DNA Desoxyribonukleinsäure

dNTP Desoxyribonukleotid-5'-triphosphat

DSM Deutsche Stammsammlung der Mikroorganismen

E-Cup Eppendorf-Reaktionsgefäß
EDTA Ethylendiamintetraessigsäure

et al. et alteri, und andere

Fa. Firmag GrammG Guanin

Gap Genomassemblierungsprogramm, genome assembly program

Glc Glucose

GLIMMER Gene Locator and Interpolated Markov Modeler

GmbH Gesellschaft mit beschränkter Haftung

Gnt Gluconat h Stunde

IPTG Isopropyl- $\beta$ -thiogalactopyranosid

IS Insertionssequenz(en)

J Joule k Kilo (10<sup>3</sup>)

kb Kilobasenpaare

l Liter

LB Luria-Bertani  $\mu$  Mikro (10<sup>-6</sup>) m Meter; Milli (10<sup>-3</sup>)

M Molar (Mol pro Liter); Mega (10<sup>6</sup>)

MCS multiple Klonierungsstelle, multiple cloning site

min Minute

mRNA Boten Ribonukleinsäure (messenger)

mol Mol

MW Mittelwert n Nano (10<sup>-9</sup>)

NAD(H) Nicotinamidadenindinukleotid (reduzierte Form)

NB Nutrient broth

NCBI National Center for Biotechnology Information

Nr. Nummer

OD Optische Dichte

ORF offener Leserahmen, open reading frame

p pico (10<sup>12</sup>)

PCR Polymerasekettenreaktion, polymerase chain reaction

PFAM Datenbank für Protein-Familien und Hidden-Markov-Modelle pH negativer dekadischer Logarithmus der Protonenkonzentration

r Resistenz

RNA Ribonukleinsäure

rRNA ribosomale Ribonukleinsäure

RT Raumtemperatur

s Sekunde s. siehe

SD Standardabweichung SDS Natriumdodecylsulfat

Sec Sekretion

SIGI score-based identification of genomic islands

sp. speciesT Thymin

TEMED N, N, N', N'-Tetramethylethylendiamid

Tm Schmelztemperatur

Tris Tris-(hydroxymethyl-)aminomethan

tRNA transfer Ribonukleinsäure

U Unit (Einheit der Enzymaktivität)

ü. N. über Nacht

Upm Umdrehungen pro Minute

UV Ultraviolett

V Volt

Vol. Volumen

v/vVolumeneinheit pro Volumeneinheitw/vMasseneinheit pro Volumeneinheit

x Mal

X-Gal 5-Brom-4-Chlor-3-Indolyl-D-Galactosid

z.B. zum Beispiel

#### Aminosäuren

Α	Ala	Alanin	M	Met	Methionin
С	Cys	Cystein	N	Asn	Asparagin
D	Asp	Aspartat	Р	Pro	Prolin
Е	Glu	Glutamat	Q	Gln	Glutamin
F	Phe	Phenylalanin	R	Arg	Arginin
G	Gly	Glycin	S	Ser	Serin
Н	His	Histidin	Т	Thr	Threonin
I	lle	Isoleucin	V	Val	Valin
K	Lys	Lysin	W	Trp	Tryptophan
L	Leu	Leucin	Υ	Tyr	Tyrosin

# 1. Einleitung

Noch bis vor etwa 10 Jahren beschränkte sich die mikrobielle Grundlagenforschung hinsichtlich der Nukleotidsequenz auf die Untersuchung einzelner Gene. Die Entwicklung halbautomatischer Sequenziermaschinen ermöglichte einen enormen Sprung zur Sequenzierung vollständiger Genome, welche in der Entschlüsselung des menschlichen Erbguts (Lander et al., 2001; Venter et al., 2001) gipfelt. In den Jahren 1995 bis 1997 wurden die ersten bakteriellen und archaeellen Genome öffentlich zugänglich, darunter die Genomsequenzen von Haemophilus influenzae (Fleischmann et al., 1995), Mycoplasma genitalium (Fraser et al., 1995), Synechocystis sp. PCC6803 (Kaneko et al., 1996), Methanococcus jannaschii (Bult et al., 1996), Mycoplasma pneumoniae (Himmelreich et al., 1996), Escherichia coli (Blattner et al., 1997), Methanobacterium thermoautotrophicum (Smith et al., 1997), Heliobacter pylori (Tomb et al., 1997) und Bacillus subtilis (Kunst et al., 1997). Bis zum heutigen Tag kann über elektronische Datenbanken auf die Sequenz von insgesamt 20 Archaea und 177 Bacteria zugegriffen werden. Aktuelle Projekte beschäftigen sich derzeit mit der Sequenzierung von weiteren 10 archaeellen und 258 bakteriellen Genomen (Quelle: Genbank, 12.11.2004). Auf diesem Wege eröffneten sich vollkommen neue Forschungsansätze auf einem genomweiten Niveau hinsichtlich der Überprüfung und Umstellung der bisher gültigen Stammbäume, der Erschließung phäno- und genotypischer Zusammenhänge, der Untersuchung evolutionärer Verknüpfungen durch horizontalen Gentransfer bis hin zur Analyse regulatorischer und metabolischer Mechanismen mittels DNA-Microarray Technologie.

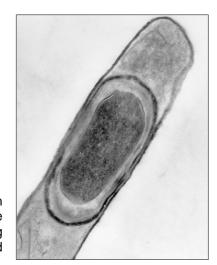
Eine kommerzielle Anwendung dieser Arbeiten findet sich in der Erschließung und Optimierung industriell nutzbarer Mikroorganismen. Einen nicht unbedeutenden Anteil der wichtigsten industriellen Enzymproduzenten nehmen modifizierte *Bacillus* Stämme, mit der Fähigkeit bis zu 20 - 25 g/l extrazellulärer Enzyme zu synthetisieren (Schallmey et al., 2004), ein. Die Produktpalette dieser industriell genutzten Bacilli umfasst alkalische Proteasen, Amylasen, Pullulanasen, Pectat-Lyasen und β-Glucanasen, die in der Reinigungsmittel-, Stärke-, Textil-, Back- und Getränkeindustrie Einsatz finden (Schallmey et al., 2004). Des Weiteren werden Antibiotika, wie Bacitracin und Surfactin, und eine Vielzahl an heterolog exprimierten Enzymen und Hormonen, wie z.B. Interferone und Proinsulin, von Bacilli synthetisiert (Schallmey et al., 2004). Bedeutende Produzenten alkalischer Serin-Proteasen (Subtilisine) sind *B. licheniformis*, *B. subtilis* und *B. pumilus*. Von besonderem Interesse ist auch die hitze- und pH-stabile α-Amylase von *B. licheniformis* (Declerck et al., 2000; Yuuki et al., 1985), welche bei Temperaturen bis zu 110°C über kurze Zeitperioden Aktivität zeigt. Andere Produkte, welche

durch *B. licheniformis* fermentativ hergestellt werden, sind das Antibiotikum Bacitracin (Froyshov and Laland, 1974) und die oberflächen- und antibiotisch-wirksamen Lipopeptide Surfactin (Fox and Bala, 2000; Moran et al., 2000; Peypoux et al., 1999) und Lichenysin (Fiechter, 1992).

B. licheniformis ist ein ubiquitäres Gram-positives Bodenbakterium, das weltweit aus sich zersetzendem Pflanzenmaterial und aus Böden isoliert werden kann (Sneath et al., 1986). Der mesophile Organismus kann sowohl aerob, als auch anaerob bis zu einer maximalen Temperatur von 55°C wachsen, wobei die optimale Wachstumstemperatur zwischen 32 und 45°C liegt (Sneath et al., 1986). Gemeinsam mit anderen bekannten Spezies der Gattung Bacillus gehört B. licheniformis zur B. subtilis Gruppe (Gruppe II) der Bacilli. Die vollständige Genomsequenz einiger Vertreter dieser Gruppe liegt bereits vor. Dazu gehören der "Heubacillus" B. subtilis (Kunst et al., 1997); der alkaliphile, marine B. halodurans (Takami et al., 2000); B. cereus ATCC 10987 (Ivanova et al., 2003) und der Milzbranderreger B. anthracis Ames (Read et al., 2003). Der Typstamm B. cereus ATCC 10987 gilt nicht als pathogen, wobei einige Stämme, dieser auf Getreideprodukten lebenden Spezies, aufgrund der Synthese mehrerer Exotoxine als pathogen eingestuft werden. B. licheniformis wird bisher von der amerikanischen FDA (Food and Drug Administration) als ungefährlich für Menschen, Tiere und Pflanzen eingestuft. Dabei soll allerdings nicht unerwähnt bleiben, dass vereinzelt über toxinproduzierende B. licheniformis Stämme berichtet wurde (Salkinoja-Salonen et al., 1999). Allen Vertretern der B. subtilis Gruppe (Gruppe II) ist die Produktion von Säuren bei Wachstum auf diversen Zuckern und die Ausbildung ovaler, zentraler oder subterminaler, in der Mutterzelle lokalisierter Endosporen unter Mangelbedingungen gemein (Priest, 1993). Mit ihrer Fähigkeit sowohl aerob, als auch teilweise anaerob zu wachsen, stellen sie eine Zwischenstufe zwischen den fakultativen Anaerobiern der Gruppe I (Bacillus polymyxa Gruppe) und den strikt aeroben Bacilli der Gruppen III (Bacillus brevis Gruppe) und IV (Bacillus sphaericus Gruppe) dar (Priest, 1993).

Basierend auf der Genomsequenzierung und der sich daraus ergebenden Menge an genetischer Information eröffneten sich neue unerwartete Dimensionen der molekularbiologischen Forschung. Für die darauf basierenden Anwendungen entstand der Begriff *"functional genomics*" (Velculescu et al., 1997). Aus den klassischen Southern und Northern Blott Untersuchungen für die Detektion spezifischer DNA und mRNA (Alwine et al., 1979; Alwine et al., 1977; Southern, 1975) entwickelte sich, über die Nutzung von Macroarrays, die DNA-

Microarray Technologie mit der einhergehenden Möglichkeit, immer umfassendere Datenmengen in einzelnen Experimenten zu generieren. Die zunehmende Automatisierung von Laborroutinen ermöglicht das Aufbringen von bis zu 50000 DNA Fragmenten auf einen einzelnen Mikroskopobjektträger (DeRisi et al., 1996; Shalon et al., 1996), welche simultan mit zwei cDNA Populationen, markiert mit zwei Fluoreszenzfarbstoffen, hybridisiert werden (Shena et al., 1995). Das Feld der Anwendung für diese Technologie ist groß und reicht von der Untersuchung der Genexpressionsmuster einzelner Organismen in vivo unter verschiedenen Umweltbedingungen, über das Screening nach neuen antimikrobiellen Wirkstoffen durch die pharmazeutische Industrie, bis hin zum genomotyping. Unter genomotyping versteht sich die Hybridisierung Gen-spezifischer Microarrays mit der genomischen DNA zu vergleichender Bakterienstämme mit dem Focus auf evolutionäre Zusammenhänge. Eine bedeutende Rolle kommt hier dem lateralen Gentransfer, als wichtigem Mechanismus zur Adaption an ökologische Nischen, zu. Genomweite Expressionsuntersuchungen hingegen ermöglichen ein wachsendes Verständnis einer globalen Genexpression als Antwort auf Umweltbedingungen (Lucchini et al., 2001). Die erhöhte Expression bekannter Gene in einem gemeinsamen Kontext mit uncharakterisierten Genen gestattet ebenfalls eine hypothetische funktionelle Zuordnung der letzteren, wobei deren genaue Rolle anschließend weiterhin über molekularbiologische und biochemische Techniken studiert werden muss.



**Abbildung 1:** Elektronenmikroskopische Aufnahme von *B. licheniformis* DSM13 mit einer zentral in der Mutterzelle lokalisierten Endospore. (freundlicherweise zur Verfügung gestellt von Michael Hoppert, Institut für Mikrobiologie und Genetik der Universität Göttingen)

Diese Arbeit beinhaltet, neben der Erstellung der Gesamtsequenz von *B. licheniformis* DSM13, den Beginn einer funktionellen Genomanalyse über DNA-Microarray Technologie. Die Basis für alle Untersuchungen stellen die vollständige Entschlüsselung der genomischen Sequenz des *Bacillus licheniformis* Wildtypstammes und die Identifizierung codierender Bereiche dar. Basierend auf diesem Fundament erfolgte die Analyse seines genetischen Potenti-

als. Mit Hilfe bioinformatischer Methoden wurde ein umfangreicher Vergleich mit der Genomsequenz des gut untersuchten *B. subtilis* und mit weiteren drei vollständigen Genomen phylogenetisch engverwandter Vertreter der *B. subtilis* Gruppe (Gruppe II) vorgenommen. Die bei der Erstellung des Genoms und der bioinformatischen Interpretation gewonnenen Erkenntnisse erfordern, soweit von Interesse, eine experimentelle Verifikation. Zur Einleitung dieser experimentellen Phase wurde ein *B. licheniformis* Genomchip erstellt und für erste genomweite Transkriptionsanalysen, deren Schwerpunkt auf den zentralen Kohlenstoff-Stoffwechselwegen und der *in vivo* Expression des metabolischen Potentials lag, genutzt.

# 2. Material und Methoden

# 2.1. Organismen und Plasmide

Im Rahmen dieser Arbeit wurden verschiedene Stämme und Plasmidvektoren verwendet. Sie sind in den Tabellen 2.1 und 2.2 dargestellt.

Tabelle 2.1.: Verwendete Organismen

Stamm	Geno-/Phänotyp	Referenz/Herkunft
B. licheniformis DSM13	Wildtyp	Deutsche Stammsammlung
		von Mikroorganismen
Escherichia coli DH5α	F supE44 lacZ M15	(Hanahan, 1983)
	lacU169 hsdR17 recA1	
	endA1 gyrA96 thi-1 relA1 $\lambda^{-}$	
Escherichia coli XL10-Gold	$Tet^r\Delta$ (mcrA) 183 $\Delta$ (mcrCB-	Stratagene, San Diego,
	hsdSMR-mrr)173 endA1 su-	Kalifornien, USA
	pE44 thi-1 recA1 gyrA96 relA1	
	lac Hte [F' proAB lac1 <sup>q</sup> ZΔM15	
	Tn10 (Tet <sup>r</sup> ) Amy Cam <sup>r</sup> ]	

Tabelle 2.2.: Verwendete Plasmide

Plasmid	Größe (kb)	Phäno- und Genotyp <sup>1)</sup>	Referenz	
pTZ19R <sup>Cm</sup>	3	lacZ, MCS, Cm <sup>r</sup>	freundlicherweise zur	
			Verfügung gestellt	
			von A. Christmann	
blaa01-	4-6	Plasmidgenbank:	freundlicherweise zur Verfügung gestellt	
blpz96		genomische DNA aus	vom Labor für Genomanalyse des Insti-	
		B. licheniformis DSM13	tutes für Mikrobiologie und Genetik,	
		in pTZ19R <sup>Cm</sup>	Göttingen	
pUC18	2,68	Amp <sup>r</sup> , lacPOZ', ColE1,	(Vieira and Messing, 1982)	
		ori R		

# 2.2. Nährmedien

Die verwendeten Medien wurden durch Autoklavieren für 20 min bei 120°C sterilisiert. Feste Medien wurden durch Zugabe von 1,5% Agar vor dem Autoklavieren hergestellt. Nicht autoklavierbare Zusätze wie Antibiotika wurden sterilfiltriert. Die Sterilisation von Glucose erfolgte durch Autoklavieren für 10 min bei 120°C.

Luria-Bertani Med	ium (LB):		<b>Nutrient brot</b>	h Medium	( <b>NB</b> ):
(Sambrook et al., 1	989)				
Trypton	10	g	Nutient broth	8	g
Hefeextrakt	5	g	$H_2O_{bidest}$	ad 1000	ml
NaCl	10	g			
$H_2O_{bidest}$	ad 1000	ml			
<b>TB-Medium:</b>			5x Phosphat l	Puffer:	
Trypton	12	g	KH <sub>2</sub> PO <sub>4</sub>	57,75	g
Hefeextrakt	24	g	$K_2HPO_4$	313,50	g
Glycerin 87% (v/v)	4	ml	$H_2O_{bidest}$	ad 500	ml
H <sub>2</sub> O <sub>bidest</sub>	ad 1000	ml			

Für die Anzucht in Flat-Bottom-Blocks (QIAGEN, Hilden) wurden zu 400 ml TB-Medium je 8 ml 5x Phosphatpuffer und Glucose-Stammlösung (10%, v/v) gegeben.

Tabelle 2.3.: Medienzusätze

Zusatz	Stammlösung	Endkonze	entration
Ampicillin	100 mg/ml in 50% EtOH <sup>a</sup>	100	μg/ml
Chloramphenicol	100 mg/ml in H <sub>2</sub> O <sub>bidest</sub>	50	μg/ml
IPTG	100 mg/ml in H <sub>2</sub> O <sub>bidest</sub>	40	μg/ml
X-GAL	40 mg/ml in N,N-Dimethylformamid <sup>b</sup>	40	μg/ml

a) Lagerung bei 4°C b) vor Gebrauch frisch angesetzt.

Zur Anzucht von *B. licheniformis* mit definierter C-Quelle wurde das in Tabelle 2.4. beschriebene Minimalmedium verwendet.

Tabelle 2.4.: Minimalmedium

Bestandteil	Konzentration der Stammlösung	Menge	Konzentration im Medium
NH <sub>4</sub> Cl	1M in H <sub>2</sub> O <sub>bidest</sub>	10 ml	0,01 M
$MgSO_4$	500 mM in H <sub>2</sub> O <sub>bidest</sub>	1 ml	0,5 mM
$CaCl_2$	100 mM in H <sub>2</sub> O <sub>bidest</sub>	1 ml	0,1 mM
Phosphatpuffer	Na <sub>2</sub> HPO <sub>4</sub> 400 mM mit KH <sub>2</sub> PO <sub>4</sub>	100 ml	
	400 mM auf pH 7,5 eingestellt		
SL9		1 ml	
C-Quelle	300 mM	X	15-50 mM
$H_2O_{bidest}$		ad 1000 ml	

#### Spurenelementelösung 9 (SL9) modifiziert nach (Tschech and Pfennig, 1984)

Titriplex I (Nitrilotriacetat)	12,8	g
FeSO <sub>4</sub> x 7 H <sub>2</sub> O	2	g
CoCl <sub>2</sub> x 2 H <sub>2</sub> O	190	mg
MnCl <sub>2</sub> x 2 H <sub>2</sub> O	100	mg
$ZnCl_2$	70	mg
$H_3BO_3$	6	mg
NiCl <sub>2</sub> x 6 H <sub>2</sub> O	24	mg
CuCl <sub>2</sub> x 2 H <sub>2</sub> O	2	mg
$Na_2MoO_4 \times 2 H_2O$	36	mg
$H_2O_{bidest}$	ad 1000	ml

Mit NaOH auf ph 6,5 einstellen und sterilfiltrieren.

Lagerung bei -20°C langfristig oder 4°C.

Phosphatpuffer und Wasser wurden gemeinsam autoklaviert, sämtliche weiteren Lösungen getrennt sterilisiert und später steril hinzu gegeben.

# 2.3. Zellanzucht und Stammhaltung

#### 2.3.1. Zellanzucht

Die Anzucht von *B. licheniformis* erfolgte aerob bei 42°C in Vollmedium (LB- oder NB- Medium). Um eine gute Durchlüftung zu gewährleisten, wurden Flüssigkulturen in Schüttelkolben mit 150 Upm inkubiert (beheizbares Schüttelwasserbad, Gesellschaft für Labortechnik

mbH, Burgwedel). Für das Wachstum auf Festmedien (LB- oder NB-Medium; bzw. Mineralmedium mit Glucose) wurden die beimpften Agarplatten ca. 2 Tage bei 37°C inkubiert.

Die Anzucht von rekombinanten *Escherichia coli* DH5α Zellen für die Erstellung der Genbank mit dem Vektor pTZ19R<sup>Cm</sup> erfolgte in Flat-Bottom-Blocks (QIAGEN, Hilden), die nach dem Animpfen mit luftdurchlässigen Folien (Aerosheet, QIAGEN, Hilden) abgeklebt wurden. Die einzelnen 96 Wells der Blöcke wurden jeweils mit 1,3 ml TB-Medium mit Chloramphenicol, Glucose und Phosphatpuffer (2.2.) befüllt. Die Inkubation erfolgte für 20 - 26 Stunden bei 37°C mit 360 Upm im Brutschrank.

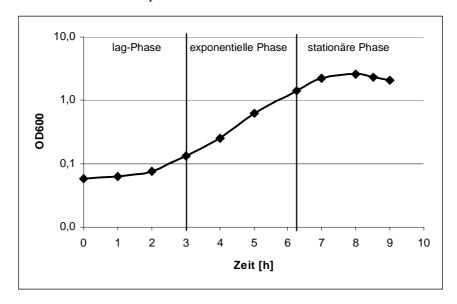
Rekombinante E. coli XL10-Gold Zellen wurden in LB-Medium unter Selektionsdruck angezogen.

# 2.3.2. Ermittlung der maximalen Wachstumsrate

Um die maximale Wachstumsrate  $\mu_{max}$  eines Organismus zu ermittelt, ist es erforderlich den Verlauf des Wachstums in einem geschlossenen System (Batch-Kultur) zu verfolgen. Das Wachstum kann in mehrere Phasen unterteilt und anhand eines typischen Kurvenverlaufes dargestellt werden (Abbildung 2.1.). In der ersten Phase, der lag-Phase, erfolgt zunächst eine Adaption der Zellen an das Medium, dass heißt an C-, N- und/oder P-Quelle. Ist der enzymatische Stoffwechsel auf die neuen Milieubedingungen eingestellt, erfolgt in der so genannten exponentiellen Phase das schnellstmögliche Wachstum unter optimaler Verwertung aller zur Verfügung stehenden Energiequellen. In dieser Phase erfolgt die Bestimmung der maximalen Wachstumsrate μ. Da in einem geschlossenen System keine Zufuhr neuer Energiequellen erfolgt, werden diese stetig, bis zur vollständigen Erschöpfung der ersten essentiellen Resource, verbraucht. An diesem Punkt stagniert das Wachstum und es erfolgt der Übergang in die stationäre Phase. Weitere Faktoren, welche das Ende des exponentiellen Wachstums bewirken, sind z.B. die Anhäufung toxischer Stoffwechselprodukte und eine zu hohe Populationsdichte. In der stationären Phase findet kein Zellwachstum, aber Zellerhaltung durch Verwertung von Speicherstoffen und Abbau von Ribosomen etc. statt. Sind alle Energiequellen verbraucht, setzen die Absterbephase und eine damit verbundene Abnahme der Zelldichte ein. In dieser Arbeit wurde die Wachstumskurve über indirekte Bestimmung der Zelldichte basierend auf der optischen Dichte der Zellkultur bei einer Wellenlänge von 600 nm in Abhängigkeit von der Zeit ermittelt. Die maximale Wachstumsrate, welche in der log-Phase erreicht wird, kann mit folgender Formel berechnet werden:

 $\mu_{max} = (lnx_2 - lnx_2)/(t_2 - t_1) \ [1/h]$   $x_2 - OD_{600} \ zum \ Zeitpunkt \ 2$   $x_1 - OD_{600} \ zum \ Zeitpunkt \ 1$   $t_2 - \ Zeitpunkt \ 2 \ [h]$   $t_1 - \ Zeitpunkt \ 1 \ [h]$ 

Die maximale Wachstumsrate  $\mu_{max}$  ist abhängig von der Art der C-, N- oder P-Quelle, der Belüftung und der Wachstumstemperatur.



**Abbildung 2.1.:** Wachstumskurve; logarithmische Auftragung der Zelldichte ( $OD_{600}$ ) über der Zeit und Unterteilung in die Wachstumsphasen

#### 2.3.3. Anzucht von *B. licheniformis* in kontinuierlicher Kultur (Chemostat)

In einer statischen Kultur ändern sich die Milieubedingungen aufgrund steigender Zelldichte, abnehmender Substratkonzentration und zunehmender Konzentration von Stoffwechselprodukten permanent. Dies sind Faktoren, welche die Schwierigkeit, Zellen reproduzierbar unter exakt den gleichen Bedingungen zu ernten, enorm erhöhen. Eine Möglichkeit, Zellen mit hoher Reproduzierbarkeit der Wachstumsbedingungen über theoretisch beliebig große Zeiträume hinweg anzuziehen und zu ernten, ist daher die kontinuierliche Kultur, bzw. das Wachstum im Chemostat. Dieses offene System basiert im Grunde darauf, dass der wachsenden Bakterienpopulation kontinuierlich frisches Nährmedium zugeführt und im gleichen Maße Bakteriensuspension abgeführt wird und sie sich somit im Fließgleichgewicht (*steady state*) befindet. Entscheidend ist hierbei, ein konstantes Volumen der Bakterienkultur. Beträgt das Volumen der Kultur V (1) und fließt die Nährlösung mit der konstanten Zuflussrate F [1/h] zu, so ergibt sich die Verdünnungsrate D= F/V [1/h]. Wobei D den Volumenwechsel pro Stunde

angibt. Die Stabilität dieses kontinuierlichen Systems beruht auf der Begrenzung der Wachstumsrate durch die Konzentration eines für das Wachstum notwendigen Substrates (N-, S-, P-oder C-Quelle). Diese Substratbegrenzung bewirkt, dass die tatsächliche Wachstumsrate der Zellen kleiner ist als die maximale Wachstumsrate unter Substratsättigung. Somit kann die Verdünnungsrate D in einem relativ weiten Bereich variiert werden. D darf jedoch die maximale Wachstumsrate  $\mu_{max}$  nicht überschreiten, in diesem Falle würden die Zellen ausgewaschen werden. Eine Verdünnungsrate kleiner als die ½ maximale Wachstumsrate hingegen, hätte ein Wachstum der Zellen in die stationäre Phase zur Folge. Somit gilt für die Verdünnungsrate D: ½ $\mu_{max}$  < D <  $\mu_{max}$ . In diesem Bereich kann über die Verdünnungsrate eine konstante Wachstumsrate im Chemostat eingestellt werden.

Eine weiterer entscheidender Grund für den Einsatz dieser Kultivierungsmethode, neben den bereits genannten Vorteilen, ist, dass sich die Zellen permanent im so genannten *balanced growth* befinden, d.h. die relativen Anteile aller Zellbestandteile, wie z.B. Proteine, DNA, RNA und Lipide, bleiben konstant.

In dieser Arbeit wurden *B. licheniformis* Zellen mit dem Ziel der RNA Präparation für die Transkriptionsanalyse in kontinuierlicher Kultur angezogen.

#### Durchführung:

Vorbereitend für eine Zellanzucht mit dem Chemostat wurde der Pumpschlauch für den Medienzulauf (Tygon-LFL, ID: 2,54 mm, ISMATEC) nach einer Stunde Einlauf an der Pumpe (ISM596 B, ISMATEC) geeicht. Hierfür wurde mindestens 3x das in 10 min tatsächlich gepumpte Volumen ermittelt. Die 20 1 Vorratsflasche für frisches Medium wurde mit 5-15 1 NBoder Minimalmedium befüllt. Wurde Minimalmedium verwendet, wurden lediglich H<sub>2</sub>O<sub>bidest</sub> und der Phosphatpuffer eingefüllt und sämtliche weiteren Zusätze erst nach dem Autoklavieren und dem Abkühlen auf RT steril zugegeben. Anschließend wurden die Gefäße des Chemostat über Schlauchsysteme miteinander verbunden (Abbildung 2.2.) und das komplette System für 30 min bei 120°C autoklaviert. Nach dem Anschluss des sterilen Systems an die Begasungs-, Magnetrühreinrichtung und die Zulauf- und Ablaufpumpe wurde das Kulturvolumen von 350 ml in den Chemostat gepumpt. Das Animpfen der Kultur erfolgte mit 10 ml einer Übernachtvorkultur, welche, in dem Versuch entsprechendem Medium, gewachsen war. Unter Begasung mit Druckluft (2 kpa/cm<sup>2</sup>), ständigem Rühren (300 Upm) und einer Temperatur von 42°C wurden die Zellen bis in die log Phase angezogen. Das Wachstum wurde etwa stündlich durch sterile Probenentnahme und über Bestimmung der Zelldichte (OD<sub>600</sub>) verfolgt. Bei einer OD<sub>600</sub> von ca. 1 wurden die Zulauf- und die Ablaufpumpe angeschaltet. Die

Volumenregulation erfolgte durch eine Ablaufrohr im Chemostat, welches an eine Pumpe angeschlossen war. Die Abflusspumpe lief mit deutlich höherer Effizienz als die Zulaufpumpe, wodurch die Höhe des Abflussrohres das Volumen im Chemosaten einstellte und sämtliches Kulturvolumen größer 350 ml abgepumpt wurde. Damit sich ein stabiles Fließgleichgewicht der Kultur einstellen konnte, wurden mindestens 3-5 Volumenwechsel, meist über Nacht, vor einer Probenentnahme abgewartet. Die Chemostatkultur wurde mindestens 1x täglich mikroskopisch und durch Ausstrich auf LB- oder NB-Platten auf Reinheit kontrolliert.

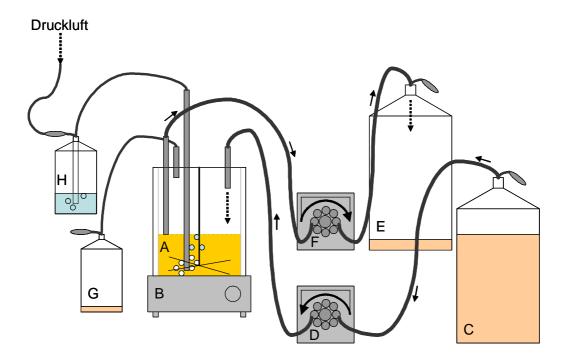


Abbildung 2.2.: Schematischer Aufbau des in dieser Arbeit genutzten Chemostat:

- A doppelwandiges Kulturgefäß, das mit vorgeheiztem Wasser umspült wird (Anschlüsse nicht dargestellt)
- B Magnetrührer
- C Behälter für das Kulturmedium
- D Pumpe für den Mediumzufluss mit definierter Verdünnungsrate
- E Auffanggefäß für abgeführte Zellkultur
- F Pumpe für das Abpumpen der überschüssigen Zellkultur
- G Auffanggefäß bei Überlaufen des Kulturgefäßes
- H Zufluss der Druckluft

#### 2.3.4. Limitierung des Wachstums im Chemostat durch die Konzentration

#### der C- oder N-Quelle

Für die Stabilität einer Kultur im Chemostat, ist die Limitierung eines Substrates entscheidend. Um in dieser Arbeit die entsprechend notwendigen Konzentrationen der N- und C-Quellen für limitiertes, bzw. unlimitiertes Wachstum zu ermitteln, wurden mehrere Chemostat gefahren. Hierbei wurde die Konzentration der zu untersuchenden Energiequelle, unter unlimitierte Zugabe der jeweils entsprechenden N- oder C-Quelle, sukzessive erhöht. Nach jeder Konzentrationserhöhung wurden mindestens 3 Volumenwechsel bis zur Einstellung des Fließgleichgewichtes abgewartet und die Zelldichte über die Messung der OD<sub>600</sub> ermittelt.

#### 2.3.5. Anaerobe Anzucht von Bacillus licheniformis

Für die anaerobe Anzucht von *B. licheniformis*, wurde sich dessen Fähigkeit sowohl aerob, als auch anaerob zu wachsen, zunutze gemacht. Für die in dieser Arbeit durchgeführten Vorversuche zum anaeroben Wachstum von *B. licheniformis* konnte auf Ent- und Begasung des Mediums verzichtet werden. Stattdessen wurden 15 ml Anaerobenröhrchen mit ca. 14 ml Medium befüllt, luftdicht verschlossen, mit steriler Spritze angeimpft und bei 37/42°C inkubiert. Auf diese Weise verbrauchte der fakultative Anaerobier *B. licheniformis* zunächst durch aerobes Wachstum den verbleibenden Sauerstoff wodurch anaerobe Kulturbedingungen entstanden.

#### 2.3.6. Reinheitskontrolle

Die Reinheitskontrolle von Kulturen erfolgte durch einen Verdünnungsausstrich auf Vollmediumplatten und wurde durch eine mikroskopische Kontrolle der Zellmorphologie ergänzt.

#### 2.3.7. Stammhaltung

Über einen Zeitraum von etwa 4 Wochen können Bakterienkulturen auf Agarplatten bei 4°C gelagert werden. Der Endosporen bildende *B. licheniformis* DSM13 kann nahezu unbeschränkt auf Agarplatten aufbewahrt werden. Zur Vermeidung von Austrocknung und Kontaminationen wurden die Petrischalen mit Parafilm verschlossen. Alle 1-2 Wochen sollten die Zellen auf eine frische Agarplatte überimpft werden.

Für die Langzeitkonservierung wurden Stammkulturen mit DMSO (Dimethylsulfoxid) oder Glycerin angelegt. 900 µl einer ÜN-Kultur wurden mit 100 µl DMSO bzw. 100 µl Glycerin

(87%) gemischt und bei -70°C eingefroren. Mit diesen Kulturen können sowohl Fest-, als auch Flüssigmedien direkt angeimpft werden.

#### 2.4. Techniken für das Arbeiten mit DNA

#### 2.4.1. Geräte und Lösungen

Um eine Kontamination durch Nukleasen zu vermeiden, wurden alle thermostabilen Lösungen, Glaswaren und andere Materialien, die für das Arbeiten mit Nukleinsäuren notwendig waren, autoklaviert oder ggf. abgeflammt. Thermolabile Medienzusätze, wie Antibiotika, wurden steril filtriert. Des Weiteren wurde der Arbeitsplatz durch regelmäßige Reinigung mit 70%igen (v/v) Ethanol sauber gehalten.

# 2.4.2. Isolierung chromosomaler DNA

Für die Isolierung chromosomaler DNA aus *B. licheniformis* wurde eine 30 ml Übernachtkultur in LB-Medium bei 37°C angezogen, diese zu 1,5 ml aliquotiert und die Zellen durch Zentrifugation bei 14000 Upm für 1 min pelletiert. Die Zellpellets konnten bis zur DNA-Präparation bei -20°C gelagert werden.

Die chromosomale DNA wurde mit Hilfe des Auqapure Kits der Firma BIORAD, München, isoliert.

Für die Lyse der Zellen wurde das Zellpellet in 90 μl sterilem TE-Puffer resuspendiert. Nach Zugabe von 10 μl Lysozym (100 mg/ml) erfolgte eine Inkubation des Ansatzes bei 37°C für 30 min unter gelegentlichem Schwenken. Anschließend wurden zu dem Ansatz 600 μl Lysis Solution und 20 μl Proteinase K (15 mg/ml) gegeben und es erfolgte eine weitere Inkubation für 45 min bei 50°C. Im Anschluss wurde die Proteinase K für 5 min bei 80°C inaktiviert. Nachdem die Proben auf Eis abgekühlt wurden, erfolgte der Verdau enthaltener RNA mittels Zugabe von 3 μl RNase A Solution. Der Ansatz wurde 25x geschwenkt und für 30 min bei 37°C inkubiert. Waren die Proben anschließend auf Raumtemperatur abgekühlt, wurden 100 μl Protein Precipitation Solution hinzugegeben, gut gemischt und die denaturierten Proteine bei 13000 Upm für 3 min abzentrifugiert. Der Überstand wurde in ein neues Eppendorfcup überführt und die enthaltene DNA mit Isopropanol gefällt (2.4.3.).

#### TE-Puffer:

 $\begin{array}{lllll} Tris\text{-HCl} & 10 & mM \\ EDTA & 1 & mM \\ H_2O_{bidest.} & ad \ 1000 & ml \\ pH \ 8,0 & & & \end{array}$ 

# 2.4.3. DNA-Fällung mit Isopropanol

Die DNA-Lösung wurde mit 0,3 Vol. Isopropanol (100%) versetzt. Nach 50x Schwenken wurde die gefällte DNA durch Zentrifugation bei 13000 Upm für 1 min pelletiert. Das Pellet wurde anschließend 1x mit 300  $\mu$ l 70%igen (v/v) Ethanol gewaschen. Darauf folgte ein weiterer Zentrifugationsschritt bei 13000 Upm für 1 min. Nach Abnahme des Überstandes wurde das verbleibende DNA-Pellet ca. 10 min an der Luft getrocknet. Gelöst wurde die DNA anschließend in 50  $\mu$ l DNA Hydration Solution bei 4°C.

# 2.4.4. Isolierung von Plasmid DNA

Zur Isolierung von Plasmid DNA, die nicht zur Sequenzierung bestimmt war, wurde das QIAprep Spin Miniprep Kit der Firma QIAGEN, Hilden benutzt.

Dieser Plasmidisolierungskit ermöglicht die Präparation von hochreiner DNA. Die Methode basiert auf der alkalischen Lyse verbunden mit der Verwendung einer Säule, an welche während der Aufreinigung Plasmid-DNA selektiv bindet. Proteine und auch andere zelluläre Verunreinigungen können nicht binden und werden somit abgetrennt. Die gewonnene DNA ist darüber hinaus frei von RNA, da bereits während der Lyse eine RNase-Behandlung erfolgt. Die Durchführung dieser Plasmid-Isolierung erfolgte gemäß der vom Hersteller mitgelieferten Versuchsvorschrift.

Plasmide für die Rohsequenzierungsphase des Genomprojektes wurden mit Hilfe von Qiagen BioRobot 8000 Robotern isoliert. Dabei kam das QIAprep 96 Turbo BioRobot Kit ebenfalls der Firma QIAGEN, Hilden zum Einsatz. Hierzu wurden die Klone wie unter 2.3.1 beschrieben im 96er Maßstab in Flat Bottom Blocks angezogen und dann ebenfalls im 96er Maßstab am BioRobot weiterverarbeitet.

Die Flat Bottom Blocks mit den angezogenen DH5α Zellen wurden für 20 Minuten bei 3300 Upm abzentrifugiert (Megafuge 2.0, Fa. Heraeus) und der Überstand verworfen. Der Block mit den Zellpellets wurde in den BioRobot 8000 eingesetzt und dann automatisch weiterbearbeitet. Die Schritte der Plasmidisolierung entsprechen der alkalischen Lyse mit anschließender Aufreinigung der DNA durch eine Silikat-Gel-Membran, die bis zu 20 μg Plasmid-DNA

binden kann. Zellreste und denaturierte Proteine werden über einen Filter vor der Auftragung des Lysats auf die Membran abgetrennt. Nach einem Waschschritt erfolgt die Elution der Plasmid-DNA mit 120 µl Millipore-Wasser in 96-Well Mikrotiterplatten mit Flachboden (UV-geeignet, Fa. Greiner).

# 2.4.5. Aufreinigung von PCR-Produkten

Zur Sequenzierung und zum Spotten von PCR-Produkten ist es erforderlich, noch vorhandene dNTPs, Primer, Puffer und Enzym-Reste aus dem PCR-Ansatz zu entfernen. Dazu wurde das QIAquick PCR Purification Kit der Firma QIAGEN, Hilden verwendet. Die Aufreinigung wurde analog zum Protokoll durchgeführt mit der Abweichung, dass die Elution der PCR-Produkte mit 70°C warmen sterilen H<sub>2</sub>O<sub>bidest</sub> erfolgte.

Für die Aufreinigung der PCR-Produkte zur Herstellung des Genomchips von *B. licheniformis* wurde das Verfahren analog im 96 Well Maßstab am BioRobot 8000 durchgeführt. Speziell für diese Anwendung wurde der QIAquick 96 PCR Purification Kit der Firma QIAGEN, Hilden entwickelt.

#### 2.4.6. Isolierung von DNA-Fragmenten aus Agarosegelen

Die Isolierung von DNA aus Agarosegelen erfolgte mit Hilfe des Gel Extraction Kits der Firma QIAGEN, Hilden nach Angaben des Herstellers.

#### 2.4.7. Scheren von DNA

Zur Erstellung von Genbanken wird häufig das Scheren von DNA angewendet. Bei diesem Verfahren wird die DNA durch mechanische Kräfte gebrochen und wobei DNA Fragmente, deren Größenverteilung weitgehend unabhängig von Erkennungsstellen ist, entstehen. Der Vorteil dieser Methode liegt darin, dass man davon ausgehen kann, dass kein Bereich der chromosomalen DNA in Genbanken über- oder unterproportional vertreten ist.

Nach der Isolierung genomischer DNA (2.4.2.) wurde diese über mechanische Scherkräfte mit Hilfe einer umgebauten HPLC Pumpe fragmentiert. Die Fragmentgröße der so gescherten DNA konnte mittels Druckänderung und/oder Dauer der Zirkulation der DNA in dem HPLC-Pumpsystem variiert werden. Für die Herstellung von Fragmenten der Größe 1,5-3 kb wurde in der Regel ein Druck von 0,5 bar und eine Zirkulationsdauer von 10 min gewählt. Allerdings konnte die gewünschte Fragmentgröße nur näherungsweise erreicht werden. Die DNA-Fragmente wurden nach der Zirkulation mit H<sub>2</sub>O aus dem HPLC-System eluiert, aufgefangen und einer Fällung mit Isopropanol zur Aufkonzentrierung unterzogen (2.4.3.).

#### 2.4.8. DNA Konzentrationsbestimmung

Zur Konzentrationsbestimmung wässriger DNA-Lösungen wurde die Absorption im UV/VISSpectrophotometer V-530 (JASCO, Groß-Umstadt) bei 260 nm gemessen. Dabei kam eine 1 ml Quarzküvette mit 1 cm Schichtdicke und einem Mindestvolumen von 50  $\mu$ l zum Einsatz. Der Nullwert wurde mit Wasser eingestellt. Eine OD<sub>260</sub> von 1 entspricht einer Konzentration von 50  $\mu$ g/ml dsDNA, 40  $\mu$ g/ml ssDNA oder 31  $\mu$ g/ml Oligonukleotiden (Sambrook et al., 1989).

Eine grobe Abschätzung der DNA-Konzentration konnte auch anhand der Intensität von Banden in der Agarose Gelelektrophorese vorgenommen werden. Die Konzentration von Plasmiden, welche mit dem BioRobot in 96 Well Platten mit flachem Boden isoliert wurden, wurde mit dem Plattenphotometer PowerWave X (Bio-Tek Instuments, Bad Friedrichshall) bestimmt.

#### 2.4.9. Agarose Gelelektrophorese

Die Agarose-Gelelektrophorese ist eine Standardmethode zur Auftrennung von DNA-Fragmenten nach ihrer Molekülgröße. Negativ geladenen Nukleinsäuren wandern im elektrischen Feld zur Anode, wobei die Migrationsgeschwindigkeit der DNA-Fragmente von mehreren Faktoren abhängt, wie z.B. der Größe der Gelporen (bestimmt durch die Agarosekonzentration), der angelegten Spannung und der Salzkonzentration des Puffers. Es erfolgt eine Auftrennung der DNA-Fragmente entsprechend ihrer Größe, da deren Laufgeschwindigkeit umgekehrt proportional zu ihrer Molekularmasse ist. Als Färbemittel diente Ethidiumbromid, welches an dsDNA bindet. Die als Banden erscheinenden DNA-Fragmente wurden im UV-Licht (254 nm) sichtbar gemacht, fotografisch festgehalten, und analysiert. Der Vergleich mit einem Standard erlaubte sowohl eine Größen-, als auch eine Mengenabschätzung der aufgetragenen DNA-Fragmente.

#### 20x TAE-Puffer, pH 8:

Tris-Base	96,80	g
Eisessig	22,82	ml
EDTA	7,44	g
$H_2O_{bidest}$	ad 1000	ml

#### Ethidiumbromid-Färbebad:

100 μl einer 10 mg/ml Ethidiumbromidlösung wurden in 100 ml H<sub>2</sub>O<sub>bidest</sub> gegeben.

#### Agaroselösung:

In Abhängigkeit von der zu erwartenden Größe der DNA-Fragmente wurde Agarose 0,8 bis 2%-ig in 1x TAE unter kurzem Aufkochen in der Mikrowelle gelöst bis eine klare homogene Lösung entstand.

#### Durchführung:

Die Agaroselösung mit einer Temperatur von etwa 50°C wurde in eine Elektrophoresewanne gegossen und ein entsprechender Kamm eingesetzt. Das auf ebener Fläche erstarrte Gel wurde dann in eine Elektrophorese-Kammer gelegt und mit 1x TAE-Puffer überschichtet. Anschließend wurden die DNA-Proben 5:1 mit Loading-buffer Lösung (Fermentas) gemischt. Dieser diente einerseits zur Beschwerung der Proben, als auch zur Markierung der Lauffront bei 500 bp. Nach vollständiger Durchmischung der Ansätze mit dem Loading-buffer wurden diese in die Geltaschen pipettiert. Für die Auftrennung der DNA-Proben im Gel wurde die Spannung bei 90 - 130 V konstant eingestellt und diese für etwa 45 - 60 min beibehalten, bis die Lauffront ca. 1,5 cm vom unteren Rand des Gels entfernt war. Nach Beendigung des Laufes, wurde die DNA 10 min im Ethidiumbromidbad gefärbt und das Gel anschließend in einem Wasserbad für 10 min entfärbt. In der Gel-Dokumentationsanlage (BioRad<sup>®</sup> GelDoc 1000) wurde die DNA bei UV-Licht durch das eingelagerte Ethidiumbromid sichtbar gemacht und fotografisch dokumentiert.

Für die gelelktrophoretische Überprüfung der PCR-Produkte, hervorgegangen aus den Chip-PCRs (2.15.2.) oder der kombinatorischen Multiplex PCR (2.12.13.), wurden große Gelkammern mit 2x 200 Taschen der Firma POLYMEHR verwendet. Diese Gele wurden mit 90 V für 90 min laufen gelassen.

Zum Vergleich wurde ein DNA-Standard mit bekannten Fragmentgrößen parallel mit der DNA-Probe aufgetragen und im Agarosegel aufgetrennt.

Für Proben mit einer Größe bis zu 10 kb wurde eine 1 kb-Leiter (MBI Fermentas) mit Größenstandards von 0.25 - 10 kb benutzt. Für Proben mit einer Größe von 100 bp bis zu 1 kb wurde eine 100 bp-Leiter (MBI Fermentas) mit Größenstandards von 80 - 1031 bp verwendet.

#### 2.5. Techniken für das Arbeiten mit RNA

# 2.5.1. Geräte und Lösungen

Das Arbeiten mit RNA erfordert äußerst sauberes Arbeiten, um eine Kontamination mit den sehr stabilen RNasen zu vermeiden.

Aus diesem Grunde wurden alle thermostabilen Lösungen, Glaswaren, Pipettenspitzen und andere Materialien, die für das Arbeiten mit RNA notwendig waren, 2x bei 120°C für 30 min autoklaviert. Die Arbeiten wurden ausschließlich mit Handschuhen durchgeführt. Der Arbeitsplatz und sämtliche notwendigen Gerätschaften wie z.B. Pipetten wurden wiederholt mit 70%igem (v/v) Ethanol gesäubert.

#### 2.5.2. Zellaufschluss von B. licheniformis Zellen mit der Zellmühle

Für die Isolierung von RNA aus *B. licheniformis* Zellen ist der Zellaufschluss mittels Lysozym ungeeignet. Allein die Inkubationszeiten und –temperaturen hätten einen vorzeitigen Abbau der RNA zur Folge. In diesem Falle ist ein zügiger Aufschluss der Zellen mit der Zellmühle im tiefgekühlten Zustand von Vorteil.

Bei der Zellmühle handelt es sich um einen Dismembrator (Mikro-Dismembratur U, Braun Biotech, Melsungen) in welchem eine Chromstahlmahlkugel mit einer hohen Geschwindigkeit in einem mit Stickstoff gekühltem PTFE-Zylinder hin und her geschlagen wird. Auf diese Weise werden im Zylinder befindliche Proben (Zellen) zu einem feinen Pulver zermahlen.

Um RNA aus *B. licheniformis* Zellen zu isolieren, wurden 7 - 7,5 x 10<sup>9</sup> Zellen (OD<sub>600</sub>=1 entspricht ca. 4 x 10<sup>8</sup> Zellen/ml) aus dem Kulturgefäß entnommen und zügig bei 4°C geerntet (5000 g, 5 min). Der Überstand wurde verworfen und das Pellet sofort in 200 - 400 μl sterilem TE-Puffer gelöst. Parallel wurden das Teflongefäß und die Stahlkugel in flüssigem Stickstoff gekühlt. Die resuspendierten Zellen wurden nun in den etwas flüssigen Stickstoff und die Metallkugel enthaltenden Teflonzylinder getropft, das Gefäß verschlossen und in die Zellmühle eingespannt. Der Aufschluss erfolgte bei 1600 Upm für 3 min. Wurde das Gefäß im Anschluss geöffnet, enthielt es ein weißes gekühltes Zellpulver. Ab diesem Moment wurde RNase frei gearbeitet und die aufgeschlossenen Zellen vorsichtig in 4 ml RLT Puffer des RNeasy Midi Kits (QIAGEN, Hilden), welchem zuvor 40 μl Mercaptoethanol zugesetzt wurden, resuspendiert. Der RLT-Puffer enthält stark Protein denaturierende Zusätze, so dass die gelösten Zellbestandteile bei 4°C über Nacht bis zur RNA-Präparation aufbewahrt werden konnten.

#### 2.5.3. Isolierung von RNA

Die Isolierung und Reinigung der RNA erfolgte mit Hilfe des RNeasy Midi Kits (QIAGEN, Hilden) nach Anleitung des Herstellers. Für die Elution der RNA wurden 2 x 150 μl RNase freies und steriles H<sub>2</sub>O<sub>bidest</sub> eingesetzt und die RNA bei -70°C bis zur weiteren Verwendung gelagert.

#### 2.5.4. DNase Verdau und Qualitätskontrolle der RNA

Um Verunreinigungen der RNA mit Resten chromosomaler DNA zur vermeiden, erfolgte eine DNase Behandlung der RNA Lösung.

Zu den maximal 300 μl wässriger RNA Lösung, hervorgegangen aus der RNA Isolation, wurden 65 μl 5x DNase Puffer und 10 μl DNase (10 U/μl) (ROCHE, Mannheim) gegeben. Während einer Inkubation des Ansatzes für 30 min bei 25°C wurde verbliebene DNA durch die DNase hydrolysiert. Im folgenden Inkubationsschritt von 10 min bei 70°C wurde die DNase inaktiviert. Um sicher zu gehen, dass sämtliche DNA abgebaut wurde, erfolgte eine Standard-PCR (2.7.1.) mit spezifischen Primern. Chromosomale DNA wurde als Positivkontrolle mitgeführt. Die PCR-Ansätze wurden anschließend gelelektrophoretisch aufgetrennt, wobei lediglich bei der Positivkontrolle ein Amplifikationsprodukt erscheinen durfte. Im anderen Falle musste der DNase Verdau wiederholt werden.

Die Qualität der RNA wurde mittels Reverser Transkription (2.7.3.), ebenfalls mit spezifischen Primern, überprüft. Auch hier erfolgte anschließend eine Auftrennung der Ansätze durch Agarosegelelektrophorese, wobei hier für jeden RNA enthaltenden Ansatz ein Amplifikationsprodukt zu sehen sein musste.

#### 5x DNase Puffer:

MgSO<sub>4</sub> 25 mM Na-Acetat, pH 5 500 mM

# 2.5.5. Phenol/Chloroform Extraktion und Fällung von RNA

Die meisten Nukleinsäurepräparationen enthalten unerwünschte Proteinkontaminationen. Eine klassische Methode zur Aufreinigung stellt die Phenol/Chloroform Extraktion dar.

Die Extraktion von RNA erfolgte unter Zugabe von 1 Vol saurem Phenol. Der Ansatz wurde 10 s gevortext, und 1 min bei 13000 Upm zentrifugiert. In diesem Schritt erfolgt eine Phasentrennung. Die obere, wässrige Phase enthält die gelöste RNA, die denaturierten Proteine hingegen befinden sich in der Interphase. Die obere Phase wurde, unter Vermeidung der Interphase.

phase, in ein neues Eppendorfcup überführt. Die RNA enthaltende, obere Phase wurde anschließend mit 1 Vol Chloroform/Isoamylalkohol (24/1) versetzt, ebenfalls 10 s gevortext, 1 min bei 13000 Upm zentrifugiert und die entstandene obere Phase in ein neues Eppendorfcup abgenommen.

Um Phenolreste zu beseitigen und die RNA aufzukonzentrieren, erfolgte als nächstes eine Alkoholfällung.

Die in Wasser gelöste RNA wurde mit 1/10 Vol 3,3 M NaAcetat pH 5 und 2,5 Vol -20°C kaltem Ethanol 96% versetzt und gemischt. Die Fällung erfolgte bei -20°C über Nacht oder 1 Stunde bei -70°C. Nach einem Zentrifugationsschritt von 30 min bei 13000 Upm und 4°C wurde der Ethanol vorsichtig mit der Pipette abgenommen, ohne das transparente RNA Pellet zu verletzen. Mit 1 ml 70%igem Ethanol wurde das Pellet gewaschen, 10 min bei 13000 Upm und 4°C zentrifugiert und der Ethanol wieder vorsichtig abgenommen. Das Eppendorfcup mit dem RNA Pellet wurde offen in einem mit Alufolien verschlossenem RNase freiem Becherglas bei 37°C für ca. 15 min getrocknet, bis alle Ethanolspuren verdampft waren. Gelöst wurde das Pellet in 40 μl RNase freiem H<sub>2</sub>O<sub>bidest</sub> unter gelegentlichem auf und ab pipettieren 1 h auf Eis und 3 h bei RT. Die Lagerung der RNA erfolgte bei -70°C.

# 2.5.6. RNA Konzentrationsbestimmung

Die Bestimmung der Konzentration von RNA erfolgte analog der Konzentrationsbestimmung von DNA (2.4.7.). Eine  $OD_{260}$  von 1 entspricht einer Konzentration von 40  $\mu$ g/ml RNA (Sambrook et al., 1989).

#### 2.6. Modifikaton von Nukleinsäuren

#### 2.6.1. Restriktion

In dieser Arbeit wurden Restriktionsnukleasen genutzt, um größere PCR-Produkte in der Endphase des Lückenschlusses zu fragmentieren. Auf diesem Wege entstandene kleinere DNA-Abschnitte mit definiertem Ende konnten subkloniert und mit Standardprimern sequenziert werden. Zum Einsatz kam dieses Verfahren, wenn der zu sequenzierende Bereich repetetive Sequenzen enthielt und somit das Generieren spezifischer Primer nicht möglich war.

#### Durchführung:

Die Enzyme wurden in den jeweiligen Puffersystemen und Temperaturspektren der Hersteller benutzt. Es kamen 3 - 5 U Enzym pro µg DNA zum Einsatz. Zur vollständigen Restriktion der PCR-Produkte und der Plasmide wurden die Ansätze für ca. 2 - 3 Stunden bei der

vom Hersteller empfohlenen Temperatur inkubiert. Zur Inaktivierung der Restriktionsenzyme wurden die Ansätze anschließend für 15 min bei 65°C inkubiert.

#### 2.6.2. Dephosphorylierung

Vor der Ligation von DNA-Fragmenten mit überhängenden Enden ("sticky ends") wurde das Insert mit alkalischer Phosphatase behandelt. Das Enzym entfernt von linearer DNA die 5'-Phosphatgruppe und verhindert so, dass Inserts miteinander ligieren.

#### Durchführung:

Die Insert DNA wurde mit 1 U alkalischer Phosphatase (Fermentas) pro μg DNA versetzt und eine Stunde bei 37°C inkubiert. Die Inaktivierung der Phosphatase erfolgte durch Inkubation bei 65°C für 15 min, oder mittels Aufreinigung mit dem QIAquick PCR Purification Kit (s. 2.4.5) (Angaben des Herstellers).

#### 2.6.3. Herstellung von blunt ends

Beim Scheren von DNA wird diese gebrochen, wobei undefinierte Enden entstehen.

Um eine Klonierung vornehmen zu können, werden jedoch definierte Enden benötigt. Zu diesem Zweck werden die Enden der gescherten DNA mit Hilfe der T4-DNA-Polymerase aufgefüllt. Nach dieser Reaktion liegen "blunt ends" vor und die so "geglättete" Insert-DNA kann weiterverarbeitet werden. Der folgende Ansatz wurde für 1,5 h bei RT inkubiert.

#### 50 μl Ansatz:

Insert DNA	ca.	38	μl
5x Puffer T4-DNA-Polymerase		10	μl
dNTP Mix je 10 mM		1,5	μl
T4-DNA-Polymerase		1	μl

Anschließend wurde der Ansatz mit dem QIAquick PCR Purification Kit (s. 2.4.5) aufgereinigt. Die Elution erfolgte mit 50  $\mu$ l  $H_2O_{bidest}$ .

#### 2.6.4. Ligation

DNA-Ligasen schließen die Phosphodiester-Bindung zwischen einer 5'-Phosphatgruppe an einem DNA-Ende und der 3'-OH-Gruppe am anderen DNA-Ende. Um eine effektive Ligation zu gewährleisten, wurde Vektor- und Insert-DNA im Verhältnis 1:3 eingesetzt. Zur Konzentrationsbestimmung von Insert- und Vektor-DNA, wurden zuvor die Restriktionsansätze

gelelektrophoretisch aufgetrennt. Der Ligationsansatz wurde mit 1U T4-DNA-Ligase (PRO-MEGA, Mannheim), sowie dem empfohlenen Puffersystem angesetzt und über Nacht bei 16°C inkubiert.

#### 2.6.5. Transformation von *E. coli* DH5α-Zellen durch Elektroporation

Die Transformation von *E. coli* mit Plasmiden ist unter anderem durch Elektroporation möglich (Fiedler and Wirth, 1988). Im Genomprojekt *B. licheniformis* wurden die Plasmidgenbanken mit elektrokompetenten *E. coli* DH5α-Zellen nach der folgenden Methode hergestellt.

#### Herstellung kompetenter Zellen:

*E. coli* DH5α wurde über Nacht in einer 3 ml Vorkultur bei 37°C angezogen. Am nächsten Morgen wurde mit dieser Vorkultur eine 300 ml Hauptkultur angeimpft und unter starkem Schütteln bis zu einer OD<sub>600</sub> von 0,7 - 0,8 inkubiert. Die Zellen wurden dann für 15 min auf Eis inkubiert, geerntet (5000 Upm, 10 min, 4°C) und zweimal mit eiskaltem H<sub>2</sub>O<sub>bidest.</sub> gewaschen. Das Pellet wurde mit 20 ml einer 10%igen Glycerin-Lösung gewaschen und in 800 μl der gleichen Lösung aufgenommen. Anschließend wurden die Zellen in Aliquots von 70 μl abgefüllt und bei -80°C gelagert.

#### Transformation:

Es wurden  $0,5..2~\mu l$  Ligationsansatz mit 70  $\mu l$  kompententen Zellen in eine auf Eis vorgekühlte Elektroporationsküvette gegeben. Zur Elektroporation wurde ein Gene Pulser (Fa. Biorad) verwendet, wobei eine Spannung von 2,5~kV, eine Kapazität von  $2,5~\mu F$  und ein Widerstand von  $200~\Omega$  eingestellt wurden. Direkt nach der Elektroporation wurde der Ansatz zusammen mit 1~ml LB-Medium in ein Eppendorfcup gegeben und zur Regeneration der Zellen eine Stunde bei  $30^{\circ}C$  und 180~Upm inkubiert. Danach wurden die Zellen auf Selektionsmedium ausplattiert und bei  $37^{\circ}C$  über Nacht inkubiert.

# 2.6.6. Hitzetransformation von *E. coli* XL10-Gold Zellen (Inoue et al., 1990)

#### Herstellung kompetenter Zellen:

Die Hauptkultur (250 ml SOB-Medium, in 1 l-Schikanekolben) wurde mit 5 ml einer über Nacht gewachsenen *E. coli*-LB-Vorkultur angeimpft und bei 18°C im Schüttelwasserbad (INFORS AG, Bottmingen) bis zu einer OD<sub>600</sub> von 0,6 - 0,7 etwa 24 h angezogen. Nach 10-

minütiger Abkühlung auf Eis wurden die Zellen durch Zentrifugation (10 min, 5000 Upm, 4°C, GSA-Rotor) geerntet, in 80 ml eiskaltem TB-Puffer vorsichtig resuspendiert, weitere 10 min auf Eis gestellt und nochmals sedimentiert. Das Pellet wurde in 20 ml kaltem TB-Puffer aufgenommen, mit 1,4 ml DMSO (Endkonzentration: 7% (v/v)) versetzt, sorgfältig gemischt, erneut 10 min auf Eis gestellt und als 200 μl-Aliquots in vorgekühlte Eppendorfcups portioniert. Die Zellen wurden sofort in flüssigem Stickstoff schockgefroren und bei -70°C gelagert.

#### SOB-Medium (Hanahan, 1983):

Trypton	2,0%	(w/v)
Hefeextrakt	0,5%	(w/v)
NaCl	10	mM
KCl	2,5	mM
$MgCl_2$	10	mM
$MgSO_4$	10	mM

Die Magnesiumsalze wurden aus getrennt autoklavierten Stammlösungen (1 M MgCl<sub>2</sub> / 1 M MgSO<sub>4</sub>) zugegeben.

#### TB-Puffer:

Pipes-Puffer	10	mM
$MnCl_2$	55	mM (getrennt autoklaviert)
CaCl <sub>2</sub>	15	mM
KC1	250	mM
pH 6,7		

#### Transformation:

Die eingefrorenen Zellen wurden auf Eis aufgetaut, mit 10 - 150 ng DNA gemischt und zur Adsorption der DNA an die Zellen 30 min auf Eis inkubiert. Die Aufnahme der DNA erfolgte durch einen Hitzeschock bei 42°C für 45 - 60 s. Nach Zugabe von 800 μl LB-Medium wurde der Ansatz 1 h bei 37°C zur Ausprägung der Selektivmarker inkubiert. Anschließend wurden 100 μl der Zellsuspension auf Agarplatten unter Selektivbedingungen ausplattiert und bei 37°C über Nacht inkubiert.

#### 2.6.7. Blau-Weiß Screening (X-Gal-Test)

Zur Selektion rekombinanter Klone ist es entscheidend, nur solche auszuwählen, die sowohl den transformierten Vektor als auch ein Insert im Vektor tragen. Hierzu wurde das so genannte" blau-weiß" Screening verwendet. Für dieses Selektionsverfahren sind E. coli Stämme notwendig, die eine Deletion im lacZ-Gen enthalten (lacZ M15) und somit keine  $\beta$ -Galactosidase mehr bilden können. Klonierungsvektoren wie pBluescript SK<sup>+</sup> und seine Derivate führen durch das lacZ Gen zur Komplementation des  $\beta$ -Galactosidase-Defekts. Zellen, die mit einem solchen Vektor transformiert wurden, können das chromogene  $\beta$ -Galactosid X-Gal spalten, wobei ein Farbstoff freigesetzt wird, der durch Luftoxidation blau wird und die Zellen anfärbt. Durch Zugabe von Isopropyl- $\beta$ -d-thiogalactopyranosid wird indirekt (Bindung des Lac-Repressors) die Transkription der  $\beta$ -Galactosidase induziert. Die bekannten Klonierungsvektoren enthalten im lacZ eine multiple Klonierungsstelle (MCS), so dass die Ligation des Vektors mit einem Insert dazu führt, dass das lacZ Gen unterbrochen und somit die Bildung einer funktionierenden  $\beta$ -Galactosidase unmöglich wird. Die Zellen bleiben also weiß, wenn ein Insert enthalten ist. Zellen ohne Vektor, bei denen die Transformation nicht erfolgreich war, werden durch eine plasmidkodierte Antibiotikaresistenz negativ selektiert.

# 2.7. Polymerasekettenreaktion (PCR)

Die PCR (polymerase chain reaction) dient zur in vitro Amplifikation von DNA. Dafür sind ein Primerpaar, Matrizen-DNA (Template) sowie eine thermostabile DNA-Polymerase notwendig. In dieser Arbeit wurde die PCR zu unterschiedlichen Zwecken eingesetzt; zum Schließen von Sequenzierungslücken (kombinatorische Multiplex PCR, Vectoretten Technik); zur Überprüfung bereits sequenzierter Bereiche im Genom von *B. licheniformis*, zur Amplifikation von ca. 300 bp großen Fragmenten der ORFs des *B. licheniformis* Genoms für die Herstellung des Genomchips; zur Expressionsanalyse mittels *Real-time* PCR und zur Überprüfung von RNA mittels Reverser Transkription. Als Primer wurden, wenn nicht anders erwähnt, Oligonukleotide zwischen 17 und 18 bp und einer Schmelztemperatur zwischen 55 und 60°C verwendet. Alle PCR Reaktionen wurden an den Cyclern PrimusHT oder Primus96<sup>plus</sup> der Firma MWG AG BIOTECH durchgeführt.

#### 2.7.1. 'Standard' PCR

Die Standard-PCR wurde für die Amplifikation von Fragmenten mit einer erwarteten Größe zwischen 500 bp und maximal 3 kb verwendet, eingesetzt wurde die Taq DNA Polymerase der Firma Quiagen, Hilden.

100 µl Ansatz	<u>∸</u>	PCR-	Programm:		
20 μΙ	Primer A (5 μM)	01	LIDTH ON		110°C
20 μΙ	Primer B (5 μM)	02	PAUSE	2 h	4°C
2 μΙ	dNTP's (10mM)	03	LIDCL	150 N	
20 μΙ	5x Q-Solution	04	TEMP	2 min	96°C
10 μl	10x PCR-Puffer	05	LOOP [	25x	
2 μΙ	chromosomale DNA (250 ng/µl)	06	TEMP	30 s	96°C
0,7 μΙ	Taq-Polymerase (5 U/µl)	07	TEMP	30 s	52°C
27,3 µl	H <sub>2</sub> O <sub>bidest</sub>	80	TEMP	2 min 30 s	72°C
		09	LOOP ]		
		10	TEMP	10 min	72°C
		11	STORE	10 s AT	8°C
		12	TEMP	FOREVER	4°C

Nach der Amplifikation wurden die PCR-Produkte gelelektrophoretisch überprüft.

# 2.7.2. Long-range PCR

Die *Long-range* PCR wurde für die Amplifikation von Fragmenten mit einer erwarteten Größe bis zu 15 kb eingesetzt. Es wurde der Triple Master Kit der Firma Eppendorf, Hamburg verwendet.

<u>100 µl</u>	Ansatz:		PCR-I	Programm:			
20	μl	Primer A (5 μM)	01	LIDHT ON			110°C
20	μl	Primer B (5 μM)	02	PAUSE	2 h		4°C
10	μl	Tuning buffer	03	LIDCL	150 N		
5	μl	10 mM dNTPs (each)	04	TEMP	2 min		96°C
0,8	μl	Triple Master Enzymmix	05	LOOP [ 10x			
2	μl	chromosomale DNA (250 ng/μl)	06	TEMP	15 s		96°C
42,2	μl	H <sub>2</sub> O <sub>bidest</sub>	07	TEMP	30 s		52°C
			08	TEMP	5 min		68°C
			09	LOOP ]			
			10	LOOP [ 10x			
			11	TEMP	15 s		96°C
			12	TEMP	30 s		52°C
			13	TEMP	5 min	+ 20 s	68°C
			14	LOOP ]			
			15	STORE	10 s	AT	8°C
			16	TEMP	FORE	VER	4°C

Nach der Amplifikation wurden die PCR-Produkte gelelektrophoretisch überprüft.

# 2.7.3. Reverse Transkription PCR

Bei einer reversen Transkription PCR wird mit Hilfe einer reversen Transkriptase RNA in cDNA (*complementary DNA*) umgeschrieben, welche anschließend mittels PCR amplifiziert werden kann. Die reverse Transkription und die anschließende PCR konnten mit dem in dieser Arbeit verwendeten OneStep RT PCR Kit der Firma Eppendorf in einem Schritt durchgeführt werden.

25 µl F	RT-PCR	Ansatz:	Progra	amm:		
1	μg	Template RNA	01	LIDHT ON		110°C
3	μl	5 μM Primer A	02	PAUSE	2 h	4°C
3	μl	5 μM Primer B	03	LIDCL	150 N	
5	μl	5x Qiagen OneStep	04	TEMP	30 min	50°C
		RT-PCR Puffer	05	TEMP	15 min	95°C
1	μl	10 mM dNTPs (each)	06	LOOP [	35x	
1	μl	Qiagen OneStep	07	TEMP	30 s	94°C
		RT-PCR Enzyme Mix	80	TEMP	45 s	50°C
ad 25	μl	H <sub>2</sub> O <sub>bidest</sub> (RNase free)	09	TEMP	1 min	72°C
			10	LOOP]		
			11	STORE	10 s AT	8°C
			12	TEMP	FOREVER	4°C

Nach der Amplifikation wurden die PCR-Produkte gelelektrophoretisch überprüft.

#### 2.7.4. Quantitative Real-time PCR

Die quantitative Real-time PCR ist eine Methode zur Quantifizierung von Nucleinsäuren und damit sehr gut zur Expressionsanalyse einzelner Gene geeignet. Ein entscheidendes Werkzeug der Real-time PCR sind unterschiedliche Farbstoffe oder Fluorophore, welche an die DNA binden und über deren Signal die Kinetik der PCR ,live' beobachtet werden kann. Die Markierung der DNA kann auf unterschiedlichen Wegen erfolgen: nach dem TaqMan Prinzip, mit molecular beacons, hybridization probes oder unter Einsatz eines Farbstoffes wie z.B. SYBR green. Der entscheidende Vorteil von SYBR green liegt darin, dass dieser Farbstoff unspezifisch in doppelsträngige DNA eingebaut wird, wohingegen für die anderen genannten Markierungsmöglichkeiten die Synthese spezifischer Sonden notwendig ist.

Bei der quantitativen *Real-time* PCR wird über das Fluoreszenzsignal indirekt die Zunahme des PCR Produktes gemessen. In den frühen Runden einer PCR Reaktion findet eine annähernd exponentielle Vermehrung der DNA Fragmente statt, welche später durch die Abnahme von dNTPs, Nachlassen der Enzymaktivität und/oder Anhäufung von Pyrophosphat gehemmt wird. Die Quantifizierung bei der *Real-time* PCR erfolgt über den so genannten Ct-Wert, hierbei handelt es sich um die Zykluszahl, bei welcher sich das Fluoreszenzsignal einer Probe das erste Mal signifikant vom Hintergrund abhebt. Anhand einer Standardkurve können definierten Templatemengen Ct-Werte zugewiesen werden. Um im Falle einer Expressionsanalyse den Regulationsfaktor eines Genes unter der Bedingung A im Vergleich zu einer Bedingung B zu ermitteln, ist die Bestimmung der Templatemenge nicht vonnöten. Nach folgender Formel kann allein über die Ct-Werte eines Gens X unter beiden Bedingungen im Vergleich zu den Ct-Werten eines konstitutiven Gens, ebenfalls unter beiden Bedingungen, der Regulationsfaktor ermittelt werden:

$$\begin{split} Regulations faktor &= 2^{-\Delta\Delta~Ct} \\ mit~\Delta\Delta~Ct~f\"{u}r~Gen~x &= (Ct_x - Ct_{konstitutiv})_{Bedingung~I} - (Ct_x - Ct_{konstitutiv})_{~Bedingung~II} \\ &\qquad\qquad\qquad (Talaat~et~al.,~2002) \end{split}$$

Die Primer für jedes Gen wurden mit einer Länge von 20 - 23 bp und einer Schmelztemperatur von 60°C generiert und so gewählt, dass die Länge des Produktes der *Real-time* PCR möglichst genau 150 bp beträgt.

#### Durchführung:

Sämtliche Arbeiten wurden unter RNase freien Bedingungen durchgeführt.

In dieser Arbeit wurden der QuantiTect SYBR Green RT-PCR Kit (QIAGEN, Hilden) und das iCycler iQ<sup>TM</sup> *Real-time* PCR Detection System der Firma BIO-RAD verwendet. Für den Ansatz der *Real-time* PCR wurde zuerst ein Mastermix aus dem Quantitect Mastermix und dem Enzymmix für alle Proben hergestellt, welcher dann auf die zu testenden RNAs aufgeteilt wurde (Premix). Die Primer wurden in 96 well 0,2 ml Thin-Wall PCR Platten (BIO-RAD) vorgelegt und 20 μl Premix zugegeben. Auf diese Weise wurde mit größtmöglicher Wahrscheinlichkeit die gleiche Konzentration aller Reaktionskomponenten in den einzelnen Ansätzen erreicht.

#### 25 µl quantitative Real-time PCR Ansatz:

2,5	μl	5 μM Primer A
2,5	μl	5 μM Primer B
12,5	μl	Quantitect Mastermix
0,25	μl	Enzymmix
400	ng	RNA
ad 25	μl	H <sub>2</sub> O <sub>bidest</sub> (RNase free)

I-Cycler Prog	gramm:		
Cycle 1:	(1x)		
Step 1:	50.0°C	30 min	(Reverse Transkription)
Cycle 2:	(1x)		
	Step 1: 95.0°C	15 min	(Aktivierung der
			DNA-Polymerase)
Cycle 3:	(45x wiederholt)		
	Step 1: 94.0°C	15 s	(PCR)
	Step 2: 55.0°C	30 s	
	Step 3: 72.0°C	30 s	
Cycle 4:	(1x)		
	Step 1: 55.0°C	1 min	(Abschließende Elongation)
Cycle 5:	(80x wiederholt)		
	Step 1: 55.0°C	10 s	(Ermittlung des Schmelzpunktes
	+0,5°C		der PCR Produkte)
Cycle 6:	(1x)		
	Step 1: 4.0°C	HOLD	

## 2.8. Herstellung einer Plasmidgenbank

Im Genomprojekt *B. licheniformis* wurde eine Plasmidgenbank mit einer Insertgröße von 2 - 3 kb verwendet, für deren Herstellung chromosomale DNA aus *B. licheniformis* mechanisch geschert wurde (2.4.7). Die gescherte DNA wurde mit Hilfe der T4-DNA-Polymerase behandelt (2.6.3), um *blunt ends* zu erzeugen, anschließend auf ein Agarsosegel aufgetragen und die DNA-Fragmente zwischen 2 - 3 kb aus dem Gel eluiert (2.4.6.) Daraufhin erfolgte eine Dephosphorylierung der Insert-DNA (2.6.2.), ihre Ligation (2.6.4.) in pTZ19R (Cm<sup>r</sup>) und letztendlich die Transformation in *E. coli* DH5α (2.6.5.). Insert enthaltende Klone wurden über Blau-Weiß-Screening (2.6.7.) selektiert. Nach Anzucht der positiven Klone unter Selektionsdruck in TB-Medium (2.3.1.) wurden die Plasmid-DNA präpariert (2.4.4.) und mit Standardprimern sequenziert (2.9.).

# 2.9. Sequenzierung

Für die Sequenzierungen im *B. licheniformis* Genomsequenzierungsprojekt kamen zwei verschiedene Sequenzierungsgeräte zum Einsatz. Die Plasmide wurden auf dem MegaBace 1000 (Amersham Biosiences) und dem ABI377 (Applied Biosystems) sequenziert. Diese Geräte

arbeiten mit dem Verfahren der Kettenabbruchmethode (Sanger et al., 1992). Bei diesem Verfahren werden floureszenzmarkierte ddNTPs (dye terminators) eingesetzt. Die markierten ddNTPs werden statt eines dNTPs eingebaut, wodurch es zum Abbruch der Kettenverlängerung kommt. Das so entstandene Fragment ist am 5'-Ende markiert. Diese fluoreszenzmarkierten Fragmente werden von einem Laser im Sequenzierautomaten zur Fluoreszenz angeregt. Das emittierte Licht wird in Detektoren erkannt und in Fluoreszenzpeaks umgewandelt, die wiederum beim "base calling" (2.10) in Sequenzen übersetzt werden.

#### 2.9.1. Sequenzierung mit MegaBace

Die Geräte der MegaBace-Reihe sind Kapillarsequenzierungseinheiten mit 96 (Mega-Bace 1000) Kapillaren. Bei diesem Typ von Sequenzierautomaten wird die elektrophoretische Auftrennung der markierten DNA-Fragmente in dünnen Glaskapillaren (ca. 100 μm Durchmesser), die eine Polyacrylamid-Matrix enthalten, durchgeführt. Einer der Vorteile dieser Methode ist, dass es nicht zu Verwechslungen der einzelnen Sequenzierungsreaktionen kommen kann, wie es bei den Plattensequenzierern z.B. ABI377 möglich ist. Außerdem wird die DNA-Probe nicht vollständig in die Kapillare gegeben, sondern es wird durch Anlegen einer Spannung nur ein Teil der DNA-Moleküle aus dem Sequenzierungsreaktion mehrmals sequenzierer kann, ohne die Reaktion neu ansetzen zu müssen, da beim Injizieren nur ein Teil der DNA in die Matrix einwandert. Nachteilig kann sich allerdings die Empfindlichkeit der Auftragungsmethode (Injektion) der DNA auswirken. Setzt man zu hohe Stromstärken ein, wählt die Injektionszeit zu hoch oder verwendet im Reaktionsansatz zu viel Template-DNA, können die Kapillaren regelrecht "verstopfen". Der folgende Sequenzansatz wurde pipettiert, wobei die DNA gereinigt vorlag.

10 µl MegaBace Sequenzier-Ansatz:		MegaE	Bace cycle seque	encing:		
400	ng	DNA	01	LIDHT ON		110°C
1	μΙ	Primer (5 µM)	02	PAUSE	FOREVER	95°C
4	μΙ	DYEnamic ET	03	TEMP	5 min	95°C
		Dye Terminator Kit	04	LOOP [	30x	
ad 10	μΙ	$H_2O_{HPLC}$	05	TEMP	20 s	95°C
			06	TEMP	15 s	50°C
			07	TEMP	1 min	60°C
			80	LOOP]		
			09	LIDHT OFF		
			10	TEMP	FOREVER	4°C

Anschließend wird das "MegaBace cycle sequencing"-Programm in den entsprechenden Thermocyclern ausgeführt. Zur Aufreinigung der Sequenzansätze wurden zwei Fällungen mit 24 µl Isopropanol (80% v/v) und 40 µl Ethanol (80% v/v) durchgeführt.

#### 2.9.2. Sequenzierung mit ABI377

Bei den ABI377-Geräten handelt es sich um Plattensequenzierer. Hier wird zwischen zwei Glasplatten ein Polyacrylamid-Gel gegossen, in dem die Elektrophorese stattfindet. 96 Proben werden manuell mit Hilfe einer Mehrkanal-Hamilton-Pipette auf das Gel aufgetragen. Der Vorteil dieser Geräte ist, dass sie relativ tolerant gegenüber der aufgetragenen DNA-Menge sind. So ist es auch möglich, bei schwachen Signalen relativ gute Sequenzen zu erhalten. Als Nachteile sind der hohe Arbeitsaufwand beim Gießen der Gele und manuellen Auftragen der Proben zu nennen, sowie die Möglichkeit, dass es durch Ausfall einzelner Spuren oder Überstrahlen einzelner Spuren (bei zu viel DNA im Ansatz) zu Verwechslungen einzelner Sequenzen kommen kann.

#### Durchführung:

Die Zusammensetzung eines ABI Sequenzieransatzes für Plasmide und das entsprechende Cycling-Programm sind im Folgenden aufgeführt. Nach dem Cycle-Schritt erfolgte eine Fällung mit Ethanol (99% v/v) und 3 M Natriumacetat (pH 4,6), wodurch überschüssige Dye-Terminatoren und Primer entfernt wurden.

20 μl ABI Plasmid Sequenzieransatz:			vcle sequencing:		
2 µl	Premix	01	LIDHT ON		110°C
2 µl	half BD	02	PAUSE	FOREVER	98°C
1 µl	Betain (5 M)	03	TEMP	2 min	98°C
1 µl	Primer (5 μM)	04	LOOP [	25x	
400 ng	DNA Template	05	TEMP	20 s	96°C
ad 20 µl	H <sub>2</sub> O <sub>bidest</sub>	06	TEMP	15 s	50°C
		07	TEMP	4min	60°C
		08	LOOP]		
		09	LIDHT OFF		
		10	TEMP	FOREVER	4°C

Für die Sequenzierung am ABI wurden die Gele wie unten beschrieben angesetzt und die Gelektrophorese durchgeführt. Vor dem Auftragen der Proben auf das Sequenziergel wurden diese mit 1 µl Stop-Mix (s. unten) versetzt.

#### ABI-Sequenziergel:

Zusammensetzung eines ABI-Sequenziergels:			ziergels:	Laufparameter für ein ABI Sequenziergel:			
Harnstoff (7M)	18	g		Geldicke	0,2 mm		
Acrylamid (30%, 29:1)	5,3	ml		Gellänge	48 cm		
10x TBE-Pufferlong rui	n	6	ml	Spannung	3.500 V	′	
H <sub>2</sub> O <sub>bidest</sub>		21,5	ml	Stromstärke	60 mA		
10% APS		250	μl	Leistung	40 W		
TEMED		25	μl	Temperatur	45°C		
				CCD	offset 0		
				CCD	gain 4		
				Vorlaufzeit	90 s		
				Laufzeit 1,5 h	+ 14,5 h		
				Laufpuffer 2 x	TTE obe	n, 1 x T	TE unten
				Minimale Anza	hl von S	cans 22	.000
loading-buffer für ABI-S	Sequenz	zansätze	<u>):</u>	Stop-Mix für ABI-Sequenzansätze:			tze:
EDTA	25	mM		Formamid		5	Teile
Blue Dextran	50	mg/ml		loading buffer		1	Teil
H <sub>2</sub> O <sub>bidest</sub>	ad 500	) ml					

# 2.10. Sequenzierungs-Software

Die Verarbeitung der Daten aus den Sequenzierungseinheiten erfolgt zu großen Teilen vollautomatisch. Hier soll nicht ausführlich auf die den Programmen zugrunde liegende Theorie

eingegangen werden, sondern nur ein allgemeiner Überblick über die Abläufe gegeben werden. Die Verarbeitung erfolgte in folgenden Schritten, nachdem die Sequenzdateien die Sequenzierungseinheiten verlassen haben:

- base calling mit dem Computerprogramm phred (Ewing and Green, 1998; Ewing et al., 1998): Zuweisen von Qualitätsparametern zu jeder Base (Fehlerwahrscheinlichkeit).
- *vector-screening*, *E. coli*-Screening und Umwandlung der Sequenzdateien mit dem Computerprogramm pregap (Bonfield et al., 1995): Maskieren von Sequenzen, die dem Sequenzierungsvektor zuzuordnen sind.
- assembly, d. h. Zusammensetzen der einzelnen reads zu Contigs durch Finden von überlappenden Sequenzen mit dem Computerprogramm gephrap (Bonfield et al., 1995).
- Erstellen einer Datenbank zur Bearbeitung der Sequenzen mit Hilfe des Sequenzeditor Programms GAP4 (Bonfield et al., 1995).

## 2.11. Editierung der Sequenzen

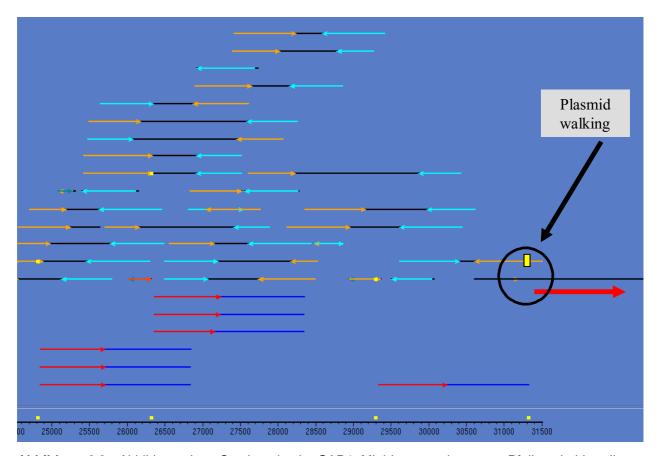
Nach dem Ende der Rohsequenzierungsphase des *B. licheniformis* Genomsequenzierungsprojekts bestand die Sequenz aus 318 Contigs. Dabei handelt es sich um zusammenhängende Sequenzabschnitte, die durch das computergestützte Zusammenlagern (*assembly*) vieler kleiner Sequenzstücke aus der Ansequenzierung von Plasmiden entstanden sind. Zwischen diesen Contigs befinden sich die Lücken, die es für eine vollständige Genomsequenz zu schließen gilt. Des Weiteren sind einzelne Bereiche der bereits bestehenden Sequenz unzureichend abgedeckt, d.h. die Qualität der Sequenzläufe ist ungenügend. In diesem Abschnitt soll vorgestellt werden, wie mit Hilfe der so genannten Editierung ein Überarbeiten bereits sequenzierter Bereiche vorgenommen wurde und basierend auf dem Potential der Plasmidgenbank erste Lücken geschlossen werden konnten.

# 2.11.1. "Plasmid walking"

Bei einer durchschnittlichen Insertgröße von 2 - 3 kb in den Plasmidgenbanken von Genomprojekten kann mit der Ansequenzierung der Inserts von beiden Seiten nicht die komplette Insertsequenz erfasst werden. Auf diese Weise können Contigenden entstehen, da nicht die vollständige Sequenzinformation eines oder mehrerer Inserts vorliegt.

Beim *Plasmid walking* wurde an den Contigenden jeweils ein Primer abgeleitet und mit diesem Primer auf dem jeweiligen Plasmid in die Lücke hineinsequenziert. Im Programm GAP4 des STADEN Programmpaketes (Staden et al., 2000) wird ein solcher Fall wie in Abbildung 2.3. dargestellt. Dabei erreicht man die jeweils schon bekannte Sequenz am anderen Ende der

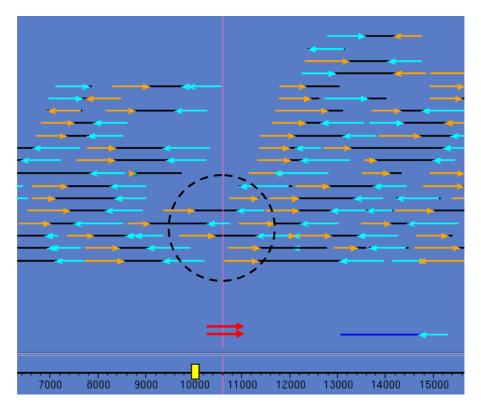
Lücke. Diese Methode ist erschöpft, wenn keine Lücken-überspannenden Plasmide mehr vorhanden sind. Eine weitere Anwendung findet das *Plasmid walking* bei der Qualitätsverbesserung der Konsensussequenz. Abbildung 2.4. zeigt einen sehr schlecht abgedeckten Konsensussereich im Contig Editor des Programms GAP4. Je dunkler die Graustufen der Konsensussequenz dargestellt sind, umso niedriger ist der so genannte *quality value* (QV) (2.13.), welcher, berechnet von dem Programm PHRED (Ewing and Green, 1998; Ewing et al., 1998), die Verlässlichkeit einer sequenzierten Base angibt. Die Qualität der Konsensussequenz ist dabei abhängig von der Qualität der Sequenzläufe. Auf den diesen Bereich überspannenden Plasmidinserts wurde ein spezifischer Primer generiert, die entsprechende Region neu sequenziert und somit die Konsensussequenz qualitativ aufgewertet. Standardmäßig wurde eine Abdeckung des Konsensus durch mindestens zwei unabhängige gute Sequenzen angestrebt.



**Abbildung 2.3.:** Abbildung eines Contigendes im GAP4. Mit blauen und orangen Pfeilen sind jeweils die Sequenzläufe auf einem Insert der Plasmidgenbank dargestellt. Die schwarze Linie, welche die Sequenzläufe verbindet, stellt den auf diesem Insert nicht sequenzierten Bereich dar. Am rechten Ende des Contigs ist als gelbes Kästchen der generierte Primer zu sehen, mit welchem auf dem dargestellten Insert aus dem Contig heraussequenziert werden kann. Die neue Sequenz wird durch den dicken roten Pfeil gezeigt.

Contig Editor: -15224 ablap40w2									×
< C: 1 > < Q: 0 >	☐ Insert Edit Modes >> ☐	Cutoffs Undo	Next Search Co	mmands >> Se	ttings >>				Quit Help >>
<< < > >>	4	j .							D
	0570 10580	10590	10600	10610	10620	10630	10640	10650	10660
-15331 bbldq18f	GCACATCCT								
-15332 gbldq18f	GCACATCCTCATCCCAC								
-15333 gblhh68f	GCACATCCTCATCCCAC	AATTCGGGA*T	<b>FAAAAGCAAGC</b>	G*CCATGGT	ACAGCTTTCC	GTTCATCTT	CATACCGTCTC	TTTGACAAA	CGGCAGCCC
15334 gblb165r								TTTGACAAA	
CONSENSUS	GCACATCCTCATCCCAC	AATTCGGGA*T	<b>FAAAAGCAAGC</b>	G*CCATGGT	ACAGCTTTCC	GTTCATCTT	CATACCGTCTC	TTTGACAAA	CGGCAGCCC
Reading:gbldq18f(#15332) Length:73	37(814) Vector:unknown Clo	ne:unknown Che	mistry:unknown	primer Primer:	forward univers	al .			

**Abbildung 2.4.:** Abbildung einer schlecht abgedeckten Konsensussequenz (unten) und von 4 Sequenzläufen, dargestellt im *Contig editor* des Programms GAP4.



**Abbildung 2.5.:** Darstellung der dem Sequenzbereich aus Abbildung 2.4. entsprechenden Region im *Template display* des Programms GAP4. Das gelbe Kästchen stellt einen in dieser Region abgeleiteten Primer dar. Neue Sequenzläufe, auf den im eingekreisten Bereich liegenden Plasmidinserts, sind als rote Pfeile dargestellt.

# 2.12. Lückenschluss (Gap closure)

Bei der Sequenzierung einer Genbank, in der das Genom statistisch verteilt kloniert vorliegt, nimmt bei steigender Anzahl von Sequenzläufen die Anzahl der Contigs ab (Lander and Waterman, 1988). Wenn die Anzahl der Contigs trotz zunehmender Anzahl von Sequenzläufen nicht mehr deutlich abnimmt, wird im Normalfall die Sequenzierstrategie auf Lückenschluss mit Hilfe unterschiedlicher PCR Verfahren umgestellt.

# 2.12.1. Lückenschluß über den Vergleich der automatisch annotierten Sequenz von *B. licheniformis* mit dem Genom von *B. subtilis*

Bereits bei einem ersten Vergleich des mit ERGO (Overbeek et al., 2003) vorläufig annotierten *B. licheniformis* Genoms mit dem vollständig sequenziert vorliegendem *B. subtilis* Genom (Kunst et al., 1997) zeigten sich deutliche Homolgien der jeweiligen ORFs und ihrer Anordnung in Clustern. Daher bot es sich an, diese Homologien für den Lückenschluß auszunutzen. Die Strategie entspricht einer Projektion der Contigs aus *B. licheniformis* auf das Genom von *B. subtilis*. Wobei die Verwendung der codierten Proteine die Sensitivität gegenüber einer Projektion auf DNA-Ebene erhöht, da *silent mutations* zwischen den beiden verschiedenen Stämmen weggefiltert werden. Es wurde folgendermaßen vorgegangen:

Die Contigenden der Contigs größer 3 kb, der bis dahin vorliegenden *B. licheniformis* Sequenz, wurden über ERGO mit dem Genom von *B. subtilis* verglichen. Konnte ein homologes ORF Cluster gefunden werden (Abbildung 2.6.A), erfolgte wiederum ein Abgleich der im *B. subtilis* Genom folgenden ORFs gegen die Sequenz von *B. licheniformis*. Im Idealfall konnte an einem anderen Ende eines *B. licheniformis* Contigs die Fortsetzung des ORF Clusters, entsprechend zu *B. subtilis*, gefunden werden (Abbildung 2.6.B).



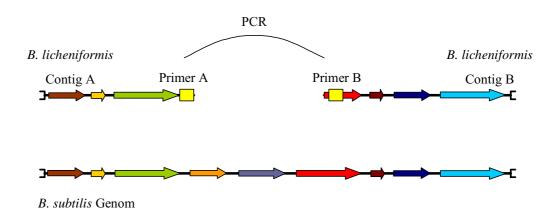
Abbildung 2.6.: Darstellung homologer ORF Cluster von B. licheniformis und B. subtilis in ERGO.

A - Ende eines Contigs X der *B. licheniformis* Sequenz mit einem ORF Cluster homolog zum Genom von *B. subtilis*.

B j- Ende eine Contigs Y der *B. licheniformis* Sequenz mit der Fortsetzung des ORF Clusters auf dem *B. subtilis* Genom.

In einem solchen Falle wurden an den Enden der beiden B. licheniformis Contigs Primer generiert und je nach der zu erwartenden Länge des Produktes eine Standard- oder Long-range

PCR durchgeführt (Abbildung 2.7.). Das entstandene PCR Produkt wurde aufgereinigt, von beiden Seiten ansequenziert und die Sequenzen in den Datensatz assembliert. Konnte mit diesen Sequenzen die Lücke, aufgrund einer Größe von mehr als ca.1 kb noch nicht geschlossen werden, wurde auf dem PCR Produkt analog zum *Plasmid walking*, mit neuen Primern entlangsequenziert.



**Abbildung 2.7.:** Schematische Darstellung zweier Contigenden von *B. licheniformis* und eines Abschnittes des Genoms von *B. subtilis* mit homologen ORF Clustern. An den Contigenden wurden Primer generiert (gelbe Kästchen) mit deren Hilfe ein lückenüberspannendes PCR-Produkt amplifiziert wurde.

# 2.12.2. Lückenschluß mit Hilfe der Vectoretten Technik (Kilstrup and Kristiansen, 2000)

Bei der Vectoretten Technik handelt es sich um ein PCR Verfahren, welches es ermöglicht, mit nur einem spezifischen Primer und einem Standardprimer ein spezifisches DNA Fragment zu amplifizieren (Abbildung 2.8.). Der genomspezifische Primer wurde am Ende eines Contigs gewählt, so dass mit dem entstandenen PCR Produkt in die Sequenzlücke hineinsequenziert werden konnte. Benötigt wird für die Vectoretten Technik verdaute chromosomale DNA, z.B. mit *Eco*RI, und eine dephosphorylierte, definierte Oligokasette (MKP) mit entsprechenden Enden (*Eco*RI). Die Fragmente der chromosomalen DNA und die Oligokassetten werden ligiert und das Ligationsprodukt in einer PCR als Template eingesetzt. Während des ersten Denaturierungsschrittes der PCR werden die DNA Doppelstränge aufgeschmolzen, wobei der an der *Eco*RI Schnittstelle dephosphorylierte 5' Einzelstrang der Oligokassette abfällt. Somit entstehen DNA Einzelstränge mit einem definierten Ende. Im folgenden Annealing und Elongationsschritt bindet der genomspezifische Primer an den DNA Einzelstrang und synthetisiert den komplementären Strang. In den folgenden Amplifikationszyklen können nun sowohl der genomspezifische Primer, als auch der Standardprimer (MKP24), welcher spezifisch für

das ligierte Fragment der Oligokassette ist, an die denaturierte DNA binden und das entsprechende Produkt amplifizieren.

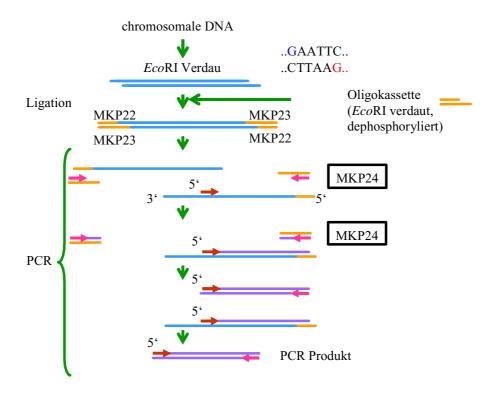


Abbildung 2.8.: Schematische Darstellung der Vectoretten Technik. Erläuterung im Text.

Blaue Linien: EcoRI verdaute chromosomale DNA

Orangefarbene Linien: EcoRI verdaute, dephosphorylierte Oligokassette (MKP)

Rote Pfeile: Genomspezifischer Primer

Pinkfarbene Pfeile: an die Oligokassette bindender Standardprimer (MKP24)

#### Durchführung:

Für die Durchführung der Vectoretten Technik wurde das Universal Vectorett System UVS1 der Firma SIGMA, Aldrich verwendet.

Die Oligokassetten mit *Eco*RI, *Hind* III bzw. *Bam*H1-Ende wurden nach Anleitung des Herstellers hergestellt. Die Kassetten konnte bis zu 2 Wochen bei 4°C gelagert werden. Der Verdau der chromosomalen DNA mit dem entsprechenden Restriktionsenzym erfolgte wie unter 2.6.1. beschrieben. Für die Ligation (2.6.5.) der verdauten DNA mit der zughörigen Oligokassette, wurden die Konzentrationen mittels Gelelektrophorese abgeschätzt und im Verhältnis 1:1 eingesetzt. Anschließend erfolgte eine Aufreinigung des Ligationsansatzes analog der Aufreinigung von PCR Produkten (2.4.4.). Für die PCR wurde die Takara Ex Taq Polymerase (Takara Biomedicals, Shuzo Co. Ltd.) eingesetzt.

30 µl Vectoretten PCR Ansatz:			Vector	etten PCR Prog	ramm:	
0,3	μl	Vektoretten-Ligation	01	LIDHT ON 1		110°C
5	μl	Primer spezifisch (5 μM )	02	PAUSE FOREVER 9		96°C
3	μl	Standardprimer (5µM)	03	TEMP	3 min	95°C
0,5	μl	dNTP-Mix Takara ( je 2,5 mM )	04	LOOP[	30x	
3	μl	10x ExTaq Puffer	05	TEMP	20 s	95°C
0,4	μl	Takara Ex Taq	06	TEMP	30 s	63°C
ad 30	μl	$H_2O_{HPLC}$	07	TEMP	2 min	72°C
			80	LOOP]		
			09	TEMP	10 min	72°C
			10	LIDHT OFF		
			11	TEMP	FOREVER	8°C

Die PCR Produkte wurden anschließend gelelektrophoretisch überprüft und aufgereinigt. Zur Sequenzierung wurden einem 20  $\mu$ l Sequenzieransatz (2.9.) 1  $\mu$ l DMSO und 1  $\mu$ l Betain zugesetzt. Sequenziert wurde sowohl mit dem genomspezifischen Primer, als auch mit dem Standardprimer.

### 2.12.3. Lückenschluß mittels kombinatorischer Multiplex PCR

Eine weitere Methode, die sehr effektiv zum Lückenschluß eingesetzt werden konnte, ist die kombinatorische Multiplex PCR. In diesem Stadium des Sequenzierungsprojektes sollten 200 Primer, lokalisiert an sämtlichen Enden der Contigs, miteinander kombiniert werden. Eine Überprüfung dieser 200 Primer in einzelnen PCR Reaktionen, hätte insgesamt 19900 PCR Ansätze zur Folge gehabt (200\*199/2).

In dieser Arbeit wurden die Primer nach zwei unterschiedlichen Kombinationsverfahren gepoolt, zum einen nach einem Pooling Algorithmus, entwickelt von (Tettelin et al., 1999) und im zweiten Lauf mit Hilfe des Kombinationsdreiecks.

### 2.12.3.1. Kombination der Primer nach Tettelin et al., 1999 (modifiziert)

Im ersten Lauf der kombinatorischen Multiplex PCR wurden die 200 miteinander zu kombinierenden Primer sehr aufwendig nach einem Pooling Algorithmus (Tettelin et al., 1999), basierend auf der so genannten 'affine planes techniqe' (Hall, 1996) in 210 Pools zu je 14 Primern zusammengefasst, welche sich aus 3 Vorpools zu 5/5/4 Primern zusammensetzten. Jeder Primer war in 7 Pools vertreten und wurde mit jeweils 13 anderen Primern kombiniert. Auf diese Weise sollte sichergestellt werden, dass, wenn mit jedem dieser Pools eine Multiplex PCR durchgeführt wird, jeder Primer mit jedem anderen Primer mindestens einmal in Kom-

bination vorliegt und alle möglichen PCR-Produkte amplifiziert werden. Die Primer mit einer Länge von 17 - 18 bp und einer Schmelztemperatur von 55 - 60°C wurden mit Hilfe des Programms GAP4 (Staden et al., 2000) generiert und durch BLAST gegen die vorliegende *B. licheniformis* Sequenz auf Spezifität mit mindestens 5 abweichenden Basen von einer homologen Sequenz überprüft.

Die Endkonzentration sämtlicher Primer im PCR Ansatz betrug 5  $\mu$ M. Die Amplifikation erfolgte in 20  $\mu$ l Ansätzen mittels *Long-range* PCR (2.7.2.). Die Auswertung der Multiplex PCR erfolgte mittels Agarosegealaktrophorese. Konnten in einem PCR Ansatz ein oder mehrere Produkte amplifiziert werden, wurden die entsprechenden Vorpools einzeln und in Kombination in 20  $\mu$ l PCR Ansätzen daraufhin untersucht, welche zwei Primer für die Amplifikation nötig waren. Nach Identifikation des Primerpaares erfolgte eine PCR im 100  $\mu$ l Maßstab, welche aufgereinigt (2.7.2.) und anschließend sequenziert (2.9.) wurde.

#### 2.12.3.2. Kombination der Primer mit Hilfe des Kombinationsdreiecks

Nach der ersten Runde kombinatorischer Multiplex PCR modifiziert nach Tettelin wurden für die verbleibenden Contigenden neue Primer mit einer Länge von 27 - 30 bp und einer Schmelztemperatur von 60 - 65°C generiert, um die Spezifität der PCR Reaktion auf chromosomaler DNA zu erhöhen. Für den zweiten Ansatz wurden die Primer in Fünfergruppen gepoolt und diese nach dem in Tabelle 2.5. dargestellten Schema kombiniert. Die *Long-range* PCRs wurden mit insgesamt 10 Primern in 20 μl Ansätzen durchgeführt und gelektrophoretisch auf Produkte überprüft. Bei positiven Reaktionen wurde durch das Weglassen von jeweils einem Primer das entsprechende Primerpaar isoliert. Anschließend wurden die Produkte in 100 μl Ansätzen amplifiziert (2.7.2.), aufgereinigt (2.4.5.) und sequenziert (2.9.).

Κ Α В С Ε F Н Ī L М N Р Q R D G 0 Α AΒ AC AD ΑE ΑF AG ΑH ΑI ΑK AL ΑM ΑN ΑO AΡ AQ AR В BC BD ΒE BG ВΙ BK BL BMВО BQ BF BH ΒN ΒP BR C CR CD CE CF CG CH CI CK CL CM CN CO CP CQ D DE DF DG DΙ DK DL DM DN DO DP DQ DR DΗ Ε EF EO ER EG EΗ ΕI ΕK EL ΕM ΕN EΡ EQ F FG FΗ FΙ FL FN FO FQ FR FΚ FΜ FΡ G GH GΙ GK GL GM GN GO GP GQ GR Н ΗΙ HK HL HM HNHO HP HQ HR ı ΙK IL IM IN Ю IΡ IQ IR K KL ΚM ΚN KO ΚP KQ KR L LM LN LO LQ LR LP M MN MO MP PQ MR Ν NO NP NQ NR 0 OP OQ OR Ρ PQ PR Q QR R

**Tabelle 2.5.:** Kombinationsschema der Primerpools mit jeweils 5 Primern. Ein Buchstabe steht für je einen Pool.

## 2.13. Qualitätsverbesserung der Sequenz (Polishing)

Die Qualität, d. h. die Verlässlichkeit, einer sequenzierten Base in einem Sequenzlauf wird durch das Programm PHRED (Ewing and Green, 1998; Ewing et al., 1998) berechnet und als so genannter *quality value* (QV) im Editierprogramm GAP4 (Bonfield et al., 1995) als Graustufe in der Hintergrundfarbe dargestellt. Die Qualität einer Base in der Konsensussequenz ergibt sich aus der Mittelung der Qualität dieser Base in den einzelnen Sequenzläufen. Der *quality value* wird wie folgt berechnet:

$$QV = -10 \cdot \log 10(P)$$

Woraus sich für einen QV von 40 ergibt, dass die Wahrscheinlichkeit P für einen Fehler pro Base kleiner gleich 1/10.000 ist, d.h. bei 10.000 sequenzierten Basen ist im Schnitt mit einem Fehler zu rechnen. Bereits während der Phase des Lückenschlusses wurde die bestehende Sequenz einem so genannten *Polishing* unterzogen. Dabei wurde im *Contig editor* nach Bereichen gesucht, die in der Konsensussequenz einen QV kleiner 40 aufwiesen. Hier wurde dann mit zusätzlichen Sequenzen auf der Basis von PCRs oder *Plasmid walkings* eine durchschnittlich höhere Qualität erreicht und somit der *quality value* verbessert.

### 2.14. Annotation und Sequenzanalyse

Seit nach der Rohsequenzierung ein Großteil der *B. licheniformis* Sequenz bekannt war, wurde diese Sequenz dem Annotationsprozess unterzogen. Im Verlauf des Sequenzierungsprojektes bis hin zum Schließen des Genoms wurde dieser Prozess regelmäßigen Updates unterzogen.

Unter Annotation versteht man das Finden aller durch die Nukleotidsequenz kodierten Gene. Soweit möglich soll den abgeleiteten Proteinen dabei eine Funktion zugewiesen werden. Das Zuweisen der Funktion beruht auf Sequenzähnlichkeit und folgt dem Paradigma, dass ab einer Ähnlichkeit von 30% mit hoher statistischer Sicherheit auf übereinstimmende räumliche Strukturen für Proteine geschlossen werden kann (Sander and Schneider, 1991).

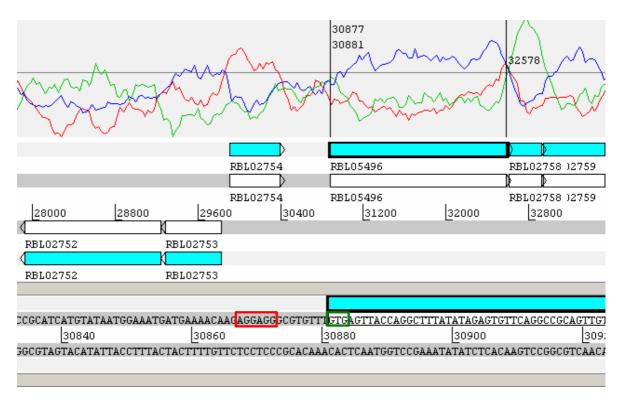
Für die Annotation im Genomprojekt *B. licheniformis* wurde nach der ORF Vorhersage das Programm ERGO (Overbeek et al., 2003) sowohl für eine automatische, als auch für eine detaillierte manuelle Annotation verwendet.

### 2.14.1. ORF Vorhersage

Bevor mit der Annotation begonnen werden konnte, mussten zuerst alle potentiellen offenen Leserahmen (ORFs, open reading frames) identifiziert werden. Diese Aufgabe wurde durch ORF-finder bewerkstelligt, Programme, die Sequenzen im einfachsten Fall nach Stop-Codons (TAA, TGA, TAG) durchsuchen und damit zuerst das Ende eines ORFs festlegen. Im nächsten Schritt wird dann im selben Leserahmen in 5'-Richtung nach einem Start-Codon gesucht. Erreicht dieser ORF eine vorher festgelegte Mindestlänge, werden die Koordination (Start und Stop) des Gens für die weitere Verarbeitung gespeichert. Für diese bioinformatische Aufgabenstellung gibt es mehrere Programme, die alle bestimmte Vor- und Nachteile mit sich bringen. Für eine optimale ORF-Vorhersage ist eine Kombination mehrere Programme und Methoden vorteilhaft. Das Programm YACOP (Tech and Merkl, 2003) erfüllt als kombinierter ORF-finder diesen Anspruch und wurde im vorliegenden Projekt verwendet. Aus der Vielzahl der so gefundenen, sich zum Großteil überschneidenden ORFs, filtern die ORFfinder solche heraus, die mit einer gewissen Wahrscheinlichkeit Gene sein können. Die verwendeten ORF-finder untersuchen dabei die interne Zusammensetzung der Sequenz in einem ORF. Dabei wird ausgenutzt, dass die Organismen nicht alle Codons des genetischen Codes verwenden, sondern nur einen Spezies spezifischen Subset.

Mit Hilfe des Programms Artemis (http://www.sanger.ac.uk/Software/Artemis/ konnten die durch YACOP gefundenen potentiellen ORFs auf Ebene der Nukleotidsequenz überprüft werden. Abbildung 2.9. zeigt die Darstellung potentieller ORFs in Artemis. Kriterien für die

Überprüfung der ORFs waren zum einen das Vorhandensein einer Ribosomenbindestelle (AAGGAGG) (Sonenshein et al., 1993) und zum anderen eine einheitliche *codon usage* vom Start- bis zum Stoppcodon des ORFs.



**Abbildung 2.9.:** Darstellung eines Genomabschnittes im Programm Artemis. Blaue und weiße Pfeile zeigen die einzelnen ORFs zu einen auf dem *leading strand* und zum anderen auf dem *lagging strand*. Die rot/grün/blaue Grafik stellt den *Corelation score* (*codon usage*), d.h. die G+C Verteilung an erster, zweiter und dritter Position in allen Codons im Bereich des mit dickem schwarzem Rahmen markierten ORFs dar. Im unteren Bereich der Abbildung ist der Beginn der Nukleotidsequenz des markierten ORFs zu sehen. Am Anfang des ORFs sind die Ribosomenbindestelle (rot) und das Startcodon (grün) gekennzeichnet.

#### 2.14.2. Automatische Annotation - ERGO (Overbeek et al., 2003)

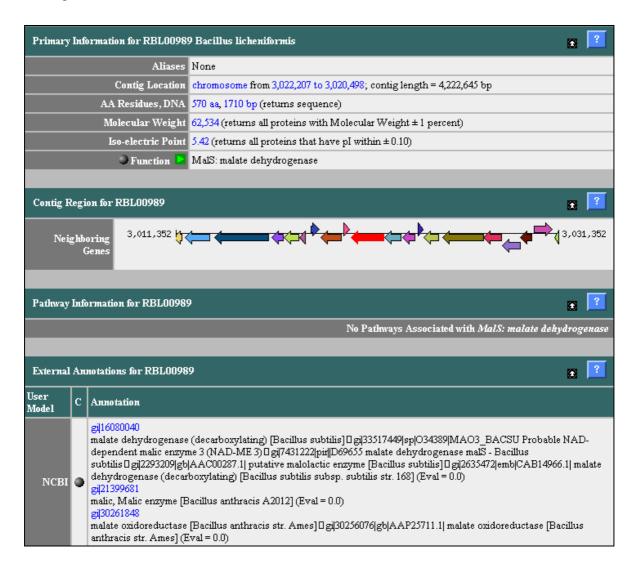
Für eine abschließende Annotation des *B. licheniformis* Proteoms wurde das ERGO Programmpaket verwendet. Nach einer automatisch durchgeführten Annotation wurde diese für jedes Protein zusätzlich manuell validiert. Die automatische Annotation durch ERGO beinhaltet die Übersetzung der, von YACOP in Form von Start- und Stoppositionen vorhergesagten, Gene in Proteinsequenzen und deren Abgleich mittels FASTA (Pearson, 1994; Pearson and Lipman, 1988) gegen eine nichtredundante Proteindatenbank bestehend aus:

- Swiss-Prot (uniprot\_sprot.\*.gz)
- TrEMBL (uniprot\_trembl.\*.gz)
- TrEMBL new (new/uniprot trembl new.\*.gz)

#### Die Dateien befinden sich unter:

ftp.expasy.org/databases/uniprot/knowledgebase/.

Um die manuelle Annotation zu erleichtern, wurde außerdem eine BLAST-Suche gegen die jeweils aktuelle Proteindatenbank (ftp.ncbi.nih.gov/blast/db/nr.tar.gz) des NCBI durchgeführt und das Ergebnis auf der jeweiligen Protein-Seite in ERGO abgebildet (s.Abbildung 2.10.). Die Suche nach tRNAs wird von dem Computerprogramm tRNAscan (Lowe and Eddy, 1997) durchgeführt.



**Abbildung 2.10.:** Screenshot aus dem Programm ERGO. Das oberste Fenster gibt die Position im Genom, die Größe und die Annotation für den aktivierten ORF (rot) wieder. Im mittleren Bereich ist der Abschnitt des Genoms von *B. licheniformis* mit den darauf angeordneten ORFs zu sehen. Als Beispiel ist hier für den ORF der Malat-Dehydrogenase im unteren Fenster das NCBI-BLAST-Ergebnis dargestellt.

# 2.14.3. Funktionelle Annotation durch Vergleich lokaler Sequenzhomologien (BLAST)

Die manuelle Annotation wurde durch funktionale Sequenzanalysen abgesichert. So wurde zur Vorhersage membranspannender Regionen das Programm TMPred (Bateman et al., 2004) (www.ch.embnet.org/software/TMPRED\_form.html) genutzt.

Des Weiteren wurde die automatische Annotation jedes ORFs über BLAST (Altschul et al., 1990) gegen Swissprot, GeneBank/European Molecular Biology Laboratory databases (Benson et al., 2004; Kulikova et al., 2004) überprüft und verifiziert.

Das Programme Pfam wurde zur Einordnung von Proteinen in Proteinfamilien (Bateman et al., 2004) (<a href="www.sanger.ac.uk/Software/Pfam/">www.sanger.ac.uk/Software/Pfam/</a>) genutzt, COG zur Einordnung von Proteinen in Cluster orthologer Gruppen (Tatusov et al., 1997) (<a href="www.ncbi.nlm.nih.gov/COG/">www.ncbi.nlm.nih.gov/COG/</a>) und Prosite (Hulo et al., 2004) zur Überprüfung funktioneller Motive in Übereinstimmung mit der zugewiesenen Funktion.

### 2.14.4. Funktionelle Annotation über Homologie zu Bacillus subtilis

B. subtilis ist ein gut abgesicherter Modelorganismus, der mit B. licheniformis sehr nahe verwandt ist. Daher wurde eine Suche aller homologen Gene beider Spezies durchgeführt, basierend auf der Idee, dass so eine Funktionszuweisung der homologen Gene in B. licheniformis möglich ist. Alle annotierten ORFs von B. licheniformis wurden mittels protein/protein BLAST mit der B. subtilis Datenbank SubtiList (http://genolist.pasteur.fr/SubtiList/) verglichen. Unter Einsatz einer speziell für diesen Zweck entwickelten Software (Ehrenreich, unveröffentlicht) wurden die BLAST Alignments jedes einzelnen ORFs überprüft. Hierbei wurde besonderer Wert darauf gelegt, dass die Alignments jeweils über die gesamte Länge der ORFs vorlagen.

Bei einem *e-value* kleiner oder gleich 10<sup>-15</sup> und einem Alignment über die gesamte Länge wurden ORFs als ortholog eingestuft und die Annotation des orthologen *B. subtilis* Proteins übernommen.

# 2.14.5. Komparative Genomanalyse von *B. licheniformis*, *B. subtilis* und *B. halodurans*

Die komparative Genomanalyse von *B. licheniformis*, *B. subtilis* und *B. halodurans* erfolgte mittels BLAST, indem jeder einzelne ORF der drei Organismen jeweils mit Datenbanken der beiden anderen Bacilli verglichen wurde. ORFs mit einem *e-value* kleiner oder gleich 10<sup>-15</sup> wurden als ortholog eingestuft.

# 2.14.6. Sequenzanalyse basierend auf der *codon usage* – SIGI (Merkl, 2004)

Mit Hilfe des Programms SIGI kann die *codon usage* eines jeden ORFs mit der durchschnittlichen *codon usage* des gesamten Genoms verglichen werden. Das Programm SIGI sortiert die Gene in die drei Kategorien "normal exprimiert", "stark exprimiert" und "Fremdgene" (*alien genes*). Gene, welche eine atypische *codon usage* aufweisen, werden anschließend durch Stammbaum und Gendatenbankanalysen daraufhin unterschieden, ob es sich um bekannte stark exprimierte Gene, oder um so genannte *alien genes* handelt. Die genomische Umgebung solcher potentieller *alien genes*, d.h. durch horizontalen Gentransfer ins Genom gelangte Gene, wird ebenfalls überprüft, da davon ausgegangen werden kann, dass keine einzelnen Gene, sondern eher Gencluster (*genomic islands*) übertragen werden. Die Übertragung ganzer Gencluster konnte sowohl für pathogene (Hacker and Kaper, 2000), als auch für nicht pathogene (Kaneko et al., 2000) Bakterien gezeigt werden. Auf diese Weise kann die Anzahl falsch positiver *alien genes* deutlich herab gesetzt werden. Eine Visualisierung der *alien genes* erfolgte mit dem Programm Artemis.

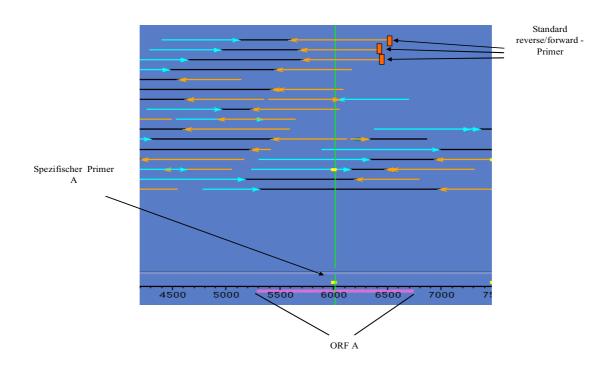
## 2.15. Erstellen des Genomchips

Für die funktionelle Analyse des Genoms von *B. licheniformis* wurde ein DNA-Microarray erstellt, auf dem nach Möglichkeit jeder ORF des Genoms von *B. licheniformis* vertreten ist. Hierfür gibt es zum einen das sehr teure Verfahren einen Abschnitt jedes ORFs als Oligonukleotid zu synthetisieren und einen so genannten Oligochip zu erstellen. Eine weitere Möglichkeit, welche in dieser Arbeit Anwendung fand, besteht darin, ca. 300 - 500 bp große, spezifische Fragmente der ORFs mittels PCR zu amplifizieren. Im Folgenden wird beschrieben, nach welchem Verfahren Primer innerhalb der ORFs abgeleitet und die Fragmente amplifiziert und auf den Microarray gespottet wurden.

## 2.15.1. Primergenerierung

Im Rahmen dieses Projektes sollten für alle ORFs, mit einer Mindestlänge von 300 bp, Primerpaare generiert werden, wobei jeder Primer eine durchschnittliche Länge von 17-18 bp und eine Schmelztemperatur von 60°C besitzen sollte. Um diesen Vorgang für die mehr als 4000 ORFs zu automatisieren, wurde ein Programm genutzt (Armin Ehrenreich, unveröffentlicht), welches unter vorgegebenen Parametern passende Primer für die Amplifikation der ca. 300 bp großen ORF Fragmente (*probes*) generierte.

Bei einem Großteil der Reaktionen konnte mit nur einem spezifischen und einem Standardprimer gearbeitet werden (siehe Ergebnisse 3.7.1.), hier wurde keine chromosomale DNA als
Template eingesetzt, sondern Plasmide der Genbank, welche zur Sequenzierung des Genoms
angefertigt wurde. In diesen Fällen wurden von dem Programm zur Primergenerierung Inserts
der Plasmide gesucht, deren Ende oder Anfang innerhalb eines ORFs lokalisiert waren (Abbildung 2.11.). Auf diese Weise konnte der Standard Forward oder Reverse Primer, der schon
zur Rohsequenzierung genutzt wurde, für eine deutlich kostengünstigere Amplifikation der
probes eingesetzt werden. ORFs, in deren Sequenzbereich kein Insertanfang oder -ende gefunden werden konnte, wurden mit zwei spezifischen Primern amplifiziert. Wenn möglich
wurde die DNA eines den 300 bp Bereich überspannden Inserts der Plasmidgenbank als
Template eingesetzt, um die Spezifität der PCR Reaktion, im Vergleich zum Einsatz chromosomaler DNA, zu erhöhen. ORFs in deren Bereich keinerlei in Frage kommendes Inserts lag,
z.B. in Regionen von Lücken, welche über PCR geschlossen wurden, kamen zwei spezifische
Primer auf chromosomaler DNA zum Einsatz. Die Synthese der Primer wurde bei der Firma
QIAGEN Operon, Köln, in Auftrag gegeben.



**Abbildung 2.11.:** Schematische Darstellung der Primerwahl für einen ORF A, dargestellt im *Template display* des Programms GAP4. Im Bereich des ORFs A liegen mehrere Inserts der Plasmidgenbank, deren forward und reverse Läufe je durch einen orangefarbenen und einen grünen Pfeil, verbunden durch eine schwarze Linie, dargestellt sind. Drei Inserts wurden als Template für die Amplifikation der *probes* ausgewählt, wobei der entsprechende Standardprimer (orangefarbene Kästchen) eingesetzt werden konnte. Ein spezifischer Primer A wurde in einem Abstand von ca. 300 bp gewählt (gelbes Kästchen).

#### 2.15.2. Amplifizierung der ,probes'

Die Amplifizierung der *probes* erfolgte im 96 well Maßstab. Die Primer A und B wurden je  $20~\mu\text{M}$  und die Plasmid DNA in 1/10~Verdünnungen der Sequenziergenbank in 96 well Platten vorgelegt und bei  $20^{\circ}\text{C}$  gelagert. Die PCR Reaktion erfolgte ebenfalls im 96 well Maßstab, so dass mit einer Achtkanalpipette lediglich Primer A, Primer B und die Template DNA zusammengeführt und 90  $\mu$ l Mastermix hinzugegeben werden mussten.

Mastermix für 96x 100 µl Chip PCR Ansätze:				Chip PCR Ansatz:			
200	μl	dNTP's (10 mM jeweils)	5	μl	Primer A (20 µM)		
1000	μl	10x PCR-Puffer	5	μl	Primer B (20 μM)		
2000	μl	5x Q-Solution	2	μl	Template DNA		
70	μl	Taq (3,5 U/100 µl PCR-Ansatz)	90	μl	Mastermix		
ad 9000µl		H <sub>2</sub> O <sub>bidest</sub>					

Die Amplifikation erfolgte mittels Standard PCR (2.7.1.), wobei die Elongationszeit auf 45 s herabgesetzt wurde. Im Anschluss an die PCR wurden die Produkte gelektrophoretisch überprüft (1,7%ig Agarose, 90 min 90 V). Die Aufreinigung erfolgte, wie unter 2.4.5. beschrieben. Für die Elution der *probes* wurden 65 μl 70°C warmes H<sub>2</sub>O<sub>bidest/steril</sub> eingesetzt.

### 2.15.3. Spotten und UV-Immobilisierung

Das Spotten der *probes* erfolgte mit dem Lucidea Spotter der Firma Amersham Biosiences nach dem Prinzip des , *contact printings*'. Ein Druckkopf mit 24 Nadeln, jede geformt wie ein winziger Füllfederhalter (Abbildung 2.12.), nimmt aus Mikrotiterplatten pro Nadel ca. 200 nl *probes* auf, um dann mit hoher Präzision 100 pl Flüssigkeit pro Spot auf die Chipoberfläche zu setzten. Auf diese Weise ist es möglich 20000 einzelne Spots auf ein Microarray mit den Maßen eines Objektträgers (2,5 cm x 7,5 cm) zu setzten, wobei jeder Spot einen Durchmesser von 150 µm aufweist und der minimale Abstand der Spots 30 µm beträgt. In dieser Arbeit wurden Aminosilan *slides* der Firma Amersham Biosiences verwendet, welche durch die Gasbehandlung von Glasslides mit 3-aminopropyltrimethoxysilan hergestellt werden. Diese Modifikation des *slides* ermöglicht das Binden von DNA durch Interaktion zwischen dem negativ geladenen Rückgrat der Nukleinsäure und der positiv geladenen Aminosilanoberfläche.



Abbildung 2.12.: Nadel des Druckkopfes zum spotten von probes auf Microarrays

Im Vorfeld wurden die *probes* aus jeweils vier 96 well Platten auf eine 384 well Platte (Firma) überführt und 1:1 mit DMSO zur Denaturierung versetzt. Das Spotten der *probes* erfolgte in Blöcken von 16 x 11 Spots, wobei jeder Block doppelt auf den Chip gesetzt wurde. Das heißt, in jeder Blockreihe, bestehend aus 4 Blöcken, sind jeweils die ersten beiden und die letzte beiden Blöcke identisch. In den ersten acht Blockreihen wurde jeweils eine so genannte *scorecard* mitgeführt. Hierbei handelt es sich um spezifische DNA (Lucedia TM Universal ScoreCard AMERSHAM Biosciences, Freiburg), welche später zur Normalisierung der Microarraydaten dient.

Nach dem Spotten der *probes* erfolgte eine so genannte UV-Immobilisierung mit dem UVC500 UV Crosslinker (HOEFER, San Francisco). Durch das Bestrahlen der gespotteten Chips mit Licht einer Energie von 500 μJ/cm² für 5 min wurde die Bindung der DNA Fragmente an die Chipoberfläche intensiviert.

Die geprinteten Microarrays sind sehr feuchtigkeitsempfindlich und werden daher im Exsikkator bei Unterdruck über einem Trocknungsmittel aufbewahrt.

# 2.16. Hybridisierung und Auswertung von Microarrays

Die Genexpressionsanalyse dient der Untersuchung der qualitativen und quantitativen Zusammensetzung zellulärer mRNA, wofür typischerweise die relativen Expressionsstärken zweier Proben miteinander verglichen werden. Bei diesen Proben handelt es sich um RNA, welche aus Zellen gewonnen wird, die zum einen unter Standard- oder Referenzbedingungen und zum anderen unter den experimentell zu untersuchenden Bedingungen gewachsen sind. Die für die Hybridisierung verwendete RNA (*target*) wird mit den fluoreszierenden Cyanin Farbstoffen Cy3 und Cy5 markiert. Die jeweils korrespondierenden Fluoreszenzsignale der beiden Farbstoffe können nach der Hybridisierung der *probes* mit den *targets* unabhängig

voneinander detektiert werden. Die Verwendung von Fluoreszenz ermöglicht, im Vergleich zu einer radioaktiven Markierung, neben höherer Sensitivität und Auflösung, das Auswerten von zwei oder mehreren verschiedenen Signalen in einem Experiment. Im Anschluss an eine Normalisierung der Daten kann die Intensität der Hybridisierungssignale miteinander verglichen werden.

### 2.16.1. Markierung von RNA mit Cy3/Cy5 Fluoreszenzfarbstoff

Die Markierung der isolierten RNA (2.5.3.) erfolgte über eine reverse Transkription mit ,*random nonamers*'. Für die Hybridisierung der *scorecard* (siehe 2.15.3.) wurde in Form eines *spike mix* spezifische RNA (Lucedia<sup>TM</sup> Universal ScoreCard<sup>TM</sup>, AMERSHAM Biosciences, Freiburg) für beide Farbreaktionen mitgeführt. Während des Umschreibens der RNA werden CyDye markierte dCTPs in die cDNA eingebaut. Die restlichen dNTPs werden ohne eine Markierung als dCTP Nukleotid Mix zu gegeben. In dieser Arbeit wurde der "Cyscribe First Strand cDNA Labeling kit" (AMERSHAM Biosciences, Freiburg) benutzt.

#### Durchführung:

Als erster Schritt des Markierens erfolgte das Annealing der "random nonamers" an die RNA, wofür 25 µg RNA benötigt wurden. Aufgrund unterschiedlicher Markierungseffizienz der Farbstoffe wurden für eine Hybridisierungsreaktion ein Ansatz Cy3 und zwei Ansätze Cy5 markiert. Es gilt zu bemerken, dass die Cyanin Farbstoffe stark lichtempfindlich sind und daher die Proben möglichst lichtgeschützt behandelt wurden. Das Annealing fand im Thermocycler statt:

Annealing-Ansatz:		Anne	aling-Programm:			
25	μg	isolierte RNA	01	LIDHT	ON	110°C
1	μl	,spike mix' für Cy3	02	PAUSE	60 min	70°C
		oder Cy5 Reaktion	03	HOLD	5 min	70°C
1	μl	,random nonamers'	04	TEMP GRADIENT	3,0°C/s	to 50°C
ad 11	μl	H <sub>2</sub> O <sub>bidest</sub>	05	TEMP GRADIENT	1,0°C/s	to 40°C
			06	TEMP GRADIENT	0,3°C/s	to 35°C
			07	TEMP GRADIENT	0,2°C/s	to 30°C
			80	TEMP GRADIENT	0,1°C/s	to 25°C
			09	TEMP GRADIENT	0,1°C/s	to 22°C
			10	HOLD	20 min	22°C
			11	TEMP	FOREVER	4°C

Im Anschluss wurden dem Annealing-Ansatz folgende Komponenten für die reverse Transkription hinzugefügt:

#### Ansatz für die reverse Transkription:

```
11
       μl
               Annealing Ansatz
4
               5x "CyScript Puffer"
       μl
               0.1 M DTT
2
       μl
1
               "dCTP Nukleotid Mix"
       μl
               dCTP CyDye-markierte Nukleotide
1
       μl
               (Cy3 oder Cy5!)
               CyScript Reverse Transkriptase (100 U/µI)
       μl
```

Die 20 μ1 Reaktionsansatz wurden kurz gemischt, abzentrifugiert und 1,5 h bei 42°C im Mastercycler Gradient (EPPENDORF, Hamburg) inkubiert.

Um die Template RNA und nicht eingebaute CyDye Nukleotide zu entfernen, erfolgte anschließend eine zügige Aufreinigung der markierten cDNA. Zunächst wurde die RNA durch Zugabe von 2 µl 2,5 M NaOH in kurze Oligomere hydrolysiert. Die Ansätze wurden 10 s gevortext, kurz abzentrifugiert und 15 min bei 37°C für die Hydrolyse im Mastercycler Gradient (EPPENDORF, Hamburg) inkubiert. Um die Lösung zu neutralisieren wurden im Anschluss 10 µl 2 M HEPES hinzugegeben, 10 s gevortext und wiederholt zentrifugiert.

Die Aufreinigung der markierten cDNA erfolgte über Säulen des CyScribe GFX Purification Kits (AMERSHAM Biosiences, Freiburg). Für jeden Hybridisierungsansatz wurde eine Säule verwendet. Auf die Säule wurden 500 μl Capture Buffer, ein Markierungsansatz Cy3 markierte cDNA und zwei Markierungsansätze Cy5 markierte cDNA gegeben und vorsichtig mit der Pipette gemischt. Es folgte ein Zentrifugationsschritt für 30 s bei 13000 Upm, wobei die cDNA an die Säulenmatrix bindet. Im Anschluss wurde der Durchfluss verworfen und mit 600 μl Waschpuffer bei 13.000 Upm für 30 s gewaschen. Der Durchfluss wurde erneut verworfen und der Waschschritt weitere zweimal wiederholt. Danach erfolgte die Elution der markierten *target* cDNA in ein neues Röhrchen, indem 70 μl 70°C warmer Elutionspuffer auf die Säule gegeben, für 5 min bei RT inkubiert und für 1 min bei 13000 Upm zentrifugiert wurde. Das markierte cDNA enthaltende Eluat konnte bis zur Hybridisierung, vor Licht geschützt, auf Eis gelagert werden.

## 2.16.2. Wavelength scan zur Überprüfung der Markierungsreaktion

Um die Effizienz der Markierungsreaktion bewerten zu können, wurde ein Wavelengthscan mit der markierten cDNA durchgeführt. Die Absorbtionsmaxima von eingebautem Cy3/Cy5 liegen bei 550/650 nm, von freiem Cy3/Cy5 bei ca. 520/600 nm.

Für die Messung wurden  $10~\mu l$  der aufgereinigten Markierungsreaktion 1:6 mit  $H_2O_{RNase~frei}$  verdünnt und das Absorbtionsspektrum in einer Quarzküvette (d=1 cm) gemessen. Das Spektrum wurde über einen Bereich von 700 - 500 nm gefahren und die Extinktion bei 550 und 650 nm ermittelt.

Die eingebauten pmol der beiden Farbstoffe konnten nun nach folgender Formel berechnet werden:

```
pmol Cy3 bzw. Cy5 in der Probe = (A/E) x (z \mu l) x Verdünnungsfaktor x 10^{12}

A = Absorbtion von Cy3 bei 550 nm bzw. Cy5 bei 650 nm

E = Extinktionskoeffizient von Cy3 (150000 \ l \ mol^{-1} \ cm^{-1})

bzw. Cy5 (250000 \ l \ mol^{-1} \ cm^{-1})

z \mu l = Volumen der cDNA-Lösung; \mu l entspricht 10^{-6}
```

## 2.16.3. Hybridisierung der probes mit markierter target-cDNA

Die Hybridisierung der *probes* mit der markierten *target*-cDNA erfolgte vollautomatisch mit dem Automated Slide Processor (ASP Lucidea, Amersham Pharmacia Biotech). Vorbereitend wurde die *target*-cDNA denaturiert und die benötigten Hybridisierungskammern des ASP gespült. Die Denaturierung der *target*-cDNA erfolgte durch 2 minütiges Inkubieren bei 95°C, anschließender Abkühlung auf Eis und der Zugabe von 50 μl *microarray hybridization buffer* des "Cyscribe First Strand cDNA Labeling kits" (AMERSHAM Biosiences, Freiburg), sowie 110 μl 100% (v/v) Formamid. Der Ansatz wurde gemischt, bei 13.000 Upm kurz abzentrifugiert und lichtgeschützt auf Eis gelagert. Anschließend wurden 200 μl des Hybridisierungsansatzes mit einer Hamiltonspritze in die Hybridisierungskammer des ASP, in welche bereits ein gespotteter Microarray eingespannt wurde, injiziert. Die Hybridisierung erfolgte mit dem folgenden Programm:

## ASP-Hybridisierungsprogramm:

1101 1	i joiraisiorang.	<u>sprogramm.</u>
00	Wait	Check dummy slides
Überpr	üfen, ob in jeder K	ammer ein dummy Slide ist
01	Flush,	Leeren der Schläuche und Kammern von alten Lösungen
02	Heat,	Kammern auf 38°C aufheizen
03	Flush	Kammern mit Isopropanol spülen und leeren
04	Flush	Kammern mit Wasser spülen und leeren
05	Air Pump	Kammern für 60 s mit Luft spülen
06	Heat	Switch Off, Heizung ausschalten
07	Prime	Schläuche mit Wash1 (Lösung 3) füllen
80	Wait	INSERT TEST SLIDES - test slides einlegen
09	Heat	Kammern auf 42°C aufheizen
10	Wait	Inject probe solution - Proben injizieren
11	Mix	mischen der Lösungen in den Kammern für 12 h
12	Flush	mit Lösung Wash1 spülen
13	Heat	Kammern auf 45°C aufheizen
14	Mix	Mischen mit Wash1 für 10 min
15	Prime	Schläuche mit Wash2 (Lösung 4) füllen
16	Flush	Mit Wash 2 spülen
17	Mix	Mischen mit Wash2 für 10 min
18	Flush	mit Lösung Wash2 spülen
19	Mix	Mischen mit Wash2 für 4 min
20	Prime	Schläuche mit wash3 (Lösung 2) füllen
21	Flush	mit Lösung Wash3 spülen
22	Prime	Schläuche mit Isopropanol füllen
23	Flush	mit Isoporpanol spülen
24	Flush	Kammern mit Luft spülen
25	AirPump Kamme	ern mit Luft trocknen
26	Heat	Heizung ausschalten
27	Wait	REMOVE TEST SLIDES
Test sli	des entfernen und	mit dummy slides ersetzen
28	Heat	Heizen auf 48°C
29	Flush	mit Lösung Wash2 spülen
30	Mix	Mischen mit Wash2 für 5 min
31	Prime	Schläuche mit Wasser füllen
32	Flush	mit 6 ml Wasser spülen
33	Mix	Mischen mit Wasser für 5 min
34	Flush	mit 6 ml Wasser spülen
35	Flush	mit Luft spülen
36	Prime	Schläuche mit Isopropanol füllen
37	Flush	mit Isopropanol spülen
38	Flush	mit Luft spülen
39	Air Pump	Kammern mit Luft trocknen
[Ende]		

#### Erläuterung der Programmbefehle:

Flush: Spülen der Hybridisierungskammern und Schläuche

Prime: Füllen der Schläuche bis zur Hybridisierungskammer

Mix: Mixen der Lösung innerhalb der Kammer durch Vor- und Zurückpumpen

Heat: Heizen der Kammer auf eine bestimmte Temperatur

Air Pump: Trocknen der Kammer durch Luft

Wait: Warten auf Bestätigung des Users

#### Lösungen:

Lösung 1: Wasser (bidest) (Min. Volumen, 200ml)

Lösung 2: Wash 3 (0.1X SSC)

Lösung 3: Wash 1 (1xSSC/0.2%SDS)

Lösung 4: Wash 2 (0.1xSSC/0.2%SDS)

Lösung 5: Isopropanol

### 2.16.4. Quantifizierung der Arraydaten mit ,GenePix Pro 4.0'

Zur Verarbeitung der Hybridisierungsdaten wurden diese mit dem Scanner GenePix 4000B (AXON Instruments, Union City) und dem Programm GenePix visualisiert. Bereits während des Scannens erfolgte eine erste Normalisierung der Arraydaten über die *scorecard*. Hierbei handelt es sich um spezifische DNA, welche mit den *probes* gespottet wurde (2.15.3.). Diese DNA wurde mit spezifischer RNA des *spike mix* (2.16.1.) hybridisiert. Während der Markierungsreaktion wurde diese RNA mit den Farbstoffen Cy3 und Cy5 im Verhältnis 1:1 markiert. Auf diesem bekannten Verhältnis beruhend, wurde nun in einem Prescann anhand der entsprechenden Spots der *scorecard* die Scannstärke (PMT 635 – Cy5; PMT 532 – Cy3) für die beiden Farbstoffe ermittelt. Anschließend erfolgte der Hauptscann mit folgenden Einstellungen:

Pixelsize 10 µm

Lines to average 1

Focus Position ( $\mu$ m) 0

Scan area 1968 x 7020 Pixel

Scan duration 351 s.

Für die Quantifizierung der Fluoreszenzwerte jedes einzelnen Spots in GenePix wurde zunächst eine Schablone auf den Chip gelegt, womit jedem Spot die entsprechende Annotation zugewiesen und der auszuwertende Bereich festgelegt wurde. Mit Hilfe eines eigens dafür entwickelten Computerprogramms (Ehrenreich, unveröffentlicht) wurde die Schablone, basierend auf einer Excel-Belegungsliste, in Form eines gal files generiert.

Die Ausrichtung dieser Schablone wurde für jeden einzelnen Spot manuell überprüft. Die Berechnung der Fluoreszenzwerte, des Backgrounds und der Standardabweichung des Backgrounds für beide Farbstoffe und des *Ratio of Medians*, *Ratio of Means* und des *Regression Ratios* erfolgte für jeden Spot automatisch durch das Programm GenePix. Bei den *Ratios of Medians*, *Means* und dem *Regression Ratio* handelt es sich um jeweils unterschiedliche mathematische Ansätze zu Ermittlung des mittleren Expressionswertes eines Spots. Jeder Spot besteht aus einer Vielzahl von einzelnen Pixeln, für welche jeweils ein Wert der Pixel-Intensität für die Wellenlänge 1 ( $I_{P,\lambda I}$ ) und der Pixel-Intensität für die Wellenlänge 2 ( $I_{P,\lambda 2}$ ) vorliegt. Die Intensitäten der zwei Wellenlängen, der den definierten Spot umgebenden Background-Pixel werden ebenfall mit einbezogen ( $I_{B,\lambda I}$  und  $I_{B,\lambda 2}$ ). Der mittlere Expressionswert eines Spots wird aus den einzelnen Pixel-Intensitäten berechnet.

Der *Ratio of Medians* wird häufig zur Berechnung von Mittelwerten herangezogen, da starke Abweichungen einzelner Werte das Endergebnis kaum beeinflussen.

Ratio of Medians:

$$\frac{\{(I_P, \lambda_2)_n\}_{med} - \{(I_B, \lambda_2)_n\}_{med}}{\{(I_P, \lambda_1)_n\}_{med} - \{(I_B, \lambda_1)_n\}_{med}}$$

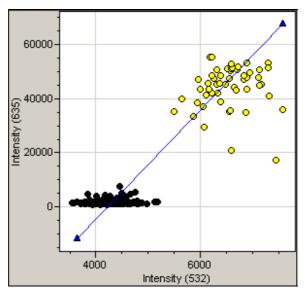
Der Ratio of Means hingegen kann von starken Abweichungen einzelner Werte verfälscht werden.

Ration of Means:

$$\frac{\sum_{i=1}^{n} \langle I_{P, \lambda 2} - \{(I_{B, \lambda 2})_{m}\}_{med} \rangle_{i}}{\sum_{i=1}^{n} \langle I_{P, \lambda 1} - \{(I_{B, \lambda 1})_{m}\}_{med} \rangle_{i}}$$

Für die Berechnung des *Regression Ratio*s ist eine Definition von Foreground- oder Background-Pixeln nicht nötig. Jeder Pixel innerhalb des zweifachen Spotdurchmessers, unabhängig von seiner Position, wird einbezogen. Das Verhältnis der beiden Wellenlängen zueinander wird mittels linearer Regression zwischen den zwei sich ergebenden Pixelwolken ermittelt (Abbildung 2.13.). Die Ergebnisse konnten in Form einer Datentabelle und eines Scatterplots eingesehen werden. An dieser Stelle erfolgte eine weitere Normalisierung der Daten. Unter der Voraussetzung, dass sich unter den zu vergleichenden Bedingungen nur die Expression vereinzelter Gene verändert und die Expression eines Großteils der Gene unverändert bleibt,

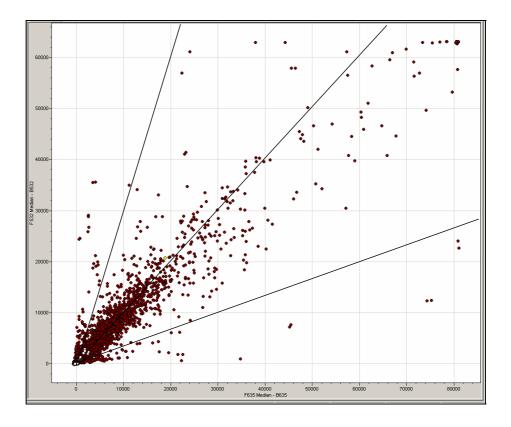
wurde im Scatterplot die Hauptwolke der Gene auf die Winkelhalbierende gelegt (Abbildung 2.14.), also in den Bereich eines Regulationsfaktors von 1. Eine logarithmische Auftragung der Expressionswerte vereinfacht dieses Verfahren zusätzlich, da Spots mit einem Expressionsfaktor von annähernd 1 gestaucht werden und sich die Dichte der Hauptwolke der Daten erhöht (Abbildung 2.15.). Eine weitere Methode zur Normalisierung der Arraydaten bestand darin, einzelne potentiell konstitutive Gene mittels *Real-time* PCR zu überprüfen und diese zur Auswertung der Chips auf ein Verhältnis von 1:1 zu normalisieren.



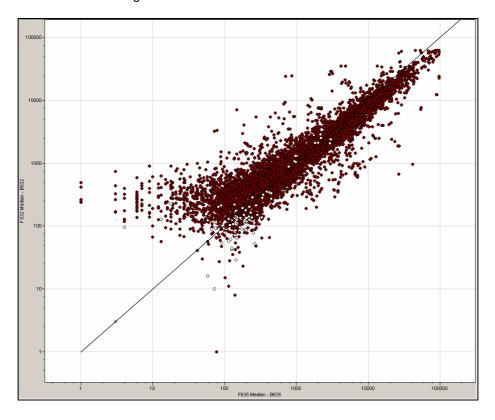
**Abbildung 2.13.:** Regression Ratio eines einzelnen Spots. Aufgetragen sind die Intensitäten der einzelnen Pixel bei 532 nm gegen die Intensitäten bei 635 nm. Die schwarzen Punkte liegen im Background, die gelben heben sich vom Background ab. Durch die blaue Linie wird das durchschnittliche Verhältnis der Pixelintensitäten und damit der Regression Ratio dargestellt.

#### 2.16.5. Analyse der Transkriptionsdaten

Für die weitere Bearbeitung und Analyse der normalisierten Transkriptionsdaten wurden diese aus GenePix in Exel exportiert. Hier wurden sämtliche Daten auf Qualität und somit Aussagekraft überprüft. Alle Daten wurden daraufhin gefiltert, dass die Fluoreszenzsignale für Rot und Grün selbst, und nach Abzug der Standardabweichung des Backgrounds einen Wert größer Null besaßen. Ein weiteres Merkmal für die Qualität eines Spots war, dass die nach unterschiedlichen mathematischen Verfahren berechneten Expressionsverhältnisse; *Ratio of Medians, Ratio of Means* und der *Regression Ratio*; für einen Spot nicht mehr als 30% voneinander abweichen. Bei starken Diskrepanzen dieser Werte, handelte es sich meist um Artefakte der Hybridisierung. Im Anschluss an die Evaluation der Daten wurden alle ORFs deren Expression sich unter den jeweiligen Bedingungen mindestens um einen Faktor 3 änderte herausgefiltert, mit anderen Chipdaten verglichen und stoffwechselanalytisch beurteilt.



**Abbildung 2.14.:** Darstellung der Expressionswerte eines Microarrays als Scatterplot im Programm GenePix. Die Hauptwolke der Gene liegt im Bereich der Winkelhalbierenden. Die Linien ober- und unterhalb der Winkelhalbierenden kennzeichnen die Bereiche, in denen eine Regulation der Expression um einen Faktor kleiner und größer 3 stattfindet.



**Abbildung 2.15.:** Logarithmische Auftragung der Expressionswerte eines Microarrays als Scatterplot im Programm GenePix. Die Hauptwolke der Gene liegt nach der Normalisierung im Bereich der Winkelhalbierenden.

# 3. Ergebnisse

Ziel dieser Arbeit war die vollständige Entschlüsselung des Genoms des Wildtypstammes *Bacillus licheniformis* DSM13 und, basierend auf dieser Sequenz, die Erstellung eines Genomchips für anschließende Transkriptionsanalysen mit dem Focus auf den C-Stoffwechsel. Im Folgenden werden der Verlauf des Sequenzierungsprojektes und die Ergebnisse der Analyse des genetischen Potentials von *B. licheniformis* DSM13 dargestellt. Des Weiteren werden die Erstellung des Genomchips, sowie die Resultate der darauf aufbauenden Transkriptionsanalysen beschrieben.

## 3.1. Das Sequenzierungsprojekt Bacillus licheniformis

Das Sequenzierungsprojekt umfasst mehrere Phasen, die Erstellung der Genbank, die Rohsequenzierung, die Editierung der Sequenzdaten, das so genannten *Gap closure* und eine abschließende Qualitätsverbesserung der vollständigen Genomsequenz.

### 3.1.1. Sequenzierungsmethode

Für die Sequenzierung des Genoms von *B. licheniformis* DSM13 wurde die *whole shotgun* Methode eingesetzt. Das Prinzip dieses Verfahrens beruht auf einer Fragmentierung der genomischen DNA, z.B. durch Scheren oder einen partiellen Verdau, und der Sequenzierung dieser Fragmente.

Im Rahmen dieses Sequenzierungsprojektes wurde die chromosomale DNA durch Scheren mechanisch gebrochen (2.4.7.), was den Vorteil einer Zerkleinerung unabhängig von spezifischen Erkennungssequenzen hat. Im Anschluss an eine Agarosegelelektrophorese (2.4.9.) wurden Fragmente mit einer Größe von 2–3 kb aus dem Gel eluiert (2.4.6.) und ihre Enden zu blunt ends aufgefüllt (2.6.3.). Die darauf folgende Dephosphorylierung der Fragmente (2.6.2.), sollte ein unspezifisches Zusammenlagern der chromosomalen DNA Abschnitte verhindern, welches bei der späteren Sequenzierung zu enormen Problemen hätte führen können. Der für die Sequenzierung verwendete Vektor pTZ19R<sup>Cm</sup> wurde mit den Inserts chromosomaler DNA von *B. licheniformis* ligiert (2.6.4.) und in *E. coli* DH5α transformiert (2.6.5.). Inserttragende Klone konnten über die Chloramphenicolresistenz des Vektors und Blau-Weiß-Screening (2.6.7.) selektiert, Stammkulturen angelegt (2.3.), die Plasmide präpariert und eine so genannte *shotgun library* für die Rohsequenzierung erstellt werden. Die für dieses Projekt erstellte *Shotgun* Genbank umfasst etwa 31.390 Klone.

#### 3.1.2. Rohsequenzierung

Im Laufe der Rohsequenzierung wurden die Inserts der *shotgun library* mit Hilfe von Vektorspezifischen Standardprimern, welche unmittelbar *down*- und *upstream* der multiplen Klonierungsstelle von pTZ19R<sup>Cm</sup> binden, von beiden Seiten ansequenziert (2.9.). Das heißt, mit den beiden Sequenzläufen, welche eine durchschnittliche Länge von 695 bp hatten, wurde in das bis zu 3 kb große Insert hineingelesen, dieses aber nicht vollständig sequenziert. Die Assemblierung dieser Einzelsequenzen, d.h. ihr Zusammensetzen anhand von Sequenzübereinstimmungen zu größeren *Contigs*, erfolgte auf bioinformatischem Wege wie unter 2.10. beschrieben. Um in etwa die Anzahl der notwendigen Rohsequenzierungsläufe abschätzen zu können, wurde folgende Formel verwendet (Lander and Waterman, 1988; Waterman, 1984):

$$N = \frac{A \times L}{\omega}$$

N stellt die statistisch zu erwartenden Sequenzläufe, für die Sequenzierung eines Genoms von der Länge L, bei einer durchschnittlichen Leselänge  $\omega$  der einzelnen Sequenzen und einer angestrebte Abdeckung A einer Base durch einzelne Sequenzläufe dar. Für *B. licheniformis* wurde aufgrund der bereits vorliegenden Genomsequenz von *B. subtilis* 4,21 Mbp (Kunst et al., 1997) von einer Genomgröße L in einem Bereich von 4-5 Mbp ausgegangen. Bei einer angestrebten Abdeckung A von 6 und einer durchschnittlichen Länge  $\omega$  = 695 bp pro Sequenzlauf, sind nach der genannten Formel 34532-43165 Rohsequenzierungsläufe notwendig. "Zwischenassemblierungen" zeigten die jeweils aktuelle Anzahl von Contigs während der Rohsequenzierung und konnten dazu verwendet werden, zu kontrollieren, ob die Sequenzierung dem Verlauf der von Lander und Waterman (Lander and Waterman, 1988; Waterman, 1984) vorhergesagten Rohsequenzierungsphase folgt. Von den Autoren wurde zur Kontrolle der Rohsequenzierungsphase folgende Formel vorgeschlagen:

$$P = e^{-N*\omega/L}$$

.

Mit dieser Formel soll die Anzahl von Contigs vorhergesagt werden, die unter idealen Bedingungen, d.h. bei konstanter Sequenzlänge und Genbanken mit statistischer Verteilung der Inserts, einer gegebenen Anzahl von Sequenzen entspricht. Dabei ist P die Wahrscheinlichkeit dafür, dass eine beliebige Base bei einer gegebenen Anzahl von Sequenzen noch nicht sequenziert wurde. Basierend auf dieser Formel kann der Verlauf der Rohsequenzierung mit

einem idealen Verlauf einer Rohsequenzierung verglichen werden, wobei mit der Zunahme der Rohsequenzen eine stetige Abnahme der Contigs und somit der Lücken im Genom verbunden ist. Im Laufe des hier beschriebenen Sequenzierungsprojektes stagnierte die Anzahl der Contigs größer als 3 kb bei einer Anzahl von 318 trotz eines Anstiegs der Sequenzläufe auf 39871. Bei einem idealen Verlauf der Rohsequenzierungsphase hätte hier bereits eine Abnahme auf ca. 40 Contigs erfolgen sollen. Grund hierfür war das Fehlen spezifischer Inserts in der Shotgun-Library und damit eine nicht optimale statistische Verteilung der Genbank-Inserts. Eine mögliche Ursache dafür liegt darin, dass die Ribosomenbindestelle der BacillusmRNA eine deutlich höhere Bindungseffizienz zu Escherichia coli Ribosomen aufweist als E.coli-mRNA. Des Weiteren verfügt E. coli über ein zusätzliches ribosomales Protein S1, welches die Bindung von mRNA an das Ribosom unterstützt. Dies hat zur Folge, dass Bacillus-Gene in E. coli höchst effektiv exprimiert werden. Befindet sich also auf einem Insert der Library ein Gen, welches ein, bei einer Überexpression, auf E. coli toxisch wirkendes Genprodukt codiert, so fehlt dieser Klon. Auf diese Weise ist es möglich, dass mit einem high copy Vektor wie pTZ19R nicht alle Bereiche des B. licheniformis Genoms in einer E. coli Genbank erfasst werden können.

Nach einer Editierung der bestehenden Sequenz, wurde daher zum Lückenschluß mittels PCR Techniken übergegangen.

## 3.1.3. Editierung der Sequenz

Die Editierung umfasst in erster Linie eine Überarbeitung der nach der Rohsequenzierung vorliegenden Sequenzdaten, basierend auf der *shotgun* Genbank. Hier galt es, die Sequenz in einzelnen Bereichen zu verbessern, Contigs zu verlängern, gegebenenfalls Lücken mit Hilfe der Plasmidgenbank zu schließen und Regionen falscher Assemblierung aufgrund repetitiver Sequenzen aufzulösen.

## 3.1.3.1. Sequenzverbesserung mittels *Plasmid walking*

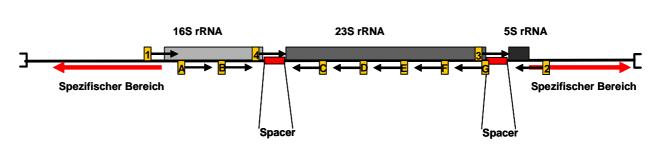
Für die Verbesserung des bestehenden Sequenzdatensatzes im Programm Gap4 wurde dieser systematisch durchgearbeitet, um Bereiche unzureichender Qualität ausfindig zu machen. Entscheidende Kriterien waren hier ein minimaler *quality value* von 40 (2.13.) für jede Base und/oder eine minimale Abdeckung von zwei qualitativ guten Sequenzläufen je Base. Bereiche, die diesen Anforderungen nicht entsprachen, wurden mit Hilfe von *Plasmid walkings* (2.11.1.) aufgewertet.

#### 3.1.3.2. Verlängerung der Contigs mittels Plasmid walking

Da wie unter 3.1.2. beschrieben die Inserts der *shotgun library* im Rahmen der Rohsequenzierung nicht vollständig, sondern nur von den Enden ausgehend sequenziert wurden, lagen einige Sequenzläufe der Inserts an den Enden eines Contigs. Mittels *Plasmid walking* (2.11.1.) konnten diese Sequenzen mit ihrem Gegenlauf verbunden, somit der entsprechende Contig verlängert und gegebenenfalls Lücken geschlossen werden. Im Laufe von zwei Editierungsrunden wurde so die Anzahl der Contigs größer als 3 kb auf 234 reduziert. An dieser Stelle war das Potential der Plasmidgenbank zum Schließen von Sequenzlücken erschöpft, und es wurden alternative Methoden, beschrieben unter 3.2.4., eingesetzt.

# 3.1.3.3. Identifizierung und Überprüfung von rRNA-Clustern und repetitiven Sequenzen

Repetitive Sequenzen, wie Insertions-Elemente, tRNA- und rRNA-Cluster, führten aufgrund sehr hoher Sequenzhomologien während der Assemblierung der Rohsequenzdaten zu Falschassemblierungen, so genannten Misassemblies. Die Consensus-Sequenz war in diesen repetitiven Bereichen aufgrund der hohen Anzahl der Einzelsequenzen von hoher Qualität. Auffällig wurden Misassemblies dieser Art zum einen durch die Anhäufung von Sequenzen und zum anderen durch Basenfehlpaarungen beim Übergang in spezifische Sequenzbereiche, welche eine starke Qualitätsabnahme der Consensus-Sequenz zur Folge hatten. Handelte es sich um einen repetitiven Bereich von einer Länge bis zu ca. 1500 bp, lag der Gegenlauf mit hoher Wahrscheinlichkeit vollständig oder zumindest teilweise im spezifischen Sequenzbereich. Auf diese Weise konnten durch Zuordnung eines Sequenzlaufes mit repetitiver Sequenz zu seinem entsprechenden Gegenlauf mit spezifischer Sequenz Misassemblies dieser Art aufgelöst werden. Ein weitaus größeres Problem stellten falsche Assemblierungen im Bereich von Clustern ribosomaler RNA dar. Der gesamte Bereich eines rRNA-Clusters in B. licheniformis umfasst ca. 5 kb und setzt sich aus der 16S rRNA mit 1549 bp, der 23S rRNA mit 2929 bp und der 5S rRNA mit 125 bp zusammen (Abbildung 3.1.). Die Sequenzabschnitte der einzelnen rRNAs sind durch Spacer mit einer Größe von ca. 125-335 bp voneinander getrennt.



**Abbildung 3.1.:** Schematische Darstellung eines rRNA-Clusters in *B. licheniformis*. Durch gelbe Kästchen sind die Primer zur Amplifikation des rRNA-Clusters (1+2), zur Sequenzierung der rRNA-Cluster spezifischen Spacer (3+4) und zur vollständigen Sequenzierung des 16S und 23S Bereiches (A-G) dargestellt.

Seit der Sequenzierung des Genoms von *B. subtilis* (Kunst et al., 1997) ist bekannt, dass dieser Organismus 10 rRNA-Cluster besitzt, womit im Falle von *B. licheniformis* ebenfalls von einer hohen Anzahl an rRNA-Clustern ausgegangen werden konnte. Neben den spezifischen Sequenzbereichen an den Enden der Cluster und im Bereich der Spacer unterscheiden sich die Sequenzen der einzelnen rRNA-Cluster lediglich in einzelnen Basenpaaren. Um die *Misassemblies* der einzelnen rRNA-Cluster aufzulösen, wurden zunächst sämtliche Bereiche im Genom isoliert, welche einen Übergang von spezifischer Sequenz in den Bereich von 16S rRNA bzw. 5S rRNA beinhalteten. Mindestens 200 bp vor Beginn der rRNA-Cluster wurden spezifisch bindende Primer generiert, paarweise kombiniert und *Long-range* PCRs (2.7.2) mit einer Elongationszeit von 5 min durchgeführt. Auf diese Weise konnten insgesamt sieben rRNA Cluster im Genom von *B. licheniformis* identifiziert werden. Die PCR Produkte, welche jeweils den Sequenzbereich der ribosomalen RNA überspannten, wurden vollständig sequenziert und die generierten Sequenzen zur Auflösung der *Misassemblies* genutzt.

## 3.1.4. Lückenschluß (, Gap closure')

1

Nach Abschluss der Rohsequenzierungsphase und der Editierung, galt es, die verbleibenden 234 Lücken zu schließen. In dieser Phase des Sequenzierprojektes wurde auf unterschiedliche Methoden der PCR-Technik zurückgegriffen.

# 3.1.4.1. Lückenschluß über den Vergleich mit homologen "ORF-Clustern" in B. subtilis

Bereits während der Editierung wurde die unvollständige Sequenz des Genoms von *B. liche-niformis* einem ORF-*Finding* mit dem Programm YACOP (Tech and Merkl, 2003) (2.14.1.) und einer automatischen Annotation mit Hilfe des ERGO Programmpaketes (Overbeek et al., 2003) unterzogen (2.14.2.). Schon zu diesem Zeitpunkt zeigten sich in vielen Bereichen er-

staunliche Homologien des *B. licheniformis* Genoms zur vollständigen Genomsequenz von *B. subtilis* (Kunst et al., 1997). Die hohe Ähnlichkeit einzelner ORFs und vor allem ihrer Anordnung in Clustern ließ genomweite Übereinstimmungen in der ORF-Verteilung bei diesen beiden Organismen vermuten. Darauf basierend wurde mit sämtlichen Contigenden des *B. licheniformis* Sequenzdatensatzes ein Abgleich gegen das *B. subtilis* Genom, wie unter 2.12.1. beschrieben, durchgeführt. Konnten zwei Contigenden als potentiell zusammengehörig identifiziert werden, wurde je nach der zu erwartenden Größe der Lücke eine Standardoder *Long-range PCR* mit an den Contigenden generierten Primern durchgeführt. Konnte ein PCR Produkt amplifiziert werden, folgte nach der Aufreinigung die Sequenzierung desselben und die Sequenz wurde dem Datensatz hinzugefügt.

Diese Methode erwies sich als äußerst effektiv, da insgesamt 132 Lücken geschlossen und somit die Anzahl der Gaps auf 102 reduziert werden konnte.

#### 3.1.4.2. Lückenschluß mittels Vectoretten Technik

Eine weitere Methode, die zum Lückenschluß eingesetzt wurde, ist die Vectoretten Technik (2.12.2.). Hierbei handelt es sich um ein Verfahren, mit dessen Hilfe das Ende eines Contigs in die noch unbekannte Sequenz der Lücke hinein verlängert werden kann. Über diese Verlängerung der Sequenz wird ein Überlappen mit bereits bekannten Sequenzabschnitten angestrebt. Das Prinzip der Vectoretten Technik (2.12.2.) beruht auf der Herstellung chromosomaler DNA Fragmente mit definierten Enden, welche mit einer synthetisch hergestellten Oligokassette ligiert werden. Für die PCR wird zum einen ein Standardprimer eingesetzt, welcher an die Oligokassette bindet und zum anderen ein für das Ende eines Contigs spezifischer Primer. Auf diese Weise generierte PCR-Produkte haben eine Länge von maximal 2-3 kb, womit diese Methode eher zum Schließen kürzerer Lücken geeignet ist. Insgesamt konnten 26 Gaps mit Hilfe dieses Verfahrens überbrückt werden.

#### 3.1.4.3. Lückenschluß über kombinatorische Multiplex PCR

Für den Großteil der verbleibenden Lücken nach dem Genomvergleich mit *B. subtilis* wurde eine Größe bis zu 10 kb erwartet. Aus diesem Grund wurde, parallel zum Schließen kürzerer Lücken mittels Vectoretten Technik, die kombinatorische Multiplex PCR vorbereitet. Die Multiplex PCR eröffnet die Möglichkeit in einem PCR Ansatz unter Verwendung mehrerer Primer mehrere Produkte zu amplifizieren. Ziel, des in dieser Arbeit verwendeten Ansatzes, war, sämtliche Primer, welche an den Enden der verbleibenden Contigs generiert wurden, gegeneinander zu testen und auf Basis der chromosomalen DNA von *B. licheniformis* die lü-

ckenüberspannenden DNA Fragmente zu amplifizieren. Um nicht jeden Primer einzeln gegen alle verbleibenden Primer zu testen, was bei 100 Lücken und somit 200 Primern 200\*199/2=19900 PCRs bedeutet hätte, sollten die Primer sinnvoll kombiniert und mittels Multiplex PCR getestet werden. In dieser Arbeit wurden zwei kombinatorische Ansätze verwendet. Im ersten Ansatz wurden die Primer, modifiziert, nach dem von TETTELIN et al. (Tettelin et al., 1999) vorgestellten System kombiniert (2.12.3.1.) und im zweiten, darauf folgenden Ansatz nach dem Prinzip des Kombinationsdreieckes gepoolt (2.12.3.2.).

Der von TETTELIN vorgeschlagene Pooling-Algorithmus stellte sich als enorm aufwendig in der Vorbereitung heraus. Darüber hinaus zeigte sich, dass entgegen der Voraussage der Autoren, bei weitem nicht jeder Primer mit jedem anderen Primer in Kombination vorlag. Dieser Erkenntnis zum Trotz, konnten auf diesem Wege 28 Lücken überbrückt und die mittels Vectoretten Technik geschlossenen Lücken bestätigt werden.

Für die zu diesem Zeitpunkt verbliebenen 48 Contigs wurden, um die Spezifität des Primerannaelings auf chromosomaler DNA zu erhöhen, neue Primer mit einer Länge von 27-30 bp und einer Schmelztemperatur von 60-65°C generiert und ein weiterer Kombinationsansatz, basierend auf dem Kombinationsdreieck, gewählt (2.12.3.2.). Mit Hilfe dieser vergleichsweise simplen Strategie konnte für alle verbleibenden 48 Lücken der fehlende DNA Abschnitt amplifiziert werden. Nach Isolation des zugehörigen Primerpaares erfolgte eine PCR im größeren Maßstab, welche aufgereinigt, sequenziert und die Sequenz in den Datensatz eingefügt wurde.

### 3.1.5. Qualitätsverbesserung (Polishing)

Bereits bevor die vollständige Sequenz des Genoms von *B. licheniformis* vorlag, wurde damit begonnen, den gesamten Datensatz einer weiteren Qualitätsüberprüfung, ähnlich der Editierung (2.11.) zu unterziehen. Der Schwerpunkt dieses *Polishings* lag ebenfalls darauf, Bereiche mit einem *quality value* kleiner als 40 (2.13.) durch zusätzliche Sequenzläufe, basierend auf dem *Plasmid walking* oder der Sequenzierung weiterer PCRs, qualitativ aufzuwerten. Als Richtlinie galt eine mindestens zweifache Abdeckung einer Base durch zwei voneinander unabhängige Sequenzen.

Ein weiterer Punkt des *Polishings* umfasste die Überprüfung von *framshifts*. Nach wiederholten Updates des ORF-*Findings* (2.14.1.) und der automatischen Annotation (2.14.2.), und dem Beginn der manuellen Annotation (2.14.3.), wiesen einige ORFs eine offensichtliche Verschiebung des Leserahmens auf. Das heißt, einige bekannte und bereits experimentell verifizierte Gene lagen je in zwei getrennten ORFs vor. Die Bereiche eines solchen potentiellen

frameshifts wurden im Datensatz überprüft. Häufig waren in diesen Sequenzabschnitten keine Fehler offensichtlich. Zur Überprüfung betroffener Sequenzabschnitte wurden diese mittels PCR amplifiziert, wiederholt sequenziert und gegebenenfalls im Datensatz berichtigt.

#### 3.1.6. Zusammenfassung

Nach der Rohsequenzierungsphase des Genomprojektes B. licheniformis und der anschließenden Editierung auf Basis der Plasmidgenbank konnten durch den Genomvergleich mit B. subtilis, die Vectorettentechnik und die kombinatorische Multiplex PCR alle Lücken im Sequenzdatensatz geschlossen werden. In Tabelle 3.1. ist der Verlauf des Genomprojektes über die einzelnen Phasen der Bearbeitung hinweg beschrieben. So nimmt die Anzahl der Sequenzläufe von der Rohsequenzierung hin bis zum Polishing von 39.871 auf 47.121 zu. Es ist ebenfalls eine Zunahme der Länge des gesamten Datensatzes auf insgesamt 4.386.868 bp zu beobachten. Die Gesamtlänge des Datensatzes am Ende der Sequenzierung entspricht jedoch nicht der Länge des Genoms von B. licheniformis, da der Datensatz einzelne Sequenzen, welche aufgrund unzureichender Qualität nicht in die Genomsequenz eingefügt wurden, beinhaltet. Die Anzahl der Contigs und damit die Anzahl der Lücken nimmt im Verlauf der Sequenzierung von 318 bis hin zur finalen Gesamtsequenz ab. Während der Editierung des Datensatzes konnten bereits 84 Lücken geschlossen werden. Sehr effektiv erwies sich das Gap closure über den Vergleich der bestehenden Sequenz mit dem Genom von B. subtilis. Die verbleibenden 102 Lücken konnten durch ein Zusammenspiel von der Vectoretten Technik und zwei unabhängigen Ansätzen der kombinatorischen Multiplex PCR geschlossen werden. Nach einem abschließenden Polishing der Genomsequenz von B. licheniformis wies diese eine durchschnittliche Abdeckung von 7,45 Sequenzläufen pro Base auf.

**Tabelle 3.1.:** Entwicklung des Sequenzdatensatzes von *B. licheniformis* von der Rohsequenzierung, über die Editierung und den Lückenschluß bis hin zum Polishing.

	Assemb- lierung der Rohse- quenzen	Editierung	Genom- vergleich mit B. subtilis	Vecto- retten Technik	1. kombi- natorische Multiplex PCR	2. kombi- natorische Multiplex PCR	Polishing
Anzahl der Sequenzläufe	39.871	44.012	44.811	45.001	45.200	46.071	47.121
Länge des gesamten Datensatztes [bp]	4.009.666	4.136.894	4.226.919	4.280.868	4.317.107	4.432.401	4.386.868
Anzahl der Contigs größer als 3 kb	318	234	102	76	48	1	1
Geschlossene Lücken	/	84	132	26	28	48	/
Abdeckung des Consensus	/	/	7,33	7,3	7,27	7,21	7,45

#### 3.2. Annotation des Genoms von Bacillus licheniformis

Nachdem die vollständige Sequenz des Genoms von B. licheniformis vorlag, wurde sie einem erneuten ORF-Finding (2.14.1.), einer Überprüfung der hinzugekommenen ORFs mit Hilfe des Programmes Arthemis (2.14.1.) und einer weiteren automatischen Annotation unterzogen (2.14.2.). Hierbei wurden jeweils die bereits vorliegenden Daten, hervorgegangen aus der Bearbeitung des unvollständigen Sequenzdatensatzes, berücksichtigt. Jeder ORF wurde anschließend durch eine manuelle Überprüfung der Annotation (2.14.13.) verifiziert. Zusätzlich erfolgte eine funktionelle Annotation über Homologien zu B. subtilis wie unter 2.14.4. beschrieben. Für orthologe ORFs wurde die B. subtilis Annotation übernommen. Insgesamt konnten 4286, auf dem Genom von B. licheniformis codierte, ORFs identifiziert werden (Tabelle 3.2.). 1409 ORFs wurden über funktionelle Annotation durch Vergleich lokaler Sequenzhomologien und über Homologien zu B. subtilis eine Funktion zugewiesen. 1704 Gene, welche ein Ortholog in B. subtilis aufweisen, deren Funktion bisher jedoch unbekannt ist, wurden analog der B. subtilis Annotation als ,Y'-Gene bezeichnet. Als ,putatitve' Gene wurden 444 ORFs annotiert, für welche aufgrund hoher Sequenzhomologien eine Funktion vermutet wird, die, bis zu diesem Zeitpunkt, jedoch nicht experimentell bestätigt wurde. ORFs, welche keine hohe Ähnlichkeit, zu in den Datenbanken abgelegten Genen, aufwiesen, wurden als ,hypothetisch' (483) bezeichnet, traten Ähnlichkeiten zu hypothetischen Genen auf, wurden diese in *B. licheniformis* als ,konserviert hypothetisch' annotiert (153). Des Weiteren konnten 7 Cluster ribosomaler RNA und insgesamt 72 tRNAs identifiziert werden.

Ein erster Blick auf die Annotation des vollständigen *B. licheniformis* Genoms zeigte, das die wesentlichen Gene der zentralen Stoffwechselwege wie Glykolyse, Pentosephosphatweg und des Tricarbonsäure Zyklus vorhanden sind. Ebenso wurden entscheidende Gene des Aminosäure- und Vitaminstoffwechsels, des Sec Proteinsekretionsystems (*secADFEGY*, *ffh*, *ftsY*), die Signalpetidasen *sipSTVW*, außer *sipU* und eine Vielzahl der Gene des Sporulationssystems identifiziert. Aufgrund des enormen Datenumfangs wird in dieser Arbeit keine detaillierte Annotationstabelle aller 4.286 ORFs wiedergegeben. Die Genomsequenz von *B. licheniformis* DSM13 ist veröffentlicht (Veith et al., 2004) und bei NCBI unter der Zugangsnummer AE017333 frei zugänglich.

Tabelle 3.2.: Übersicht über die "Annotationsgruppen" der B. licheniformis ORFs

ORFs	Anzahl	%
- insgesamt	4.286	100
- mit zugewiesener Funktion	1.409	32,9
- ,Y'-Gene	1.704	39,8
- putativ	444	10,4
- hypothetisch	483	11,3
- konserviert hypothetisch	153	3,6
- rRNA Cluster	7	0,5
- tRNAs	72	1,6

## 3.3. Analyse des Genoms von Bacillus licheniformis

## 3.3.1. Allgemeine Eigenschaften des *B. licheniformis* Genoms

Im folgenden Abschnitt soll auf die allgemeinen Eigenschaften des Genoms von *B. licheniformis* eingegangen werden, welches in Form einer zirkulären ORF Karte in Abbildung 3.2. dargestellt ist. Die gesamte Sequenz von *B. licheniformis* umfasst 4.222.748 bp, davon bilden 87,9% die codierende Sequenz für 4.286 ORFs. 74,3% der ORFs sind auf dem *leading strand* lokalisiert. Sehr deutlich ist die co-direktionale Replikation und Transkription anhand der ORF Verteilung auf die beiden DNA Stränge, ausgehend vom Replikationsursprung (Position

0 in Abbildung 3.2.), zu erkennen, welche typisch für Bacilli und auch Clostridien ist. Der durchschnittliche G+C Gehalt des *B. licheniformis* Genoms liegt bei 46,2%. Insgesamt konnten 4 Prophagen identifiziert werden, welche zumeist von IS-Elementen flankiert werden. Bei drei dieser Prophagen ist eine lokale Abweichung des G+C Gehaltes zu beobachten. Insgesamt 72 tRNAs sind in der Sequenz von *B. licheniformis* zu finden. Sowohl ein Großteil der tRNAs, als auch 4 der 7 rRNA Cluster sind in der näheren Umgebung des Replikationsursprungs lokalisiert. In Tabelle 3.3. sind die allgemeinen Daten des *B. licheniformis* Genoms noch einmal zusammengefasst. Ebenfalls in Abbildung 3.2. dargestellt, sind alle *B. licheniformis* Gene welche ein orthologes Gen in *B. subtilis* aufweisen. Auf den Vergleich dieser beiden Organismen auf genomischer Ebene soll unter 3.4.2. detailliert eingegangen werden. Ein erster Blick auf die Verteilung der zu *B. subtilis* orthologen ORFs in *B. licheniformis* (Abbildung 3.2., vierter Zirkel von außen in grün) weist jedoch bereits auf eine sehr hohe Homologie der beiden Genome hin.

Tabelle 3.3.: Allgemeine Eigenschaften des Genoms von Bacillus licheniformis

Gesamtlänge der Sequenz	4.222.748 bp	100%
- kodierende Sequenz	3.711.795 bp	87,9%
- durchschnittlicher G+C Gehalt		46,2%
Gesamtanzahl der ORFs	4.286	100%
- auf dem <i>leading strand</i> lokalisierte ORFs	3.185	74,3%
Anzahl der Prophagen	4	
Anzahl rRNA Cluster	7	
Anzahl der tRNAs	72	

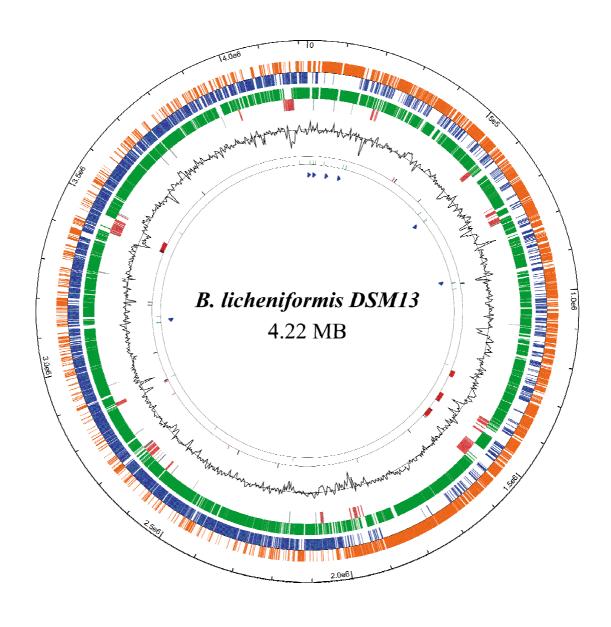


Abbildung 3.2.: Zirkuläre ORF-Karte des Genoms von B. licheniformis

Die codierende Sequenz des Genoms ist, je nach Orientierung auf dem *leading* oder *lagging strand*, in orange bzw. blau dargestellt. *B. licheniformis* ORFs mit einem Homologon in *B. subtilis* sind in dem grünen Kreis aufgezeigt. Gene, welche eine abweichende *codon usage* aufweisen (*alien genes*), sind im 5. Kreis von außen rot abgebildet. Der nach innen folgende schwarze Graph gibt lokale Abweichungen des durchschnittlichen G+C Gehaltes wieder, wobei Peaks nach außen, erhöhten Werten entsprechen. Prophagen sind in Form roter Kästchen hervorgehoben und IS-Elemente als kurze schwarze Linien. Die tRNA Gene sind als kurze grüne Linien dargestellt und im inneren des Zirkels sind die rRNA Cluster als blaue Pfeile zu finden.

#### 3.3.2. Ermittlung des Replikationsursprunges

Eine Möglichkeit für die Bestimmung des Replikationsursprunges stellt die Berechnung des GC-Skews dar (Lobry, 1996). Die meisten *Bacteria* weisen im *leading strand* mehr Guaninals Cytosin-Basen auf (Mrazek and Karlin, 1998). Das Verhältnis von Guanin und Cytosin kann mit Hilfe des GC-Skews ermittelt werden. Aufgrund der Unverteilung dieser beiden Basen, welche vermutlich auf einem unterschiedlichen Selektionsdruck auf den *leading* und auf den *lagging strand* beruht (Lobry, 1996), erfolgt im Idealfall ein Vorzeichenwechsel sowohl am Replikationsursprung, als auch am Terminus der Replikation. Im Falle des Chromosomes von *B. licheniformis* wurde die kumulative GC-Skew Analyse (Grigoriev, 1998) verwendet, wobei das Ergebnis eine Kurve mit einem Minimum am Origin und einem Maximum am Terminus liefern sollte:

$$GCSkew = \frac{G(i,i+n) - C(i,i+n)}{G(i,i+n) + C(i,i+n)}$$

Das Ergebnis für das *B. licheniformis* Genom ist in Abbildung 3.3. dargestellt. Im Vorfeld wurde das *B. licheniformis* Chromosom, in Erwartung des Replikationsursprunges in der Nähe von *dnaA*, an dieser Stelle linearisiert. Aus diesem Grund befindet sich das Minimum der Kurve, welches den Ori (Origin of Replication) beschreibt, an Position 0. Das Maximum liegt bei ca. 2.000.000 bp, dem Terminus der Replikation.

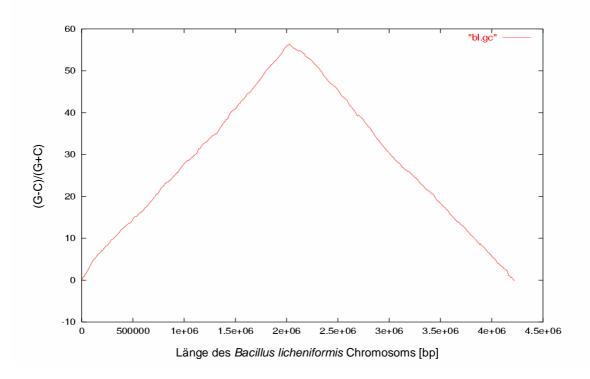
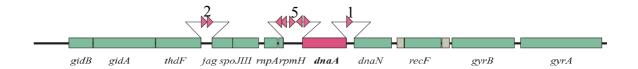


Abbildung 3.3.: Kumulativer GC-Skew des B. licheniformis Genoms

Der kumulative GC-Skew bestätigte die Lokalisation des *Ori* kurz vor dem Gen *dnaA*. Abbildung 3.4. zeigt die Umgebung des *Origin of Replication*. Eine Analyse der das Gen *dnaA* umgebenden Sequenz ließ eine charakteristische Anordnung von insgesamt 8 DnaA-Bindemotiven (TTATCCACA), angeordnet in drei Boxen, erkennen. Das Arrangement der, den Replikationsursprung von *B. licheniformis* umgebenden Gene und der DnaA Bindeboxen ist homolog zur Anordnung in *B. subtilis* (Ogasawara et al., 1985; Yoshikawa and Ogasawara, 1991).



**Abbildung 3.4.:** Umgebung des *dnaA* Gens am *Ori* von *B. licheniformis*Das Gen *dnaA* ist als pinkfarbene Box, Gene in der Umgebung mit bekannter Funktion als grüne Boxen und Gene unbekannter Funktion als graue Boxen dargestellt. Bindemotive für DnaA sind in Form pinkfarbener Pfeile gekennzeichnet.

#### 3.3.3. Identifizierung von Fremdgenen und hochexprimierten Genen

Ein immer wieder interessanter Aspekt ist die Betrachtung einer Genomsequenz unter evolutionären Gesichtspunkten. Der Transfer von Genen zwischen verschiedenen Spezies, bezeichnet als horizontaler Gentransfer, wird als eine der treibenden Kräfte in der mikrobiellen Evolution gesehen (Koonin et al., 2001). Eine Möglichkeit Gene aufzuspüren, welche potentiell durch horizontalen Gentransfer in ein ursprüngliches Genom gelangt sind, so genannte Fremdgene (alien genes), ist die Untersuchung der codon usage. Die codon usage ist die Speziestypische Verwendung bestimmter Triplets zur Codierung der einzelnen Aminosäuren, was durch den degenerierten Code ermöglicht wird. Es wird davon ausgegangen, dass die codon usage in phylogenetisch verwandten Spezies ähnlich ist (Hori and Osawa, 1987; Lawrence and Ochman, 1997).

Ein Aspekt der Analyse des *B. licheniformis* Genoms soll die Untersuchung der ORFs mit Hinblick auf potentielle Fremdgene sein. Dazu wurde das Programm SIGI (Merkl, 2004) verwendet (2.14.6.), welches nicht nur auf die Suche einzelner Fremdgene, sondern insbesondere auf das Auffinden von *genomic islands*, d.h. ganzer Cluster von *alien genes*, spezialisiert ist. Als auffällig, bezüglich ihrer *codon usage*, erwiesen sich insgesamt 426 ORFs, welche anschließend von SIGI auf ihre Herkunft hin untersucht wurden. So konnten 127 Gene der 426 ORFs als hochexprimierte Gene und 263 Gene als potentielle *alien genes* eingestuft wer-

den. Im Anhang A.4. und A.5. findet sich eine vollständige Auflistung dieser ORFs. In den Tabellen 3.4. und 3.5. ist jeweils eine Übersicht der hochexprimierten Gene und der potentiellen Fremdgene wiedergegeben. 86 der hochexprimierten Gene konnte während der Annotation (3.2.) eine Funktion zugewiesen werden, wobei 52 Gene davon ribosomale Proteine codieren. 14 Gene wurden als ,Y'- Gene - analog zu *B. subtilis* (siehe 3.2.)-, 7 Gene mit einer putativen Funktion und 20 Gene als hypothetisch oder konserviert hypothetisch annotiert (Tabelle 3.4.). Von den 263 potentiellen Fremdgenen konnte 24 Genen eine Funktion zugewiesen werden, 81 wurden mit einer putativen Funktion annotiert und 120 als hypothetisch oder konserviert hypothetisch eingestuft (Tabelle 3.5.). Bei 38 Genen aller potentiellen *alien genes* handelt es sich um so genannte ,Y'-Gene (siehe 3.2). Eine Assoziation mit Phagen wird für insgesamt 40 der 263 Gene angenommen.

An dieser Stelle sei bemerkt, dass sich unter den potentiellen Fremdgenen sechs ORFs befinden, bei denen es sich um 2 Operonstrukturen für Typ I Restriktionssysteme handelt, worauf im Kapitel 3.4.1.2. eingegangen wird.

In Abbildung 3.2. ist im 5. Zirkel von außen die Verteilung der *alien genes* in rot dargestellt. Verblüffend ist die Übereinstimmung der von *alien genes* überspannten Bereiche mit den Regionen, in denen *B. licheniformis* keine Homologien zu *B. subtilis* aufweist. Darüber hinaus gehen lokale Abweichungen des G+C Gehaltes mit den Regionen der Fremdgene konform. Weiterhin ist sehr auffällig, dass 3 der 4 Prophagen in Bereichen des Genoms lokalisiert sind, in welchen nur geringe bis keine Homologien zu *B. subtilis* ORFs ersichtlich sind. Hier sind des Weiteren deutliche Abweichungen des lokalen G+C Gehaltes zu vermerken und ein Großteil der ORFs ist als potentielle Fremdgene markiert. Die Gene des vierten Phagen an Position 1.320.000 weisen hingegen eine hohe Homologie zu dem PBSX Prophagen von *B. subtilis* auf und sind vollkommen unauffällig bezüglich ihrer *codon usage* im Vergleich zur *codon usage* des gesamten Genoms von *B. licheniformis*.

Tabelle 3.4.: Übersicht der als hochexprimierte Gene eingestuften ORFs

Hochexprimierte Gene	Anzahl	%
- insgesamt	127	100
- mit zugewiesener Funktion	34	26,8
- ribosomale Proteine	52	40,9
- ,Y'-Gene	14	11
- putativ	7	5,5
- hypothetisch / konserviert hypothetisch	20	15,7

Tabelle 3.5.: Übersicht der als potentielle Fremdgene eingestuften ORFs

Potentielle Fremdgene	Anzahl	%
insgesamt	263	100
mit zugewiesener Funktion	24	9,1
- davon Phagen assoziiert	2	
,Y'-Gene	38	14,4
- davon Phagen assoziiert	5	
putativ	81	30,8
- davon Phagen assoziiert	33	
hypothetisch/konserviert hypothetisch	120	45,6

### 3.4. Genomevergleich von B. licheniformis mit B. subtilis

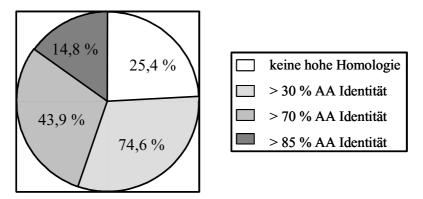
Bereits während der Sequenzierung von *B. licheniformis* gab es Hinweise auf eine hohe Homologie zu *B. subtilis* auf genomischer Ebene (3.1.4.1.). Nach Kenntnis der vollständigen Genomsequenz von *B. licheniformis* wird im folgenden Kapitel auf Gemeinsamkeiten und Unterschiede dieser beiden Organismen bezüglich ihrer Aminosäure- und Nukleotidsequenz eingegangen.

## 3.4.1. Genomvergleich auf Ebene der Aminosäuresequenz

#### 3.4.1.1. Orthologe Gene in B. licheniformis und B. subtilis

Ein Vergleich der Genome von *B. licheniformis* und *B. subtilis* auf Ebene der Aminosäuresequenz (2.14.4.) zeigte eine erstaunliche Konservierung der Funktionalität dieser beiden Organismen. Knapp drei Viertel aller *B. licheniformis* ORFs weisen eine mindestens 30%ige Ähnlichkeit zu Genen von *B. subtilis* auf (Abbildung 3.5.). Ganze 43,9% sind zu mindestens 70% identisch und 14,8% zeigen eine Identität von mehr als 85% zu in *B. subtilis* codierten Proteinen. Die orthologen Gene der beiden Organismen umfassen unter anderem die zentralen Stoffwechselwege der Glykolyse, des Pentosephosphat-Zyklus und des Tricarbonsäure-Zyklus. Sowohl *B. licheniformis*, als auch *B. subtilis* sind in der Lage auf Mineralmedium mit definierten Energiequellen ohne Zusatz von Vitaminen oder Aminosäuren zu wachsen. Diese Tatsache bestätigt sich durch ihre Fähigkeit der Synthese aller zum Wachstum notwendiger Bausteine und Cofaktoren. Weiterhin verfügen beide Organismen über alle Gene des Sec Proteinsekretionssystems: *secA* (BLi03773), *secDF* (BLi02891), *secE* (BLi00118), *secG* (BLi03643), *secY* (BLi00153), *ffh* (BLi01818) und *ftsY* (BLi01816). In beiden Genomen sind

die vier Signalpeptidasen *sipS* (BLi00675), *sipT* (BLi01655), *sipV* (BLi01122) und *sipW* (BLi02638) vorhanden. Lediglich ein Ortholog zu *sipU*, einer fünften Signalpeptidase in *B. subtilis*, konnte in *B. licheniformis* nicht identifiziert werden. Aufgrund der Tatsache, dass *B. licheniformis* am Ende der logarithmischen Wachstumsphase sporuliert, ist es nicht überraschend, dass der Großteil aller für die Sporulation notwendigen Gene identifiziert werden konnte.



**Abbildung 3.5.:** Kreisdiagramm zur Darstellung der prozentualen Anteile der ORFs mit einer Aminosäureidentität von weniger als 30%, mehr als 30%, mehr als 70% und mehr als 85% zu *B. subtilis* Genen bezogen auf alle in *B. licheniformis* identifizierten ORFs.

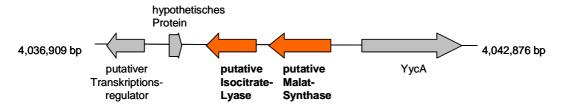
#### 3.4.1.2. Gene von B. licheniformis ohne entsprechendes Ortholog in B. subtilis

An dieser Stelle soll es um Gene von *B. licheniformis* gehen, welche kein entsprechendes Gegenstück in *B. subtilis* aufweisen. Trotzdem sei erwähnt, dass die auffälligste Abweichung, auf Seiten des *B. subtilis* Genoms, das Vorhandensein von Polyketidsynthetasen ist. Diese Enzyme synthetisieren, ähnlich der Fettsäuersynthese, makromolekulare Sekundärmetabolite. Dazu gehören Antibiotika, wie z.B. Tetracycline und Rifamycine. Poliketidsynthetasen umfassen beinahe 4% des *B. subtilis* Genoms und konnten in *B. licheniformis* nicht identifiziert werden.

Insgesamt 1091 *B. licheniformis* Gene (25,4%) sind *unique* im Vergleich zu *B. subtilis*, das heißt, es liegt eine Aminosäure Identität von weniger als 30% vor (Abbildung 3.5.). 52% der *B. licheniformis* eigenen Gene sind hypothetisch, 11% sind in Transportfunktionen involviert, 8% in regulatorische Prozesse und 17% in verschiedenste enzymatische Reaktionen. Eine vollständige Liste der *unique*-Gene von *B. licheniformis* kann im Anhang der Veröffentlichung des *B. licheniformis* Genoms eingesehen werden (Veith et al., 2004). Vier Gene davon (BLi00401, BLi00402, BLi00403, BLi00404) codieren für den nicht-ribosomalen Peptidsynthetasekomplex der Lichenysin-Biosynthese (Konz et al., 1999). Lichenysin ist ein antimikrobiell und Oberflächen-wirksames Lipopeptid, ähnlich dem sehr prominenten Lipopeptid Bacitracin aus *B. subtilis*. Einige *B. licheniformis* Stämme besitzen ebenfalls das genetische Potential zur Synthese von Bacitracin. Dafür in Frage kommende Gene konnten in *B. licheni-*

formis DSM13 nicht gefunden werden. Während der Untersuchung des Wachstums von B. licheniformis DSM13 in Minimalmedium, versetzt mit unterschiedlichen C-Quellen, konnte in verschiedenem Ausmaß die Bildung eines rotbraunen Farbstoffes beobachtet werden. Involviert in die Synthese dieses potentiellen Siderophores könnten mehrere Gene sein: BLi01185, BLi01186, BLi01187 und BLi01188 (Martinez et al., 1994; Moss et al., 1999). Für B. licheniformis wurden Schwierigkeiten bei der Transformation mit heterologer DNA beschrieben (O'Sullivan et al., 2000). In einem möglichen Zusammenhang mit dieser Beobachtung steht die Identifizierung von zwei Operonstrukturen, welche für Typ I Restriktionssysteme codieren (BLi04315, BLi04316, BLi04318 und BLi00743, BLi00746, BLi00744, BLi00745). Die beiden Operone zeigen eine hohe Ähnlichkeit zu dem Hsd-System (Host specificity of DNA). Hierbei handelt es sich zumeist um Heteropentamere, welche DNA ausgehend von einer asymmetrischen Erkennungssequenz zufällig, aber mindestens 1000 bp entfernt, schneiden (Wilson and Murray, 1991).

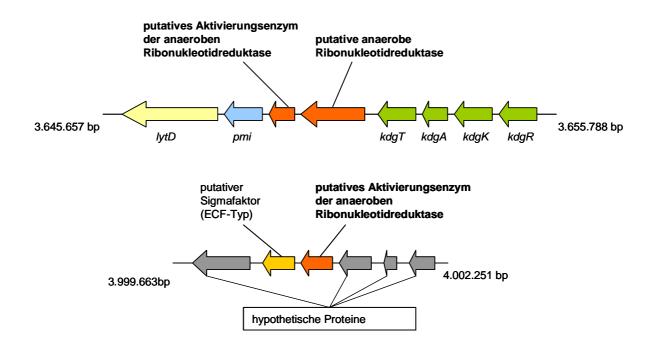
Die Sequenz von *B. licheniformis* DSM13 eröffnet Einblicke in die metabolischen Fähigkeiten dieses Organismus. So konnten z.B. zwei Gene identifiziert werden, welche starke Homologien, zum einen zur Isocitrat-Lyase (BLi04207) (Sharma et al., 2000) und zum anderen zur Malat-Synthase (BLi04208) (Sandeman et al., 1991) aufweisen. Diese beiden Gene sind essentiell für den Glyoxylatzyklus, welcher zum Wachstum mit C2-Substraten befähigt (Gottschalk, 1986). Die Gene der Isocitrat-Lyase und der Malat-Synthase sind direkt nebeneinander auf dem Chromosom von *B. licheniformis* angeordnet und werden, wie in Abbildung 3.6. dargestellt, *upstream* von dem ,Y-ORF' *yyc*A und *downstream* von einem putativen Transkriptionsregulator flankiert. Eine funktionelle Isocitrat-Lyase und Malat-Synthase würden *B. licheniformis* zum Wachstum auf eigenen Endprodukten der unvollständigen Oxidation oder Gärung befähigen.



**Abbildung 3.6.:** Schematische Anordnung der Gene für die putative Isocitrat-Lyase, die putative Malat-Synthase und angrenzender Gene im Genom von *B. licheniformis*.

Des Weiteren wurden im Genom von *B. licheniformis* putative Gene für eine anaerobe Ribonukleotidreduktase (BLi03824) und zwei zugehörige potentielle Aktivierungsenzyme (BLi03823, BLi04172) identifiziert (Torrents et al., 2000), deren Funktionalität ein anaerobes

Wachstum auf Glucose ermöglichen würde. Ribonukleotidreduktasen sind essentiell für die Reduktion von Ribonukleosid-Triphosphaten, um eine Versorgung der DNA-Synthese mit Desoxyribonukleotiden zu gewährleisten. In Abbildung 3.7. ist die Anordnung der entsprechenden ORFs in *B. licheniformis* dargestellt. Die Gene der anaeroben Ribonukleotidreduktase und eines der Aktivierungsenzyme sind direkt nebeneinander lokalisiert und werden *upstream* von einem Gencluster der Pectinverwertung (*kdgTAKR*) flankiert. Unmittelbar *downstream* befinden sich die Gene der Mannose-6-Phosphat-Isomerase (*pmi*) und des Autolysins *N*-Acetylglucosaminidase (*lytD*). Das Gen des zweiten putativen Aktivierungsenzyms befindet sich etwa 344 kb *upstream* und wird von mehreren hypothetischen Genen und einem putativen Sigmafaktor begrenzt. Die im Bezug auf den Glyoxylatzyklus und die anaerobe Ribonukleotidreduktase durchgeführten Wachstumsversuche werden unter 3.6.1. beschrieben.



**Abbildung 3.7.:** Schematische Anordnung der Gene für die putative anaerobe Ribonukleotidreduktase, die zwei identifizierten putativen Aktivierungsenzyme und die jeweils angrenzenden Gene im Genom von *B. licheniformis*. (Erläuterung im Text)

#### 3.4.1.3. Industriell relevante Exoenzyme mit und ohne Ortholog in *B. subtilis*

B. licheniformis ist ein wichtiger Produzent von Exoenzymen, worauf seine hauptsächliche industrielle Nutzung aufbaut. Sein genetisch verankertes Potential zur Exoenzymproduktion wurde im Vergleich zu B. subtilis untersucht. In diesem Zusammenhang konnten einige prominente Gene identifiziert werden. Eine Vielzahl dieser Enzyme besitzt ein entsprechendes Homolog in B. subtilis (Tabelle 3.6.), über andere Exoenzyme verfügt B. subtilis nicht (Tabel-

le 3.7.). Besonders erwähnt sei im ersten Fall die Glutaminsäure-spezifische Protease (BLi00340) (Kakudo et al., 1992; Svendsen and Breddam, 1992).

**Tabelle 3.6.:** Industriell relevante Exoenzyme von *B. licheniformis* mit einem entsprechendem Ortholog in *B. subtilis* 

Gen ID	Funktion	B. subtilis
BLi03021	α-L-Arabinofuranosidase	abfA
BLi01295	Arabinan-endo-1,5-L-Arabinase	abnA
BLi04220	Arabinan-endo-1,5-α-L-Arabinosidase	yxiA
BLi04276	Arabinogalactan-endo-1,4-β-Galactosidase	yvfO
BLi00447	β-Galactosidase	lacA
BLi04214	β-Glucosidase	bglH
BLi02088	Endo-1,4-β-Glucanase	bglC
BLi00655	Endo-1,4-β-Xylanase	yjeA
BLi00340	Glutamat-spezifische Protease	mpr
BLi02827	Levanase	sacC
BLi03707	Levanase	yveB
BLi03706	Levansucrase	sacB
BLi04019	minor extrazelluläre Serine-Protease	vpr
BLi01123	minor extrazelluläre Serine-Protease	epr
BLi01404	Pectat-Lyase	pel
BLi03053	Pectat-Lyase	pelB
BLi03741	Pectat-Lyase	yvpA
BLi04177	PeptidaseT	pepT
BLi01399	Polysaccharid-abbauendes Enzym	ykfC
BLi02863	Protease	yrrN
BLi02862	Protease	yrrO

Unter den Exoenzymen, welche kein Ortholog in *B. subtilis* aufweisen, befinden sich eine vielfach beschriebene alkalische Serin Protease - das Subtilisin Carlsberg - (BLi01109) (Jacobs et al., 1985; Smith et al., 1968; Syed et al., 1993), die maltogene α-Amylase (BLi00658) (Kim et al., 1992) und die hitze- und pH-stabile α-Amylase (BLi00656) (Gray et al., 1986; Kandra et al., 2002; Stephens et al., 1984; Yuuki et al., 1985). Weiterhin konnten im Genom von *B. licheniformis* eine Cellulose-1,4-β-Cellobiosidase (BLi01881), eine Endo-

1,4-β-Mannosidase (BLi01883) und eine Endo-1,4-Glucanase (BLi01880) identifiziert werden. Putative Gene, welche für ein Cellobiose spezifisches Phosphotransferasesystem (PTS) und β-Glucosidasen codieren, wurden ebenfalls gefunden. Ein weiterer Hinweis darauf, dass *B. licheniformis* nicht nur Cellulose abbauen, sonder auch Chitin verwerten kann, liegt in der Existenz zweier Gene, welche für Chitinasen codieren (BLi00338, BLi00339).

Tabelle 3.7.: Industriell relevante Exoenzyme von B. licheniformis ohne Ortholog in B. subtilis

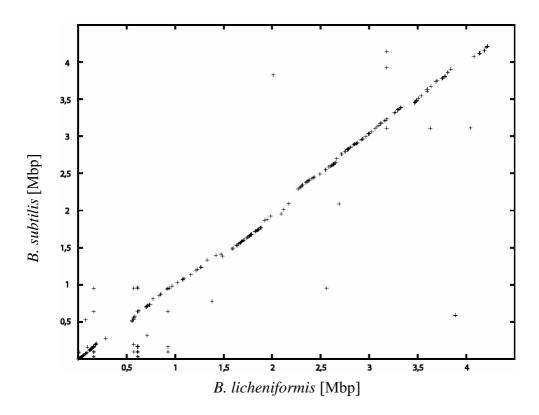
Gene ID	Funktion
BLi00656	α-Amylase (EC 3.2.1.1)
BLi03543	α-Glucosidase
BLi02117	α-Glucosidase (EC 3.2.1.20)
BLi01882	Cellulase (EC 3.2.1.4)
BLi01881	Cellulose-1,4-β-Cellobiosidase
BLi00338	Chitinase (EC 3.2.1.14)
BLi00339	Chitinase (EC 3.2.1.14)
BLi01883	Endo-1,4-β-Mannosidase
BLi01880	Endo-1,4-Glucanase (EC 3.2.1.4)
BLi00545	Esterase/Lipase
BLi03370	Esterase/Lipase
BLi02821	Esterase/Lipase
BLi00658	maltogene $\alpha$ -Amylase (EC 3.2.1.1)
BLi04129	Pectat-Lyase
BLi03498	Pectin-Methylesterase
BLi01109	Subtilisin Carlsberg (EC 3.4.21.62)
BLi01909	Zink-Protease (EC 3.4.99)

#### 3.4.2. Genomvergleich auf Ebene der Nukleotidsequenz

Der Vergleich des *B. licheniformis* Genoms mit dem *B. subtilis* Genom auf Ebene der Aminosäuresequenz hatte eine erstaunliche Übereinstimmung der Funktionalität von drei Vierteln der Gene ergeben (3.4.1.). Da die Reihenfolge von Genen weniger konserviert ist als ihr Vorhandensein (Huynen and Bork, 1998; Lathe et al., 2000), erfolgte eine weitere Analyse des *B. licheniformis* Genoms im Vergleich zu *B. subtilis* auf der Ebene der Nukleotidsequenz.

#### 3.4.2.1. Alignment der Genomsequenzen

Ein sehr effizientes Verfahren des *whole genome alignments* stellt das Programm MUMmer (Delcher et al., 2002; Kurtz et al., 2004) dar. Mit dieser Software war es möglich, sich sehr schnell einen allgemeinen Überblick über die Ähnlichkeit der Genomsequenzen von *B. licheniformis* und *B. subtilis* zu verschaffen. Die graphische Darstellung des Ergebnisses ist in Abbildung 3.8 als so genannter MUMmer Plot gezeigt. Anhand der diagonal eingezeichneten Treffer des Alignments lässt sich eine sehr hohe Colinearität der beiden Genome erkennen. Lediglich in einzelnen Bereichen erfolgt eine Unterbrechung dieser Diagonale und damit der Übereinstimmung der DNA-Sequenz. Die Analyse der Genome mit Hilfe des Programms MUMmer zeigt eine hohe Colinearität der Nukleotidsequenzen von *B. licheniformis* und *B. subtilis* über weite Bereiche.



**Abbildung 3.8.:** Vergleich der Genome von *B. licheniformis* und *B. subtilis* anhand der Nukleotidsequenz mittels der MUMmer Software. Die diagonal eingezeichneten Treffer zeigen eine hohe Colinearität der *B. licheniformis* und *B. subtilis* Genome.

#### 3.4.2.2. Konservierung funktioneller Nukleotidsequenzen

Nachdem sowohl der Vergleich von *B. licheniformis* und *B. subtilis* auf Ebene der Nukleotidsequenz, als auch auf Ebene der Aminosäuresequenz enorme Ähnlichkeiten aufzeigte, lag die Vermutung nahe, dass orthologe, co-linear angeordnete Gene auch einer ähnlichen Regulation

unterliegen. Exemplarisch wurden daher die Bindemotive zweier Sigmafaktoren untersucht. Mit Hilfe des Sigmafaktors L ( $\sigma^L$ ) wird in *B. subtilis* die Transkription eines Regulons, welches unter anderem den Katabolismus einiger Aminosäuren, wie z.B. Arginin, Ornithin, Leucin und Valin beinhaltet, reguliert (Debarbouille et al., 1999; Gardan et al., 1997). Das Operon des Acetoin Katabolismus (Huang et al., 1999) und die Induktion des Levanase Gens (Debarbouille et al., 1991) unterliegen ebenfalls einer Regulation durch  $\sigma^L$ . Die Consensussequenz des  $\sigma^L$  Bindemotivs (Helmann and Moran Jr., 2002) wurde in der DNA-Sequenz von *B. licheniformis* gesucht und vor den in Tabelle 3.8. aufgelisteten Genen gefunden. Dazu gehören *levD* (Levanase), *rocD* (Arginin/Ornithin Katabolismus), *acoA* (Acetoin Katabolismus), *ptb* (potentielle Phosphat-Butyryltransferase) und *yveP* (Funktion unbekannt; Ähnlichkeit zu einer Glycosyltransferase).

**Tabelle 3.8.:** Sigmafaktor L (σ<sup>L</sup>) (BLi03678) Bindemotiv

Gene	Promotorsequenz <sup>a</sup>	5' UTR <sup>b</sup>
<i>lev</i> D (BLi02831)	cagttttgt <b>tggca</b> cgatac <b>ttgcat</b> tataaataagcgaacg	39
rocD (BLi00422)	$\verb attaagagttggca   csttgcat   attaagatagtgtgaaa  $	67
acoA (BLi00849)	$\verb tcggagagc  \verb tggca  cgattt  \verb ttgcat  cataacgtgtgagtaa $	38
Ptb (BLi02586)	ttttcatac <b>tggca</b> cgaaac <b>ttgcat</b> gatatagtgggcggat	39
yveP (BLi03689)	tgggataca <b>tggca</b> gc-ttc <b>ttgcat</b> tccgccgcttgatttt	83
B. licheniformis		
Consensus <sup>c</sup>	TGGCAccTTGCAT	38-83
B. subtilis		
Consensus <sup>c</sup>	TGGcACTTGCAT	

<sup>&</sup>lt;sup>a</sup> Die Basen der Consensussequenz sind fett hervorgehoben. Großbuchstaben stehen für eine hohe, Kleinbuchstaben für eine geringere Konservierung.

Ein zweites Beispiel lieferte die Suche nach dem Sigmafaktor D ( $\sigma^D$ ) Bindemotiv.  $\sigma^D$  ist in *B. subtilis* in die Regulation eines Regulons der späten Flagellar-Gene (Helmann et al., 1988), einiger Chemotaxisproteine und einem Großteil der vegetativen Autolysine (Lazarevic et al., 1992; Margot et al., 1994; Margot et al., 1999; Marquez et al., 1990) involviert. Das Bindemotiv dieses Sigmafaktors (Helmann and Moran Jr., 2002) konnte ebenfalls vor einigen Genen in *B. licheniformis* identifiziert werden (Tabelle 3.9.): yvyF und yvyC (Funktion jeweils unbekannt, Ähnlichkeit zu Flagellarproteinen), hag (Flagellinprotein), mcpA (Chemotaxispro-

<sup>&</sup>lt;sup>b</sup> Die angegebene Zahl steht für die Anzahl der Basenpaare von der ersten Base hinter dem Consensus bis zur ersten Base des Startcodons.

<sup>&</sup>lt;sup>c</sup> Großbuchstaben stehen für eine hohe, Kleinbuchstaben für eine geringere Konservierung.

tein), *lytD* (N-Acetylglucosa-minidase, Autolysin), *cheV* (Modulation der Aktivität von *CheA*, Chemotaxis) und *motA* (Motility Protein).

Erwähnt sei an dieser Stelle ebenfalls die Untersuchung des Replikationsursprungs in Kapitel 3.3.2.. Auch hier konnten hohe Übereinstimmungen des DnaA Bindemotives und der Anordnung von insgesamt 8 Bindestellen in 3 Boxen von *B. licheniformis* und *B. subtilis* festgestellt werden.

**Tabelle 3.9.:** Sigmafaktor D ( $\sigma^D$ ) (BLi01868) Bindemotiv

Gen	Promotorsequenz <sup>a</sup>	5' UTR <sup>b</sup>
<i>yvyF</i> (BLi03788)	ttagaagc <b>taaa</b> tcattttt-cttccct <b>gccgatat</b> aaaaa	35
hag (BLi03780)	ttttgcta <b>taaa</b> caatttca-aggcacc <b>tccgatat</b> aaaat	87
yvyC (BLi03779)	aagatatt <b>taat</b> ctcattcgccctta <b>gccgatat</b> aagtt	35
mcpA (BLi03295)	aatgcata <b>taaa</b> aacttttaaaagaa <b>tccgatat</b> taggg	36
<i>lytD</i> (BLi03821)	ttttattg <b>taaa</b> aaaataat-ccggctc <b>gccgataa</b> aaaga	57
cheV (BLi01614	ccggagca <b>tcaa</b> cttttaac-aaaagac <b>gccgatat</b> aaaaa	38
motA (BLi01524)	aaacagcc <b>taaa</b> gttcctatcacgcaac <b>accgatat</b> taacg	39
B. licheniformis		
Consensus <sup>c</sup>	<b>-TaAa</b> N (14-16) <b>gCCGATA</b> t	35-87
B. subtilis		
Consensus <sup>c</sup>	gCCGATAt	

<sup>&</sup>lt;sup>a</sup> Die Basen der Consensussequenz sind fett hervorgehoben.

<sup>&</sup>lt;sup>b</sup> Die angegebene Zahl steht für die Anzahl der Basenpaare von der ersten Base hinter dem Consensus bis zur ersten Base des Startcodons.

<sup>&</sup>lt;sup>c</sup> Großbuchstaben stehen für eine hohe, Kleinbuchstaben für eine geringere Konservierung.

# 3.5. Genomvergleich von *Bacillus licheniformis* mit anderen Bacilli der *B. subtilis*-Gruppe II

#### 3.5.1. Bacilli der Bacillus subtilis Gruppe II

*B. licheniformis* gehört gemeinsam mit anderen bekannten Vertretern der Bacilli, deren vollständige Genomsequenz ebenfalls vorliegt, zur *B. subtilis* Gruppe II. Dabei handelt es sich um *B. halodurans* (Takami et al., 2000), *B. cereus* ATCC 10987 (Ivanova et al., 2003), *B. anthracis* Ames (Read et al., 2003) und natürlich *B. subtilis* (Kunst et al., 1997). Bacilli der *B. subtilis* Gruppe II zeichnen sich durch die Produktion von Säuren bei Wachstum auf diversen Zuckern aus und bilden ovale, zentral oder subterminal in der Mutterzelle lokalisierte Endosporen aus. *B. halodurans* ist ein marines alkaliphiles Bakterium, welches Augrund seiner Fähigkeit der β-Galactosidase- und Xylanase-Produktion nicht nur von wissenschaftlichem, sondern auch von industriellem Interesse ist. Um einen pathogenen Vertreter der Spezies *Bacillus* handelt es sich bei *B. cereus* ATCC 10987. Dieses Bakterium ist häufig auf Getreideprodukten zu finden und einige Stämme führen durch die Produktion eines Exotoxins zu Lebensmittelvergiftungen. Einer der wohl bekanntesten pathogenen Bacilli ist *B. anthracis*, der Erreger des Milzbrandes.

#### 3.5.2. Vergleich der Genomdaten

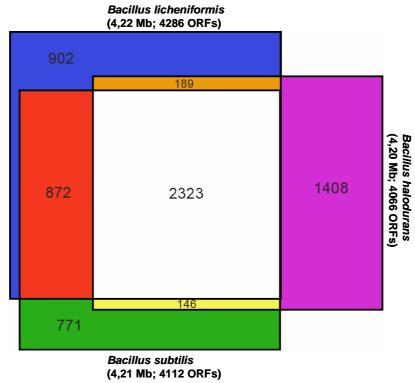
In diesem Kapitel soll eine erste Gegenüberstellung der Genomdaten der phylogenetisch nahverwandten Bacilli *B. licheniformis*, *B. subtilis*, *B. halodurans*, *B. cereus* ATCC 10987 und *B. anthracis* Ames erfolgen (Tabelle 3.10.). Bereits mit einem ersten Blick auf die allgemeinen Genomdaten kann eine Unterteilung dieser 5 Organismen in zwei Gruppen, die *subtilis/licheniformis/halodurans*-Gruppe und die *cereus/anthracis*-Gruppe, vorgenommen werden. Die beiden letzteren verfügen jeweils über ein (*B. cereus* ATCC 10987) bzw. zwei (*B. anthracis* Ames ) Plasmide, auf welchen im Fall von *B. anthracis* Ames der Hauptanteil der Virulenzgene lokalisiert ist (Okinaka et al., 1999a; Okinaka et al., 1999b). Hinsichtlich der Chromosomengröße weisen *B. licheniformis*, *B. subtilis*, und *B. halodurans* ein etwa eine Megabase kleineres Genom auf. Dies spiegelt sich auch in der Anzahl der ORFs, welche bei *B. licheniformis*, *B. subtilis* und *B. halodurans* mit ca. 4000 Genen von den etwa 5500 in *B. cereus* ATCC 10987 und *B. anthracis* Ames codierten Genen abweicht. Der durchschnittliche codierende Anteil der Sequenz des Chromosoms liegt bei allen fünf Organismen zwischen 84,3% (*B. anthracis* Ames) und 87,9% (*B. licheniformis*). Hinsichtlich des G+C Gehaltes

	B. licheniformis	B. subtilis	B. halodurans	B. cereus ATCC 10987	.CC 10987	B. an	B. anthracis Ames	
	Chromosom	Chromosom	Chromosom	Chromosom	Plasmid pBc10987	Chromosom	Plasmid pXO1	Plasmid pXO2
Größe [bp]	4.222.748	4.214.810	4.202.353	5.224.283	208.369	5.227.293	181.677	94.829
Anzahl der Gene	4.286	4.112	4.066	5.642	242	5.508	217	113
codierende Sequenz [%]	87,9	87,0	85,0	85,0	80,9	84,3	77,1	76,2
G+C Gehalt [%]	46,2	43,5	43,7	35,6	33,5	35,4	32,5	33,0
rRNA Operons	7	10	ω	12	0	1	0	0
tRNA Gene	72	98	28	86	0	95	0	0
Referenz	(Veith et al., 2004)	(Kunst et al., 1997)	(Takami et al., 2000)	(Ivanova et al., 2003)	al., 2003)	(Reac	(Read et al., 2003)	

lässt sich wieder eine Teilung in die oben genannten Gruppen beobachten. Hier weist *B. li-cheniformis* mit 46,2% den höchsten Anteil im Vergleich zu *B. subtilis* und *B. halodurans* mit 43,5% bzw. 43,7% auf. Der G+C Gehalt von *B. cereus* ATCC 10987 und *B. anthracis* Ames hingegen ist mit 35,6% bzw. 35,4% deutlich geringer. Im Vergleich zu den anderen vier Bacilli konnte in *B. licheniformis* die geringste Anzahl an Clustern ribosomaler RNA (7) und tRNA Genen (72) identifiziert werden.

## 3.5.3. Vergleich der Proteome von *B. licheniformis*, *B. subtilis* und *B. halodurans*

Aufgrund der hohen Übereinstimmungen der allgemeinen Genomdaten (3.5.2.) erfolgte eine weiterführende komparative Analyse mittels BLAST für *B. licheniformis*, *B. subtilis* und *B. halodurans* auf der Ebene der *open reading frames* (ORFs) (2.14.5.) Das Ergebnis ist in Abbildung 3.9. in Form eines Schnittmengendiagrammes dargestellt. Als Grenzwert für die Bestimmung orthologer Gene wurde der BLAST *e-value* kleiner oder gleich 10<sup>-15</sup> gewählt.



**Abbildung 3.9.:** Schnittmengendiagram der vergleichenden Genomanalyse von *B. licheniformis*, *B. subtilis* und *B. halodurans*.

weiß - gemeinsame ORFs von B. licheniformis, B. subtilis und B. halodurans

orange - gemeinsame ORFs von *B. licheniformis* und *B. halodurans* gelb - gemeinsame ORFs von *B. halodurans* und *B. subtilis* rot - gemeinsame ORFs von *B. subtilis* und *B. licheniformis* 

ORFs von B. licheniformis ohne Homolog in B. subtilis oder B. halodurans
 ORFs von B. halodurans ohne Homolog in B. licheniformis oder B. subtilis
 ORFs von B. subtilis ohne Homolog in B. halodurans oder B. licheniformis

Aus der Analyse ging hervor, dass diesen drei Bacilli ein Kerngenom von 2.323 Genen gemeinsam ist. Die Schnittmengen der Gene, welche *B. licheniformis* und *B. halodurans* besitzen, aber kein Ortholog in *B. subtilis* aufweisen, bzw. welche *B. subtilis* und *B. halodurans* gemeinsam sind, aber kein Ortholog in *B. licheniformis* aufweisen, liegt bei lediglich 189 bzw. 146. *B. licheniformis* und *B. subtilis* hingegen haben weitere 872 Gene gemeinsam, für die keine homologe Entsprechung in *B. halodurans* zu finden ist. Möglicherweise handelt es sich hier um einen Hinweis auf einen höheren Verwandtschaftsgrad von *B. licheniformis* und *B. subtilis* im Vergleich zu *B. halodurans*. In dieser Dreiergruppe weist *B. subtilis* mit 771 Genen die geringste Anzahl an *unique* ORFs, d.h. jeweils ohne orthologes Gegenstück in *B. licheniformis* oder *B. halodurans*, auf. Für *B. licheniformis* sind 902 Gene und für *B. halodurans* 1.408 Gene *unique*. Eine Übersicht über die 1.091 Gene von *B. licheniformis* ohne eine Ortholog in *B. subtilis* ist in Kapitel 3.4.1.2. gegeben.

#### 3.6. Wachstum von Bacillus licheniformis in kontinuierlicher Kultur

Im Vorfeld der Transkriptionsanalyse galt es ein Verfahren für die Anzucht von *B. licheniformis* zu finden, welches eine höchstmögliche Reproduzierbarkeit der das Wachstum beeinflussenden Faktoren erlaubt. Eine optimale Anzucht in diesem Sinne ermöglicht das kontinuierliche Wachstum im Chemostat (2.3.3.). Mittels einer konstanten Verdünnungsrate, bestimmt durch ein konstantes Volumen und einen definierten Zufluss frischen Mediums, wird eine Milieuveränderung durch die Anhäufung von Stoffwechselprodukten verhindert. Weiterhin erfolgt eine ständige Ausdünnung synthetisierter Proteine, was eine permanente Transkription und Synthese der benötigten Proteine zur Folge hat. Andere Regulationsmechanismen, wie etwa allosterische Regulationen, sollten unter diesen Bedingungen keine allzu hohe Relevanz besitzen. Des Weiteren können die Zellen auf eine konstante Wachstumsrate in Abhängigkeit von der Verdünnungsrate eingestellt werden. Die Zellen einer kontinuierlichen Kultur befinden sich stetig in der logarithmischen Wachstumsphase und alle entscheidenden Regulationsmechanismen werden auf Transkriptionsebene aktiv.

Bevor das Wachstum von *B. licheniformis* im Chemostat etabliert werden konnte, erfolgten Vorversuche hinsichtlich der optimalen Wachstumstemperatur und geeigneter C-Quellen.

# 3.6.1. Optimierung des Wachstums von *B. licheniformis* und die Verwertung verschiedener C-Quellen

Zur Optimierung der Wachstumsparameter Temperatur und pH-Wert wurden für *B. licheniformis* DSM13 die Wachstumsraten in einem Temperaturbereich von 32 - 52°C und einem Bereich von pH 7 - 8 bestimmt. Aufgrund der ermittelten Daten (nicht dargestellt) erfolgte im Weiteren eine optimierte Anzucht von *B. licheniformis* bei 42°C und pH 7,5.

Als potentielle C-Quellen für die Anzucht im Chemostat wurden Glucose, Gluconat, Xylose, 2,3-Butandiol, Acetat und Citrat, jeweils aerob, und Glucose anaerob getestet und die Wachstumsraten ermittelt (2.3.2.). Die Konzentrationen der verwendeten C-Quellen und die ermittelten Wachstumsraten sind in Tabelle 3.11. dargestellt.

Tabelle 3.11.: Das Wachstum von B. licheniformis auf unterschiedlichen C-Quellen

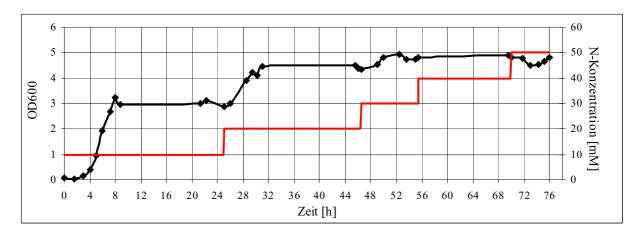
C-Quelle	Konzentration im	Wachstum	Wachstumsrate
	Minimalmedium (2.2.)		μ [1/h]
Glucose	30 mM	+	0,95
Gluconat	30 mM	+	0,68
Xylose	30 mM	+	0,30
2,3-Butandiol	30 mM	+	0,15
Acetat	60 mM	+	0,13
Glucose anaerob	30 mM	+	0,045
Citrat	30 mM	+	0,50

### 3.6.2. Limitierung des Wachstums durch C- und N-Quelle im Chemostat

Für die Stabilität einer kontinuierlichen Kultur ist die Limitierung eines das Wachstum beeinflussenden Faktors notwendig. Wahlweise kann das die Stickstoff-, Kohlenstoff-, Phosphatoder Schwefelquelle sein. Da für die Anzucht von *B. licheniformis* mit einem Phosphat gepufferten Minimalmedium gearbeitet wurde, entfiel eine Limitierung mit Phosphat. Aufgrund dessen wurden für die N- und C-Quellen die das Wachstum limitierenden Konzentrationen ermittelt. Die Bestimmung erfolgte durch die Messung der Zelldichte bei Wachstum im Chemostat unter sukzessiver Erhöhung der Konzentration der jeweiligen Energiequelle (2.3.4.). Für die C-Quellen Glucose und Gluconat wurde die Verdünnungsrate D = 0,55/h gewählt, womit sie jeweils zwischen der maximalen und der ½ maximalen Wachstumsrate liegt (siehe

Tabelle 3.11.). Für die Untersuchung der Wachstumslimitierung durch die Xylose-Konzentration wurde aufgrund der niedrigen Wachstumsrate die Verdünnungsrate D=0.21/h gesetzt.

Beispielhaft sind in Abbildung 3.10. die experimentellen Daten der Abhängigkeit der Zell-dichte von der N-Konzentration über der Zeit dargestellt.



**Abbildung 3.10.:** Abhängigkeit der Zelldichte von der N-Konzentration. Veränderung der Zelldichte, schwarz; N-Konzentration, rot.

Basierend auf den Daten aus Abbildung 3.10. ist der Einfluss der N-Konzentration auf die Zelldichte in Abbildung 3.11. dargestellt. Bis zu einer 30 mM Konzentration von NH<sub>4</sub>Cl erfolgte eine stetige Zunahme der Zellmasse. Bei einer weiteren Erhöhung der N-Konzentration blieb die Zelldichte konstant, das Wachstum der Zellen war nicht mehr durch die N-Konzentration limitiert.

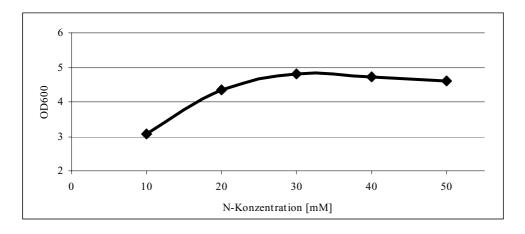


Abbildung 3.11.: Abhängigkeit der Zelldichte von der N-Konzentration.

Analog zur Untersuchung der Wachstumslimitierung durch Stickstoff wurde der Einfluss der Glucose-, Xylose- und Gluconat-Konzentration auf das Zellwachstum ermittelt. In Tabelle 3.12. sind der Schwellenwert für limitiertes und unlimitiertes Wachstum und die letztendlich eingesetzte Konzentration der einzelnen Energiequellen aufgeführt.

**Tabelle 3.12.:** Limitierung des Wachstums im Chemostat durch die Konzentration der C- oder N-Quelle.

Energie-Quelle	Glucose	Gluconat	NH <sub>4</sub> Cl	Xylose
Schwellenwert für limitiertes/unlimitiertes Wachstum	20 mM	35 mM	30 mM	50 mM
Konzentration für limitiertes Wachstum	10 mM	10 mM	10 mM	30 mM
Konzentration für unlimitiertes Wachstum	50 mM	50 mM	40 mM	66,24 mM

•

## 3.6.3. Anzucht von *B. licheniformis* mit unterschiedlichen C-Quellen im Chemostat

Für eine RNA-Präparation und anschließende Transkriptionsanalysen wurde *B. licheniformis* DSM13 in kontinuierlicher Kultur unter unterschiedlichen Bedingungen angezogen:

- Glucose unlimitiert / NH<sub>4</sub>Cl limitiert
- Glucose limitiert / NH<sub>4</sub>Cl unlimitiert
- Gluconat unlimitiert / NH<sub>4</sub>Cl limitiert
- Xylose unlimitiert / NH<sub>4</sub>Cl limitiert

Für das Wachstum mit Glucose und Gluconat wurde eine durchschnittliche Verdünnungsrate D von 0,55/h analog zur Wachstumslimitierung (3.6.2.) eingestellt. Für Xylose wurde aufgrund der deutlich niedrigeren Wachstumsrate von 0,33/h eine entsprechend geringere Verdünnungsrate D von 0,21/h gewählt.

Für die RNA-Präparation (2.5.3.) der auf Glucose bzw. Gluconat gewachsenen Zellen, wurde der Chemostat, wie unter 2.3.3. beschrieben, angeimpft, das Fließgleichgewicht eingestellt und nach konstantem Wachstum über mindestens 3 Volumenwechsel die Zellen geerntet und aufgeschlossen (2.5.2.).

Die Anzucht von *B. licheniformis* im Chemostat unter Verwertung von Xylose zeigte sich problematisch. Bereits bei der Limitierungsuntersuchung für Xylose (3.6.2.) konnten auf den beimpften NB-Platten zur Reinheitskontrolle zwei unterschiedliche Zellmorphologien ausgemacht werden, zum einen klar abgegrenzte Kolonien und zum anderen schwärmende Kolo-

nien. Ein wiederholter Ausstrich auf Minimalmedium-Platten zeigte das gleiche Ergebnis, womit die Wahrscheinlichkeit, dass es sich um eine Kontamination handeln könnte, verringert wurde. Des Weiteren konnten mikroskopisch keine morphologischen Unterschiede beobachtet werden. Bei der Anzucht von *B. licheniformis* mit Xylose für eine RNA-Präparation zeigte sich, sowohl mikroskopisch als auch auf den NB- und Minimalmedium-Platten, dasselbe Bild. Zusätzlich stellte sich über einen Zeitraum von 43 h, was bei einer Verdünnungsrate D=0,21/h und einem Kulturvolumen von 350 ml neun Volumenwechseln entspricht, keine konstante Zelldichte ein. Stattdessen nahm die Zellmasse, entgegen dem Prinzip einer Kultur im Fließgleichgewicht, stetig zu. Diese Beobachtungen lassen vermuten, dass der erhöhte Selektionsdruck im Chemostat bei Wachstum mit Xylose über einen relativ langen Zeitraum eine Adaption der Zellen an diese Bedingungen zur Folge hat. Aufgrund dessen wurde auf eine RNA-Präparation unter diesen Wachstumsbedingungen verzichtet.

Da *B. licheniformis* bei Wachstum mit Acetat bzw. 2,3-Butandiol eine noch geringere maximale Wachstumsrate im Vergleich zu Xylose aufweist, wurde für diese beiden C-Quellen ebenfalls keine Anzucht in kontinuierlicher Kultur vorgenommen. Um Transkriptionsanalysen mit RNA von *B. licheniformis*, gewachsen mit einer dieser beiden C-Quellen, durchführen zu können, wurden die Zellen wie unter 2.3.1. beschrieben in 500 ml Batch-Kulturen angezogen, bei einer OD<sub>600</sub> von 0,5 in der mittleren logarithmischen Phase geerntet und die RNA präpariert.

## 3.7. Transkriptionsanalyse von Bacillus licheniformis DSM13

#### 3.7.1. Erstellen des Genomchips

Für die Erstellung des Genomchips von *B. licheniformis* wurden 300 - 500 bp große Fragmente jedes ORFs größer 300 bp amplifiziert. Die Primergenerierung, die Amplifizierung und das Spotten der *probes* erfolgten wie unter 2.15. beschrieben. Für ca. 96% aller durchgeführten PCRs konnte Plasmid-DNA der Sequenzierung von *B. licheniformis* (3.1.) als Template verwendet werden. Für die verbleibenden 4% wurde chromosomale DNA als Matrize eingesetzt. Unter Verwendung der Plasmidgenbank wurde des Weiteren für 76% aller PCRs nur ein spezifischer Primer benötigt, da als 2. Primer der Standard *forward*- oder *reverse*-Primer des Sequenzierungsprojektes genutzt werden konnte. Für die restlichen 24% der Chip-PCRs hingegen wurden zwei spezifische Primer generiert. Insgesamt sind 81,5% aller in *B. licheniformis* identifizierter ORFs auf dem Genomchip vertreten. Hierbei gilt es zu berücksichtigen, dass 14,1% aller 4286 *B. licheniformis* ORFs kleiner als 300 bp sind. Bei einer Mindestgröße der

*probes* von 300 bp konnten diese ORFs für die Erstellung des Genomchips nicht berücksichtigt werden. Somit wurden letztendlich 95% aller *B. licheniformis* ORFs größer als 300 bp auf den Genomchip gespottet.

# 3.7.2. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit Acetat

Für eine vergleichende Expressionsanalyse des Wachstums mit Glucose und mit Acetat wurde RNA, gewonnen aus jeweils zwei biologisch unabhängigen Experimenten, verwendet. Die Glucosekultur wurde im Chemostat mit einer durchschnittliche Wachstumsrate von 0,55/h angezogen, nach Einstellen des Fließgleichgewichtes geerntet (2.5.2.) und die RNA isoliert (2.5.3.). Aufgrund der im Vergleich sehr niedrigen Wachstumsrate von B. licheniformis mit Acetat als C-Quelle, erfolgte hier die Anzucht in einer 500 ml Batch-Kultur und die Zellernte in der mittleren logarithmischen Phase. Die Auswertung der Transkriptionsanalysen erfolgte anhand von mindestens 5 Hybridisierungen (2.16.3.), wobei eine Kombination der unterschiedlichen RNAs und für ein Hybridisierungsexperiment ein dye switch vorgenommen wurden (Tabelle 3.13.). Unter dye switch versteht man die Wiederholung eines Versuchs mit einer umgekehrten Markierung der eingesetzten RNA (2.16.1.). Die Quantifizierung, Normalisierung, Qualitätssicherung und Analyse der Transkriptionsdaten erfolgte wie unter 2.16.4. und 2.16.5. beschrieben. An dieser Stelle sei darauf hingewiesen, dass es sich bei sämtlichen Angaben von Expressionsdaten, basierend auf der DNA-Microarrayanalyse, um gemittelte Werte von mindestens 5 unabhängigen Hybridisierungen und nicht um absolute Werte handelt.

**Tabelle 3.13.:** Hybridisierungs- und Markierungsschema für die DNA-Microarrayanalyse Für die Hybridisierungen Nr. 1 und 2 wurde RNA des ersten biologisch unabhängigen Experimentes unter den Bedingungen A (RNA A1) mit Cy3 markiert und jeweils mit den RNAs des ersten und zweiten biologisch unabhängigen Experimentes unter den Bedingungen B (RNA B1; RNA B2), markiert mit Cy5, kombiniert. Für die RNA des zweiten biologischen Experimentes unter den Bedingungen A (RNA A2) erfolgte ebenfalls eine Hybridisierung (Nr. 3 und 4) in Kombination mit den RNAs der zwei Experimente unter den Bedingungen B. Die Hybridisierung Nr. 5 ist eine Wiederholung der Hybridisierung Nr. 4, wobei die Farbstoffe ausgetauscht wurden (*dye switch*).

Hybridisierung	Markierung mit Cy3	Markierung mit Cy5		
1	RNA A1	RNA B1		
2	RNA A1	RNA B2		
3	RNA A2	RNA B1		
4	RNA A2	RNA B2		
dye switch	RNA B2	RNA A2		

#### 3.7.2.1. Expression der Gene des C-Stoffwechsels

Für die Untersuchung der Gene des Kohlenstoff-Stoffwechsels wurde der Focus auf die ORFs gerichtet, deren Genprodukt in direktem funktionellem Zusammenhang mit der Glykolyse und dem Tricarbonsäurezyklus steht. Die Ergebnisse der Transkriptionsanalyse mittels DNA-Microarray-Technologie sind für die Glykolyse in Abbildung 3.12. und für den Tricarbonsäurezyklus in Abbildung 3.13. dargestellt (Angaben ,nicht fett'). Die in den Abbildungen 3.12. und 3.13. verwendeten Genkürzel sind in der Tabelle 3.14. erläutert. Wachsen B. licheniformis Zellen mit Acetat als C-Quelle wird im Vergleich zum Wachstum mit Glucose, die Genexpression der Glyceraldehyd-3-Phosphat-Dehydrogenase A (GapA), der Phosphoglycerat-Kinase (Pgk), der Phosphoglycerat-Mutase (Pgm), der Enolase (Eno) und der Triose-Isomerase (Tpi) um etwa den Faktor 2 gedrosselt (Abbildung 3.12.). Die Expression der Gene des Glyoxylsäurezyklus, der Isocitrat-Lyase und der Malat-Synthase, hingegen wird um einen Faktor von ca. 100 hoch reguliert (Abbildung 3.13.). Auf diese Weise werden Oxalacetat und Malat, welche durch die Phosphoenolpyruvat-Carboxykinase (PckA) zu Phosphoenolpyruvat und das potentielle Malat-Enzym (MalS) zu Pyruvat umgesetzt werden, für die Gluconeogenese bereitgestellt. Diese beiden Enzyme werden ebenfalls um einen Faktor von etwa 55 bzw. 4 höher exprimiert als bei Wachstum mit Glucose. Eine Erhöhung der Expression von GapB, der Glyceraldehyd-3-Phosphat-Dehydrogenase B, zur Umsetzung von 1,3-di-Phosphoglycerat zu Glyceraldehyd-3-Phosphat, um das 6-fache wurde ebenfalls detektiert (Abbildung 3.12.). Während des Wachstums mit Glucose ließ sich eine im Vergleich höhere Expression der Acetolactat-Synthase (AlsS) und der Acetolactat- Decarboxylase (AlsD), um einen Faktor 5, bzw. 2, zur Synthese von Acetoin ermitteln (Abbildung 2.13.). Dient hingegen Acetat als C-Quelle, erfolgt eine höhere Expression der Acetoin-Dehydrogenasen AcoABC und AcuABC zur reversiblen Synthese von Acetyl-CoA aus Acetoin. Für die Gene acoABC konnte eine Expressionserhöhung um etwa das 6fache, für acuABC eine Erhöhung um den Faktor 8 ermittelt werden. Essentiell für die Verwertung von Acetat ist die Acetyl-CoA-Synthetase (AcsA). Für das Gen acsA wurde eine Erhöhung der Expression von ca. 17 detektiert. Insgesamt bestätigt die beobachtete Veränderung des Expressionsmusters der Gene von Glykolyse und Tricarbonsäurezyklus bei Wachstum mit Acetat im Vergleich zum Wachstum mit Glucose die Erwartungen.

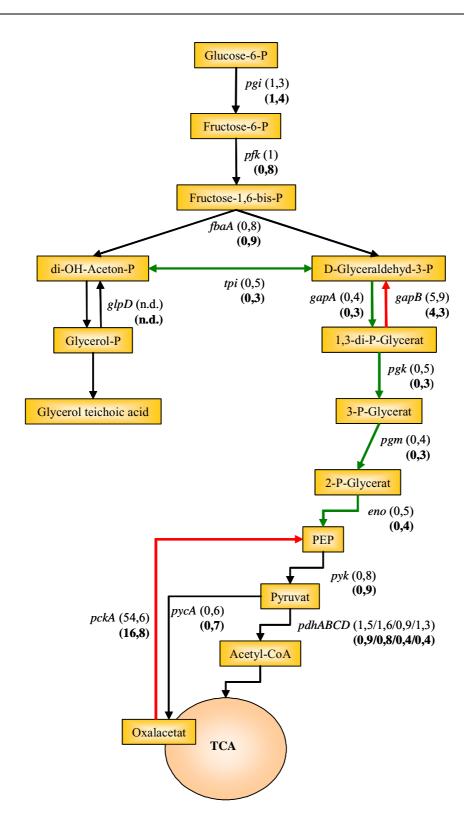
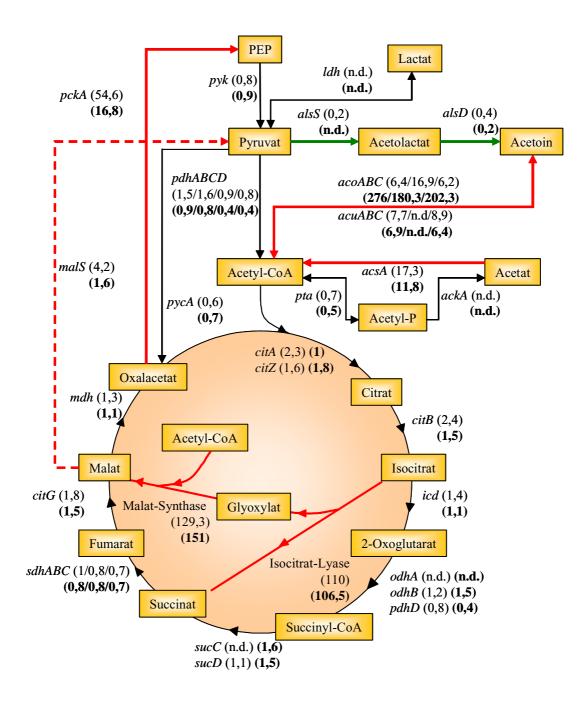


Abbildung 3.12.: Schematische Darstellung der Glykolyse unter Angabe der Expressionsverhältnisse der beteiligten Gene bei Wachstum auf Acetat, bzw. Butandiol im Vergleich zum Wachstum auf Glucose: Für jedes Gen ist in Klammern "nicht fett" das Expressionsverhältnis Acetat/Glucose angegeben. Fett gedruckte Werte zeigen das Expressionsverhältnis von 2,3-Butandiol/Glucose. Grüne Pfeile stellen eine deutlich höhere Genexpression bei Verwertung von Glucose im Vergleich zu Acetat bzw. 2,3-Butandiol dar, rote Pfeile eine entsprechend höhere Genexpression bei Wachstum mit Acetat oder 2,3-Butandiol. Die Funktionen der einzelnen Gene sind in Tabelle 3.14. angegeben. n.d. – nicht bestimmt



**Abbildung 3.13.:** Schematische Darstellung des Tricarbonsäurezyklus unter Angabe der Expressionsverhältnisse der beteiligten Gene bei Wachstum auf Acetat, bzw. Butandiol im Vergleich zum Wachstum auf Glucose. Für jedes Gen ist in Klammern 'nicht fett' das Expressionsverhältnis Acetat/Glucose angegeben. Fett gedruckte Werte zeigen das Expressionsverhältnis von 2,3-Butandiol/Glucose. Grüne Pfeile stellen eine deutlich höhere Genexpression bei Verwertung von Glucose im Vergleich zu Acetat bzw. 2,3-Butandiol dar, rote Pfeile eine entsprechend höhere Genexpression bei Wachstum mit Acetat oder 2,3-Butandiol. Die Funktionen der einzelnen Gene sind in Tabelle 3.14. angegeben. n.d. – nicht bestimmt

**Tabelle 3.14.:** Genkürzel und Funktionen der in den Abbildungen 3.12., 3.13., 3.14 und 3.15. dargestellten Gene der Glykolyse bzw. des Tricarbonsäurezyklus

Gen	Funktion
ackA	Acetat-Kinase
acoABC	Untereinheiten der Aco Acetoin-Dehydrogenase
acs	Acetyl-CoA-Synthetase
acuABC	Untereinheiten der Acu Acetoin-Dehydrogenase
alsD	Acetolactat-Decarboxylase
alsS	Acetolactat-Synthase
citA	Citrat-synthase I (minor)
citB	Aconitat-Hydratase
citG	Fumarat-Hydratase
citZ	Citrat-synthase II (major)
eno	Enolase
fbaA	Fructose-1,6-bisphosphat-Aldolase
gapA	Glyceraldehyde-3-Phosphat-Dehydrogenase
gapB	Glyceraldehyde-3-Phosphat-Dehydrogenase
glpD	Glycerol-3-Phosphat-Dehydrogenase
icd	Isocitrat-Dehydrogenase
ldh	Lactat-Dehydrogenase
malS	potentielles Malat-Enzym
mdh	Malat-Dehydrogenase
odhAB	Untereinheiten der 2-Oxoglutarat-Dehydrogenase
pckA	Phosphoenolpyruvat-Carboxykinase
pdhABC	Untereinheiten der Pyruvat-Dehydrogenase
pdhD	Untereinheit der Pyruvat-Dehydrogenase / Untereinheit der 2-Oxoglutarat-
	Dehydrogenase
pfk	Fructose-6-Phosphat-Kinase
pgi	Glucose-6-Phosphat-Isomerase
pgk	Phosphoglycerat-Kinase
pgm	Phosphoglycerat-Mutase
pta	Phosphotransacetylase
pycA	Pyruvat-Carboxylase
pyk	Pyruvat-Kinase
sdhABC	Untereinheiten der Succinat-Dehydrogenase
sucCD	Untereinheiten der Succinyl-CoA-Synthetase
tpi	Triosephosphat-Isomerase

## 3.7.2.2. Expressionsanalyse der Isocitrat-Lyase und der Malat-Synthase mittels *Real-time* PCR

Das besondere Augenmerk während der Transkriptionsanalyse von *B. licheniformis* Zellen, angezogen mit Acetat im Vergleich zur Anzucht mit Glucose, lag auf dem Glyoxylsäurezyklus. Die Analyse der Genexpression mittels DNA-Microarray-Technologie hatte bereits qualitativ eine deutliche Erhöhung der Expression der Isocitrat-Lyase und der Malat-Synthase aufgezeigt (3.7.2.1.). Unter Einsatz der *Real-time* PCR (2.7.4.) erfolgte eine weitere Untersuchung zur Quantifizierung der Transkriptionsänderung der beiden Gene. Für diese Analyse wurde ebenfalls die RNA aus *B. licheniformis* Zellen angezogen auf Acetat, bzw. auf Glucose aus jeweils zwei unabhängigen biologischen Experimenten verwendet. Für jede RNA wurden 6 konstitutive Gene mitgeführt (Daten nicht angegeben). Nach Ermittlung der C<sub>T</sub>-Werte, erfolgte mit Hilfe der unter 2.7.4. beschriebenen Formel, die Berechnung der Expressionsfaktoren. In Tabelle 3.15. sind die Mittelwerte der C<sub>T</sub>-Werte jedes Gens für die Acetat- und Glucose-RNA und die sich daraus ergebenden Expressionsfaktoren, im Vergleich zu den DNA-Microarray-Daten, angegeben.

**Tabelle 3.15.:** Expressionsfaktoren für die Isocitrat-Lyase und die Malat-Synthase bei Wachstum von *B. licheniformis* mit Acetat im Vergleich zum Wachstum mit Glucose ermittelt durch DNA-Microarray-Technologie und *Real-time* PCR.

ORF	Funktion	Expressionsfaktor	MW C <sub>T</sub>	SD	MW C <sub>T</sub>	SD	Expressionsfaktor
		Acetat/Glucose	Acetat	[%]	Glucose	[%]	Acetat/Glucose
		DNA-Microarray					Real-time PCR
BLi04207	Isocitrat-	110,3	10,3	7,3	18,4	0,3	305,2
	Lyase						
BLi04208	Malat-	129,3	10,7	7,4	18,6	3,7	265,7
	Synthase						

Im Rahmen dieses Versuches zeigte sich eine hohe Reproduzierbarkeit der einzelnen *Real-time* PCR-Experimente mit unterschiedlichen RNAs, wie die Standardabweichungen der C<sub>T</sub>-Werte (Tabelle 3.15.) belegen. Die Transkriptionsanalyse ergab sowohl mittels DNA-Microarray Technologie, als auch mittels *Real-time* PCR eine Regulation der beiden untersuchten Gene im 100fachen Bereich. Für die Isocitrat-Lyase konnte im Rahmen der quantitativen *Real-time* PCR eine 305fache Expressionserhöhung bei Wachstum mit Acetat im Vergleich zum Wachstum mit Glucose bestimmt werden und für die Malat-Synthase eine Erhöhung um den Faktor 266. Die Expressionswerte, ermittelt über die qualitative DNA-

Microarray Analyse, weichen um einen Faktor 2 - 2,8 von den über die quantitative *Real-time* PCR bestimmten Expressionsfaktoren ab.

#### 3.7.2.3. Expressionsdaten anderer auffällig regulierter Enzyme

Aufgrund der mittels DNA-Microarray-Analyse erstellten Expressionsdaten konnten insgesamt weitere 105 Gene ermittelt werden, deren Expression sich unter den gegebenen Bedingungen um mindestens den Faktor 3 ändert. Diese Gene sind mit den jeweils detektierten Expressionsfaktoren in Tabelle A.1. im Anhang dargestellt. Darunter befinden sich 41 ,Y'-ORFs und 13 putative Gene. Den verbleibenden 52 Genen konnte eine Funktion zugewiesen werden.

Unter Verwertung von Glucose ist auffällig, dass mehrere Gene, welche in die Purin-Biosynthese involviert sind (purA, purB, purC, purK, purQ, yabJ, xpt), höher exprimiert werden. Für die Gene pyrP und pyrAA, die an der Pyrimidin-Biosynthese beteiligt sind, wurde ebenfalls eine Expressionserhöhung detektiert. Weitere Gene, welche im Vergleich zum Wachstum mit Acetat erhöht transkribiert werden, sind die der vegetativen Katalase 1 (katA), der Katalase 2 (katE) und der Alkylhyperoxid-Reduktase (ahpCF). Die Funktion dieser Genprodukte liegt in der Detoxifikation der Zelle. Die Transkription der Gene des DNAbindenden Stress-Proteins MrgA und des Hitzeschock-Proteins GrpE wird ebenfalls erhöht. Steht der Zelle ausschließlich Acetat als C-Quelle zur Verfügung, so ist gleichfalls eine deutliche Erhöhung der Expression einer Vielzahl von Genen zu beobachten. Das gilt zum Beispiel für Gene, deren Produkte in die Differenzierung der Endospore bzw. ihrer Auskeimung involviert sind, wie gerPE und gerPC (Keimung der Spore), cotXYZ (Proteine der Sporenhülle), CoxA (Cortexprotein der Spore) und cwlC (N-Acetylmuramoyl-L-Alanine-Amidase cwlL Vorläufer). Die gleichzeitige Höherexpression dieser Gene ist verwunderlich, da es sich um gegenläufige Prozesse in der Zelle handelt. Es konnte jedoch für die genannten ORFs eine Erhöhung der Transkription in einem Bereich von 70 - 200 detektiert werden, ausgenommen coxA mit einer Erhöhung der Expression um den Faktor 10. Weitere Proteine, deren Gene erhöht transkribiert werden, sind MelA (α-D-Galactosid-Galactohydrolase), AmyD (involviert in den Zuckertransport), Kbl (2-Amino-3-Ketobutyrat-CoA-Ligase), AprX (intrazellulare alkalische Serin-Protease), FhuD und FhuC (Ferrichrom ABC-Transporter), EftA und EftB (α- und β-Untereinheiten des Elektronentransfer-Flavoproteins), MmgA (Acyl-CoA-Dehydrogenase) und MmgC (Acetyl-CoA-Acetyltransferase), CydC und CydD (ABC-Transporter für die Expression von Cytochrom bd), sowie CydB (Cytochrome bd Ubiquinol-Oxidase (Subunit II)).

# 3.7.3. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit 2,3-Butandiol

Die Transkriptionsanalyse von *B. licheniformis* für den Vergleich des Wachstums mit 2,3-Butandiol und Glucose wurde analog zur Untersuchung des Wachstums mit Acetat und Glucose vorgenommen. Hier wurde ebenfalls mit der RNA aus jeweils zwei biologisch unabhängigen Experimenten gearbeitet. Aufgrund der im Verhältnis niedrigen Wachstumsrate von *B. licheniformis* unter Verwertung von 2,3-Butandiol, wurde die RNA aus der mittleren logarithmischen Phase einer Batchkultur gewonnen. Bezüglich der Hybridisierung und Datenauswertung wurde wie bei der vergleichenden Expressionsanalyse des Wachstums mit Glucose und Acetat (3.7.2.) verfahren.

#### 3.7.3.1. Expression der Gene des C-Stoffwechsels

Bei der Untersuchung der Genexpression des C-Stoffwechsels wurde, analog zur Untersuchung des Wachstums mit Acetat, der Schwerpunkt auf die Gene gelegt, deren Produkte in die zentralen Stoffwechselwege der Glykolyse und des Tricarbonsäurezyklus involviert sind. Die auf der DNA-Microarray Analyse basierenden Expressionsfaktoren der entsprechenden Enzyme sind in fett gedruckten Lettern in den Abbildungen 3.12. und 3.13. wiedergegeben. Die in diesen Abbildungen verwendeten Genkürzel sind in der Tabelle 3.14. erläutert. Für die Verwertung von 2,3-Butandiol zeigte sich ein Expressionsmuster, ähnlich zur Verwertung von Acetat. Im Vergleich zum Wachstum mit Glucose findet auch hier eine Repression der Genexpression der Glyceraldehyd-3-Phosphat-Dehydrogenase A (GapA), der Phosphoglycerat-Kinase (Pgk), der Phosphoglycerat-Mutase (Pgm), der Enolase (Eno) und der Triose-Isomerase (Tpi) um etwa den Faktor 2,5 - 3 statt (Abbildung 3.12.). Die Expression der alternativen Glyceraldehyd-3-Phosphat-Dehydrogenase B (GapB) wird ebenfalls um den Faktor 4 erhöht. Die Funktion dieses Enzyms liegt in der Umkehrung der von GapA katalysierten Umsetzung von Glyceraldehyd-3-Phosphat zu 1,3-di-Phosphoglycerat, um die Gluconeogenese von Mehrfachzuckern zu gewährleisten. Für die Gene des Glyoxylsäurezyklus, die Isocitrat-Lyase und die Malat-Synthase konnte eine Expressionserhöhung um das 100, bzw. 150fache detektiert werden (Abbildung 3.13.). Das Gen der Phosphoenolpyruvat-Carboxykinase (PckA), zuständig für die Bereitstellung von Oxalacetat für die Gluconeogenese, wird, im Vergleich zur Nutzung von Glucose, um einen Faktor 17 höher transkribiert. Für das potentielle Malat-Enzym MalS hingegen konnte keine Änderung der Expression, wie bei der Verwertung von Acetat, detektiert werden. Die Expression der Acetolactat-Decarboxylase (AlsD) wird bei Nutzung von 2,3-Butandiol, analog zur Nutzung von Acetat, um einen Faktor 3 reprimiert. Offensichtlich wird der *overflow* Metabolismus zur Synthese von Acetoin, betrieben während des Wachstums mit Glucose, eingestellt. Für die Acetolactat-Synthase (AlsS) konnten keine, den qualitativen Maßstäben entsprechenden, Daten ermittelt werden. Entsprechend der Verwertung von Acetat als C-Quelle wird unter Nutzung von 2,3-Butandiol die Expression der Acetoin-Dehydrogenasen AcoABC und AcuABC erhöht. Für die Gene *acu*ABC wurde eine Expressionserhöhung um den Faktor 6 ermittelt, für die *aco*ABC Gene hingegen eine etwa 200fache Expressionsteigerung. Die Expression der Acetyl-CoA-Synthetase (AcsA) wird ungefähr um das 12fache erhöht.

#### 3.7.3.2. Expressionsdaten anderer auffällig regulierter Enzyme

Neben den unter 3.7.3.1. erwähnten Genen konnte für weitere 177 Gene eine mindestens 3fache Expressionsänderung detektiert werden. Darunter befinden sich 61 so genannte ,Y-ORFs' und die Gene von 24 putativen Proteinen. In Tabelle A.2. im Anhang sind diese ORFs mit ihren ermittelten Expressionsdaten unter den gegebenen Bedingungen erfasst. Bei Wachstum von B. licheniformis mit 2,3-Butandiol im Vergleich zum Wachstum mit Glucose zeigt sich, analog zur Transkriptionsanalyse von mit Acetat gewachsenen Zellen, eine Expressionsreprimierung einiger in die Purin-Biosynthese involvierter Enzyme (PurABCDEFHKLMNQ, YabJ, Xpt). Für einige Gene der Pyrimidin-Biosynthese (pyrAA, pyrAB, pyrK, pyrP) konnte ebenfalls eine Drosselung der Genexpression detektiert werden. Des Weiteren zeigte sich unter der Verwertung von Glucose eine erhöhte Expression der Glutamat-Synthase (GltA und GltB) und einiger Gene, deren Produkte an der Histidin-Biosynthese beteiligt sind (hisF, hisG, hisH). Für die Alkylhyperoxid-Reduktase (ahpCF) und die vegetative Katalase 1 (katA) wurde unter Verwertung von Glucose ebenfalls eine höhere Expression als bei der Nutzung von 2,3-Butandiol gemessen. Auffällig ist, dass mit Glucose wachsende Zellen Gene, deren Produkte in den Transport von Mangan (mntA, mntB, mntD und mntH) und Glycin/Betain/Arnitin und Cholin (opuCA und opuCC) involviert sind, erhöht exprimieren. Weitere Gene, deren Transkription unter Verwertung von Glucose induziert wird, sind nasC (katalytische Untereinheit der assimilatorischen Nitrat-Reduktase), nasE (Untereinheit der assimilatorischen Nitrit-Reduktase), nrgB (Stickstoff reguliertes PII-ähnliches Protein), folD (involviert in die Folat-Biosynthese), glnR (Repressor der Transkription der Glutamin-Synthetasegene), perR (Repressor der Transkription des Peroxidregulons), pucR (Regulator der puc Gene), serA (Phosphoglycerat-Dehydrogenase), ctaF (Untereinheit IV der Cytochrom caa3 Oxidase), cypX (Cytochrom P450-ähnliches Enzym) und qoxD (Untereinheit IV der Cytochrom aa3 Quinol-Oxidase).

B. licheniformis Zellen, denen ausschließlich 2,3-Butandiol als C-Quelle zur Verfügung steht, erhöhen die Expression einer Vielzahl von Genen im Vergleich zum Wachstum mit Glucose. Dazu gehören, wie auch beim Wachstum mit Acetat, amyD (involviert in den Zuckertransport), cotXYZ (Proteine der Sporenhülle), gerPE und gerPC (Keimung der Spore), cwlC (N-Acetylmuramoyl-L-Alanin-Amidase cwlL Vorläufer), cydC (ABC-Transporter für die Expression von Cytochrom bd), fhuD und fhuC (Ferrichrom ABC-Transporter), eftA und eftB (α- und β-Untereinheiten des Elektronentransfer-Flavoproteins), kbl (2-Amino-3-Ketobutyrat-CoA-Ligase) und melA (α-D-Galactosid-Galactohydrolase). Des Weiteren wird eine Vielzahl von Genen höher exprimiert, deren Genprodukte in den Lipid-Metabolismus involviert sind, wie die Acyl-CoA-Dehydrogenase (AcdA), die Leucin-Dehydrogenase (Bcd), die Untereinheiten der α-Ketosäure-Dehydrogenase (BkdAA, BkdAB, BkdB), die Butyrat-Kinase (Buk), die Membranphospholipid-Desaturase (Des), Enzyme der Glycerolverwertung (GlpD, GlpF,GlpQ), eine weitere potentielle α-Ketosäure-Dehydrogenase (LpdV) und eine potentielle Phosphatbutyryl-Transferase (Ptb). Eine Steigerung der Expression konnte gleichfalls für einige Gene, deren Produkte in die Arginin-Biosynthese involviert sind, detektiert werden, wie argB, argC, argD, argG, argH, argJ und carA. Weitere Gene, welche unter den gegebenen Bedingungen höher exprimiert werden, sind hxlR (positiver Regulator der hxlAB Gene, involviert in den Ribulose-P-Weg), ispA (Major intrazellulare Serin-Protease), licT (Antiterminator der bglPH Gene zum Celluloseabbau), mmgD (Citrat-Synthase III), pbpE (Penicillinbindendes Protein), pmi (Mannose-6-P-Isomerase), sigW (RNA-Polymerase Sigmafaktor W zur Regulation der Genexpression von Genen der Detoxifikation, wie z.B. pbpE), sipV (Signalpeptidase I) und das Gen, welches für das Subtilisin Carlsberg codiert. Wobei hier wohlmöglich ein Zusammenhang zwischen der Expression der Signalpeptidase SipV und dem Subtilisin Carlsberg bezüglich des Transportes aus der Zelle besteht und der Expression des Penicillin-bindenden Proteins PbpE zum Schutz der Zelle, induziert durch den Sigmafaktor W.

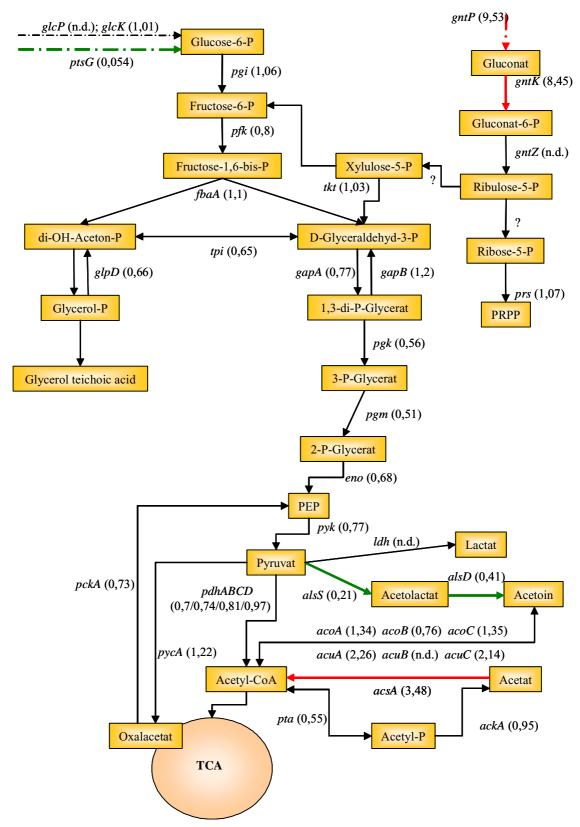
# 3.7.4. Genexpression bei Wachstum mit Glucose im Vergleich zum Wachstum mit Gluconat

Analog zu den Transkriptionsanalysen bezüglich des Wachstums mit Acetat, bzw. 2,3-Butandiol im Vergleich zum Wachstum mit Glucose, wurden für die vergleichende Untersuchung der Genexpression unter der Verwertung von Gluconat und Glucose RNAs aus jeweils zwei biologisch unabhängigen Experimenten verwendet. Die sich aus fünf Hybridisierungen des Genomchips von *B. licheniformis* ergebenden Expressionsdaten werden im Folgenden,

zum einen bezogen auf den Kohlenstoffstoffwechsel und zum anderen hinsichtlich aller weiteren auffällig regulierten Gene, dargestellt.

# 3.7.4.1. Expression der Gene des C-Stoffwechsels

Für die zentralen C-Stoffwechselwege zeigte sich eine Adaptation an die C-Quelle Gluconat (Abbildung 3.14). So wird die Expression des Gens ptsG, welches für das Glucosespezifische Enzym II des Phosphotransferase-Systems (PTS) codiert, um etwa das 20fache reduziert. Die Expression des Gens gntP, der Gluconat-Permease, hingegen wird um etwa das 10fache erhöht. Das Gen gntK, der Gluconat-Kinase, welche Gluconat in den Pentosephosphat-Weg einschleust, wird ebenfalls um das 8fache erhöht exprimiert. Mit Hilfe der Transketolase (Tkt) werden anschließend drei Moleküle Xylose-5-Phosphat zu zwei Molekülen Fructose-6-Phosphat und einem Molekül Glyceraldehyd-3-Phosphat umgeformt und für die Glykolyse bereitgestellt. Die Transketolase wird unter diesen Bedingungen offenbar konstitutiv exprimiert. Während des Wachstums mit Glucose betreiben die B. licheniformis Zellen offenbar einen overflow-Metabolismus, welcher sich durch eine erhöhte Expression der Acetolactat-Synthase (AlsS) und der Acetolactat-Decarboxylase (AlsD) zur Synthese von Acetoin äußert. Unter Verwertung von Gluconat wird die Expression der beiden entsprechenden Gene reprimiert und stattdessen die Expression der Acetyl-CoA-Synthetase (AcsA) zur Synthese von Acetyl-CoA aus Acetat erhöht. Für andere Gene, welche in die zentralen Stoffwechselwege der Glykolyse und des Tricarbonsäurezyklus involviert sind, konnte keine auffällige Veränderung der Expression detektiert werden. Lediglich die Gene der Acetoin-Dehydrogenase acuABC werden teilweise leicht höher transkribiert. Die in der Abbildung 3.14. verwendeten Genkürzel sind in Tabelle 3.14. erläutert.



**Abbildung 3.14.:** Schematische Darstellung der Glykolyse unter Angabe der Expressionsverhältnisse der beteiligten Gene bei Wachstum auf Gluconat im Vergleich zum Wachstum auf Glucose: Für jedes Gen ist in Klammern das Expressionsverhältnis Gluconat/Glucose angegeben. Grüne Pfeile stellen eine deutlich höhere Genexpression bei Verwertung von Glucose im Vergleich zu Gluconat dar, rote Pfeile eine entsprechend höhere Genexpression bei Wachstum mit Gluconat. Die Funktionen der einzelnen Gene sind in Tabelle 3.14. angegeben. n.d. – nicht bestimmt

# 3.7.4.2. Expressions analyse des Gluconat-Operons mittels Real-time PCR

Die Gene der Gluconat-Verwertung sind im Genom von *B. licheniformis* in einem Operon angeordnet, welches *upstream* von einem putativen Transkriptions-Repressor und *downstream* von einer putativen Alkohol-Dehydrogenase flankiert wird (Abbildung 3.15.). Für diese Gene wurde zusätzlich zur qualitativen Expressionsanalyse mittels DNA-Microarray Technologie eine quantitative Untersuchung der Expression unter Einsatz der *Real-time* PCR (2.7.4.) vorgenommen.

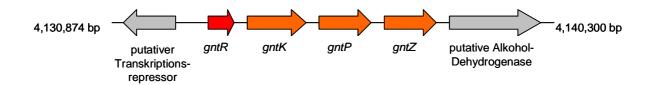


Abbildung 3.15.: Das Gluconat-Operon und umgebende Gene im Genom von B. licheniformis.

Für diese Analyse wurde die RNA aus *B. licheniformis* Zellen, angezogen auf Gluconat, bzw. auf Glucose aus jeweils zwei unabhängigen biologischen Experimenten verwendet. Für jede RNA wurden 6 konstitutive Gene mitgeführt (Daten nicht angegeben) und nach Ermittlung der C<sub>T</sub>-Werte mit Hilfe der unter 2.7.4. beschriebenen Formel der jeweilige Faktor der Expressionsänderung unter den zu vergleichenden Zuständen berechnet. In Tabelle 3.16. sind die Mittelwerte der C<sub>T</sub>-Werte jedes Gens für die Gluconat-RNA und die Glucose-RNA und die sich daraus ergebenden Expressionsfaktoren im Vergleich zu den DNA-Microarray-Daten angegeben.

**Tabelle 3.16.:** Expressionsfaktoren für die Gene *gnt*KPRZ des Gluconat-Operons bei Wachstum von *B. licheniformis* mit Gluconat im Vergleich zum Wachstum mit Glucose ermittelt durch DNA-Microarray-Technologie und *Real-time* PCR. n.d. - nicht bestimmt

ORF	Gen	Expressionsfaktor	MW C <sub>T</sub>	SD	MW C <sub>T</sub>	SD	Expressionsfaktor	
		Gluconat/Glucose	Gluconat		Glucose	[%]	Gluconat/Glucose	
		DNA-Microarray					Real-time PCR	
BLi04286	GntR	7,5	15,7	4,5	19,4	6,9	6,6	
BLi04287	GntK	8,5	14,2	9,5	19,3	5,7	18,4	
BLi04288	GntP	9,5	14,6	9,9	20,0	4,7	21,5	
BLi04289	GntZ	n.d.	14,2	9,1	19,4	9,6	18,7	

Die Transkriptionsanalyse ergab sowohl mittels DNA-Microarray Technologie, als auch mittels *Real-time* PCR einen ähnlichen Regulationsfaktor von 7,5, bzw. 6,6 für den Repressor des

Gluconat-Operons (GntR). Für die Gluconat-Permease (GntP) konnte im Rahmen der quantitativen *Real-time* PCR eine 18,4fache Expressionserhöhung bei Wachstum mit Gluconat im Vergleich zum Wachstum mit Glucose bestimmt werden, für die Gluconat-Kinase (GntK) eine Erhöhung um den Faktor 21,5 und für die Gluconat-6-Phosphat-Dehydrogenase (GntZ) eine Erhöhung der Expression um das 18,7fache. Die Expressionswerte, ermittelt über die qualitative DNA-Microarray Analyse, weichen für die Gene *gntP* und *gntK* um einen Faktor 2,2 von den mittels quantitativer *Real-time* PCR bestimmten Expressionsfaktoren ab.

# 3.7.4.3. Expressionsdaten anderer auffällig regulierter Enzyme

Im Vergleich zu den Expressionsdaten, welche sich für den Vergleich des Wachstums mit Acetat, bzw. 2,3-Butandiol und Glucose ergeben haben, ändert sich das Expressionsmuster unter Verwertung von Gluconat nur bei relativ wenigen Genen. Neben den unter 3.7.4.1. erwähnten Genen konnte lediglich für weitere 19 Gene eine Änderung der Expression um mindestens den Faktor 3 detektiert werden (Tabelle 3.17.). Darunter befinden sich ein putatives Gen und acht ,Y-Gene'. Bezogen auf Glucose als C-Quelle ist die Expression der Asparagin-Synthetase (AspO), des *Multidrug*-Efflux Transporters Blt, des Cytochrom P450-ähnlichen Enzyms CypX, des Transkriptions-Regulators der *puc*-Gene, zuständig für den Purin-Abbau, PucR und einiger Gene, deren Genprodukt in die assimilatorische Nitrat-Reduktion involviert sind, erhöht. Unter Verwertung von Gluconat konnte lediglich eine Expressionserhöhung für den Ferrichrom ABC Transporter FhuC und eine Translokations-abhängige antimikrobielle Sporenkomponente (TsaA) ermittelt werden. Eine mögliche Ursache für diese relativ geringe Änderung des Expressionsmusters unter diesen Bedingungen, liegt in der annähernd so guten Verwertbarkeit von Gluconat im Vergleich zu Glucose. Die Zelle kann sich mit geringem regulatorischem Aufwand an die neue C-Quelle anpassen.

**Tabelle 3.17.:** Expressionsdaten anderer Gene, welche mindestens um einen Faktor 3 beim Vergleich von Wachstum mit Gluconat und Glucose reguliert werden. Die Expressionsfaktoren sind jeweils als Verhältnis der Expressionswerte bei Wachstum mit Gluconat zu den Expressionsdaten bei Wachstum mit Glucose angegeben (Gluconat/Glucose). Gene deren Expression unter Verwertung von Gluconat erhöht ist, sind grau unterlegt.

ORF-Nr.	Produkt	Funktion	Gnt/Glc
BLi01174	AsnO	asparagine synthetase	0,27
BLi03437	Blt	multidrug-efflux transporter	0,34
BLi03567	СурХ	cytochrome P450-like enzyme	0,14
BLi03519	FhuC	ferrichrome ABC transporter (ATP-binding protein)	5,95
BLi00365	NasA	nitrate transporter	0,3
BLi00482	NasB	assimilatory nitrate reductase (electron transfer subunit)	0,19
BLi00483	NasC	assimilatory nitrate reductase (catalytic subunit)	0,22
BLi00485	NasE	assimilatory nitrite reductase (subunit)	0,31
BLi03436	PucR	transcriptional regulator of puc genes	0,27
BLi01175	putative protein	putative ammonium transporter	0,2
BLi02637	TasA	translocation-dependent antimicrobial spore component	6
BLi00856	YfiA	unknown; similar to unknown proteins	5,55
BLi01096	YhgE	unknown; similar to phage infection protein	0,32
BLi01593	YkvW	unknown; similar to heavy metal-transporting ATPase	0,3
BLi02639	YqxM	unknown	8,04
BLi02766	YrhE	unknown; similar to formate dehydrogenase	0,25
BLi03472	YusV	unknown; similar to iron(III) dicitrate transport permease	4,07
BLi03566	YvmC	unknown; similar to unknown proteins	0,07
BLi03569	YvnA	unknown; similar to unknown proteins from B. subtilis	0,2

# 3.7.5. Genexpression des Wachstums unter Glucoselimitierung im Vergleich zum Wachstum unter Stickstofflimitierung

Für die Transkriptionsanalyse von *B. licheniformis* Zellen, angezogen mit unlimitierter N-Quelle und limitierter C-Quelle im Vergleich zu einer Anzucht unter N-Limitierung und unlimitierter Glucose-Zugabe, wurde RNA aus jeweils zwei biologisch unabhängigen Experimenten verwendet. Es wurden fünf Hybridisierungen des Genomchips von *B. licheniformis* durchgeführt, wobei in 4 Experimenten die unterschiedlichen RNAs miteinander kombiniert wurden und in einer weiteren Analyse ein *dye switch* erfolgte. Die darauf beruhenden Expressionsdaten werden im Folgenden, zum einen bezogen auf den Kohlenstoff-Stoffwechsel und zum anderen hinsichtlich aller weiteren auffällig regulierten Gene, dargestellt.

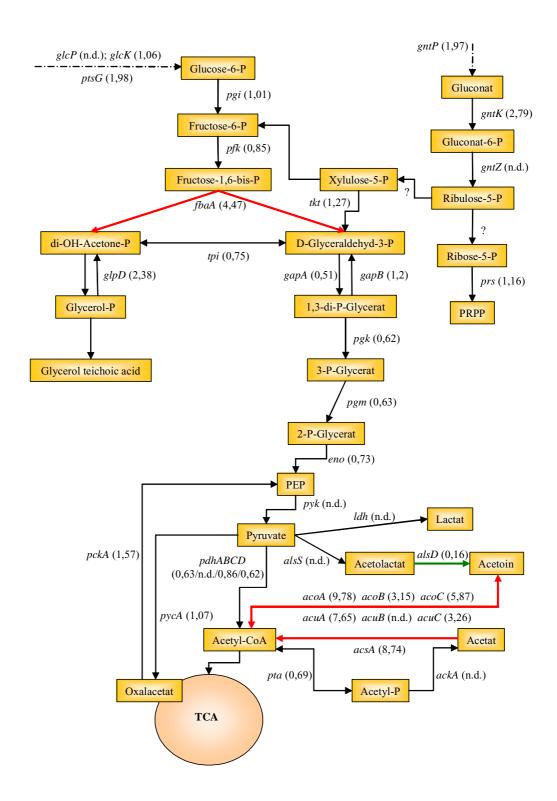
# 3.7.5.1. Expression der Gene des C-Stoffwechsels

Bezüglich der Gene, deren Produkte an der Glykolyse beteiligt sind, konnte eine erhöhte Trankription der Fructose-1,6-Bisphosphat-Aldolase um einen Faktor 4,5 unter Glucosemangel detektiert werden. Des Weiteren wird die Transkription der Acetolactat-Decarboxylase AlsD bei Glucose-limitiertem Wachstum um etwa den Faktor 6 reprimiert. Auch in diesem Experiment zeigt sich, dass offensichtlich unter Glucoseüberschuss ein *Overflow*-Metabolismus bezüglich der Synthese von Acetoin aus Acetolactat betrieben wird. Die Expression der Acetoin-Dehydrogenase-Gene *aco*ABC hingegen erhöht sich unter Glucosemangel um einen Faktor von 3,2 - 9,8. Dasselbe ist für die Acetoin-Dehydrogenase *acu*ABC zu beobachten. In diesem Falle konnte eine Expressionserhöhung in einem Bereich von 3,3 - 7,7 detektiert werden. Auch in diesem Experiment finden sich unter Glucosemangel Hinweise auf eine erhöhte Verwertung von Acetat durch die Acetyl-CoA-Synthetase AcsA. Die Expression von *acs*A erhöht sich unter diesen Bedingungen etwa um einen Faktor 8,7. Die Gene der Glykolyse mit den unter diesen Bedingungen bestimmten Transkriptionsdaten sind in Abbildung 3.16. dargestellt. Die in dieser Abbildung verwendeten Genkürzel sind in der Tabelle 3.14. erläutert.

# 3.7.5.2. Expressionsdaten anderer auffällig regulierter Enzyme

Zusätzlich zu den in den zentralen Stoffwechsel der Glykolyse und des Tricarbonsäurezyklus involvierten Genen (3.7.5.1.) konnte unter den gegebenen Bedingungen eine mindestens 3fache Expressionsänderung für weitere 51 Gene detektiert werden. Darunter befinden sich 13 so genannte "Y-ORFs" und die Gene von 12 putativen Proteinen. In Tabelle A.3. im Anhang sind diese ORFs mit den jeweils ermittelten Expressionsdaten angegeben.

Bei Wachstum von *B. licheniformis* unter Glucose-Limitierung im Vergleich zum Wachstum mit unlimitierter Zugabe von Glucose zeigt sich, analog zur Transkriptionsanalyse von mit Acetat oder 2,3-Butandiol gewachsenen Zellen, eine Expressionsreprimierung der in die Purin-Biosynthese involvierter Enzyme PurA, PurK und Xpt. Der Transkriptionsregulator PucR der *puc*-Gene, deren Funktion im Abbau von Purinen liegt, wird bei Glucosemangel ebenfalls geringer exprimiert. Dasselbe gilt für AlsR, den Transkriptionsregulator des α-Acetolactat-Operons; GlnR, den Transkriptionsrepressor der Glutamin-Synthetase Gene und SerA, die Phosphoglycerat-Dehydrogenase, welche in die Serin-Biosynthese involviert ist. Weiterhin konnte eine verringerte Expression der Ribonuklease HIII (RnhC) und einer Untereinheit der assimilatorischen Nitrit-Reduktase (NasE) detektiert werden. Eine Erhöhung der Expression



**Abbildung 3.16.:** Schematische Darstellung der Glykolyse unter Angabe der Expressionsverhältnisse der beteiligten Gene bei Stickstoff-unlimitiertem Wachstum zu Glucose-unlimitiertem Wachstum. Für jedes Gen ist in Klammern das Expressionsverhältnis N-unlimitiert/Glucose-unlimitiert angegeben. Grüne Pfeile stellen eine höhere Genexpression bei Glucose-unlimitiertem Wachstum dar, rote Pfeile eine entsprechend höhere Genexpression bei N-unlimitiertem Wachstum. Die Funktionen der einzelnen Gene sind in Tabelle 3.14. angegeben. n.d. – nicht bestimmt

ist hingegen für *eft*A und *eft*B (α- und β-Untereinheiten des Elektronentransfer-Flavoproteins), *kbl* (2-Amino-3-Ketobutyrate-CoA-Ligase), und *isp*A (Major intrazellulare Serin-Protease) zu vermerken. Für die Acetyl-CoA-Acetyltransferase MmgA und die Acyl-CoA-Dehydrogenase AcdA, involviert in die Lipidsynthese, konnte ebenfalls eine erhöhte Expression ermittelt werden. Weitere Gene, deren Expression unter Glucoselimitierung gesteigert wird, sind *dpp*E (Dipeptide ABC Transporter), *mal*A (6-Phospho-α-Glucosidase, *mms*A (Methylmalonate-Semialdehyd-Dehydrogenase), *prk*A (Serin-Protein-Kinase, *sac*T (Antiterminator, involviert in die positive Regulation von *sac*A und *sac*P des Sucrose Metabolismus) und *spo*IIIAH. Wie bereits bei dem Wachstum mit 2,3-Butandiol, ist auch hier unter Glucosemangel eine erhöhte Transkription der Gene für das Subtilisin Carlsberg und die Signalpeptidase SipV zu beobachten.

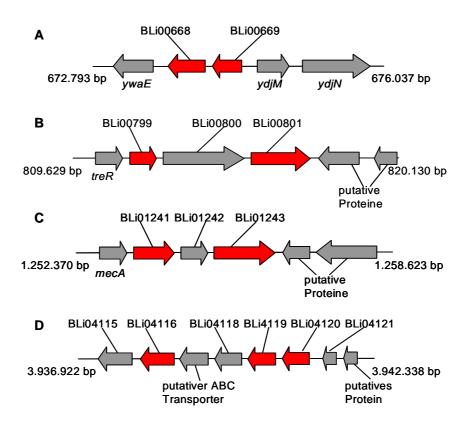
# 3.7.6. Expression hypothetischer und konserviert hypothetischer ORF's

Im Rahmen der Annotation des B. licheniformis Genoms (3.2.) konnte insgesamt 636 Gene keine Funktion über den Vergleich mit den aktuellen Sequenzdatenbanken zugewiesen werden. Davon wurden 483 ORFs, für welche keine bekannte homologe Entsprechung gefunden werden konnte, als hypothetisch annotiert. Weitere 153 ORFs zeigten Ähnlichkeiten zu hypothetischen Genen in anderen Genomen auf und wurden dementsprechend als konserviert hypothetische ORFs annotiert. Für die Existenz dieser Gene gibt es bislang keinen Nachweis, außer dem Vorhandensein eines Start- und Stopp-Codons, einer Ribosomenbindestelle und einer sinnvollen codon usage im Bereich des hypothetischen ORFs. Im Rahmen der Erstellung des Genomchips von B. licheniformis wurden auch diese ORFs bei einer Mindestlänge von 300 bp amplifiziert und gespottet. Basierend auf den Transkriptionsdaten, welche für B. licheniformis unter Verwertung unterschiedlicher C-Quellen ermittelt wurden, können Aussagen über die Expressionsänderung einiger hypothetischer und konserviert hypothetischer Gene getroffen werden. Bei insgesamt 38 ORFs konnte eine Expressionsänderung um mindestens einen Faktor 3 unter den gegebenen Bedingungen detektiert werden (Tabelle 3.18.). Für 19 hypothetische oder konserviert hypothetische ORFs wurde eine Änderung der Expression bei Wachstum mit Acetat im Vergleich zum Wachstum mit Glucose beobachtet, für 27 ORFs bei Wachstum mit 2,3-Butandiol im Vergleich zum Wachstum mit Glucose und bei einem ORF unter Verwertung von Gluconat im Vergleich zur Glucoseverwertung. Unter Glucoselimitierung, bzw. einer unlimitierten Zugabe von Stickstoff im Vergleich zum Wachstum mit limitierter N- und unlimitierter Glucosequelle konnte für 6 ORFs eine deutliche Expressionsänderung detektiert werden. 13 hypothetische Gene werden sowohl unter Verwertung von Acetat, als auch unter Verwertung von 2,3-Butandiol induziert, bzw. reprimiert.

**Tabelle 3.18.:** Hypothetische und konserviert hypothetische Proteine, deren Genexpression sich um mindestens einen Faktor 3 im Rahmen der Transkriptionsanalyse ändert. In fetten Lettern sind die Expressionswerte der Gene dargestellt, welche sowohl bei Wachstum mit Acetat, als auch mit 2,3-Butandiol eine Expressionserhöhung, bzw. –reprimierung aufzeigen. Die Expressionsdaten hypothetischer Gene, die nebeneinander im Genom von B. licheniformis lokalisiert sind, wurden grau unterlegt.

Gen-ID	Produkt	N-unlimitiert/	Gluconat/	Acetat/	Butandiol/
		Glc-unlimitiert	Glucose	Glucose	Glucose
BLi00198	hypothetical protein	6,44			
BLi00323	cons. hypothetical protein				3,25
BLi00443	hypothetical protein				6,95
BLi00491	cons. hypothetical protein	4,72		5,02	
BLi00654	cons. hypothetical protein			0,33	
BLi00668	hypothetical protein				3,88
BLi00669	hypothetical protein				3,67
BLi00719	hypothetical protein	0,25			0,12
BLi00725	cons. hypothetical protein			18,77	
BLi00799	cons. hypothetical protein				8,72
BLi00801	hypothetical protein			68,81	49,65
BLi00858	cons. hypothetical protein				19,53
BLi00930	cons. hypothetical protein	5,31			
BLi00933	cons. hypothetical protein			0,13	0,25
BLi01021	cons. hypothetical protein			13,58	
BLi01128	cons. hypothetical protein			5,03	4,69
BLi01150	hypothetical protein			0,06	0,09
BLi01241	cons. hypothetical protein				7,26
BLi01243	cons. hypothetical protein				8,65
BLi01271	hypothetical protein			467,15	133,79
BLi01389	hypothetical protein				18,59
BLi01488	hypothetical protein	0,08			
BLi01532	hypothetical protein	0,32			
BLi02005	hypothetical protein				4,01
BLi02206	hypothetical protein			11,04	4,38
BLi02352	hypothetical protein			14,82	8,17
BLi02709	cons. hypothetical protein				0,29
BLi02844	cons. hypothetical protein				5,07
BLi02908	hypothetical protein			43,01	17,35
BLi02966	hypothetical protein			181,32	22,15
BLi03073	hypothetical protein			3,56	
BLi03127	hypothetical protein			8,94	
BLi03570	hypothetical protein		0,224		
BLi04116	hypothetical protein			12,84	9,79
BLi04119	hypothetical protein			10,54	6,68
BLi04120	cons. hypothetical protein				3,66
BLi04184	hypothetical protein			0,08	0,25
BLi04326	hypothetical protein			12,98	7,4

Diese ORFs sind in Tabelle 3.18. in fetten Lettern hervorgehoben. Einige der aufgeführten ORFs sind im Genom von *B. licheniformis* nebeneinander lokalisiert. Die Expressionswerte der entsprechenden ORFs sind in Tabelle 3.18. grau unterlegt. In diesen Fällen kann vermutet werden, dass es sich um ein Operon oder Teile eines Operons handelt. Hierfür wurde die Umgebung dieser hypothetischen ORFs im Genom von *B. licheniformis* untersucht. Im Falle der hypothetischen Gene BLi00668 und BLi00669 befindet sich *upstream* der ,Y-ORF' *ywa*E, dessen Genprodukt Ähnlichkeit zu einem Transkriptionsregulator der MarR Familie aufweist, *downstream* befinden sich zwei weitere ,Y-ORFs' *ydl*M und *ydj*N (Abbildung 3.17.A).



**Abbildung 3.17:** Anordnung hypothetischer ORFs, deren Transkription bei Wachstum mit 2,3-Butandiol induziert wird, und angrenzender Gene. Höher exprimierte ORFs sind rot dargestellt. Für hypothetische Gene ist nur die ORF-Nummer im Genom von *B. licheniformis* angegeben. Erläuterungen im Text.

Für die hypothetischen ORFs BLi00799 und BLi00801 wurde ebenfalls eine deutliche Expressionserhöhung bei Wachstum mit 2,3-Butandiol beobachtet. Zwischen ihnen ist ein weiteres hypothetisches Protein codiert (BLi00800). *Upstream* dieser drei ORFs ist das Gen des Transkriptionsregulators TreR des Trehalose-Operons zu finden, downstream sind zwei Gene putativer Proteine lokalisiert (Abbildung 3.17.B). Hierbei handelt es sich um ein putatives Protein der Polysaccharide-Synthese für die Sporenhülle (BLi00802) und eine putative O-Acetyltransferase (BLi00803). Zwischen den induzierten hypothetischen Genen BLi01241

und BLi1243 liegt ein weiterer hypothetischer ORF BLi1242, welcher sich unter den gegebenen Bedingungen jedoch nicht durch eine Expressionsänderung hervorhebt (Abbildung 3.17.C). Flankiert werden diese Gene *upstream* von einem negativen Regulator der Kompetenz (MecA) und *downstream* von zwei Genen, deren putative Funktionen die eines Zwei-Komponenten-Responsregulators und einer Sensor Kinase (BLi01244, Bli01245) sind. Drei weitere Gene, die potentiell zu einem Operon gehören, sind BLi04116, BLi04119 und BLi04120 (Abbildung 3.17.D). Zwischen diesen drei ORFs sind ein weiteres hypothetisches Protein (BLi04118) und ein putativer ABC Transporter (BLi04117) codiert. Angrenzend an diese Gene liegen *upstream* die ORFs eines putativen Transkriptionsregulators (BLi04122) und eines konserviert hypothetischen Proteins (BLi04121). Ein weiteres hypothetisches Protein (BLi4115) wird *downstream* codiert. Insgesamt kann die detektierte Expressionsänderung der in Tabelle 3.18. aufgeführten hypothetischen und konserviert hypothetischen Gene als ein weiterer Beleg für ihre Existenz gewertet werden.

# 3.7.7. Zusammenfassung

In diesem Kapitel wurden die Ergebnisse der Transkriptionsanlyse von B. licheniformis mittels DNA-Microarray Analyse und teilweise durch Einsatz der Real-time PCR dargestellt. Die Expressionsmuster der B. licheniformis Gene während des Wachstums mit Acetat, 2,3-Butandiol und Gluconat im Vergleich zu Glucose konnten in einem genomweiten Kontext untersucht werden. Weiterhin wurde die Genexpression unter unlimitierter Zugabe von Stickstoff und Glucoselimitierung der Genexpression bei N-Limitierung und unlimitierter Zugabe von Glucose gegenübergestellt. Unter Verwertung von Acetat und 2,3-Butandiol konnte eine Umstellung des Kohlenstoffwechsels in Anpassung an die Verwertung von C2-Substraten beobachtet werden. Während der Nutzung dieser beiden Substrate erfolgte eine erhöhte Genexpression der Gene des Glyoxylatzyklus um etwa das 100fache. Mittels Real-time PCR konnte für die Malat-Synthase ein Expressionsfaktor von 305, 2 und für die Isocitrat-Lyase ein Faktor von 265,7 während der Verwertung von Acetat im Vergleich zu Glucose detektiert werden. Weitere Gene, deren Genprodukte essentiell in die Bereitstellung von Pyruvat, Phosphoenolpyruvat und Glyceraldehyd-3-Phosphat für die Gluconeogenese eingreifen, wie die Phosphoenolpyruvat-Carboxykinase PckA, die alternative Glyceraldehyd-3-Phosphat-Dehydrogenase GapB und teilweise das potentielle Malat-Enzym MalS, werden ebenfalls im erhöhten Maße exprimiert. Steht der Zelle hingegen Gluconat als einzige C-Quelle zur Verfügung, erfolgte eine erhöhte Expression des Gluconat-Operons. Die Quantifizierung der Genexpression mittels Real-time PCR ergab für die Gluconat-Permease einen Expressionsfaktor

von 21,5, für die Gluconat-Kinase 18,4, für die 6-Phosphogluconat-Dehydrogenase 18,7 und für den Transkriptionsregulator des Gluconat-Operons 6,6, jeweils im Vergleich zur Genexpression bei Wachstum mit Glucose. Die Referenz RNA für die beschriebenen Untersuchungen stammte jeweils aus Zellen, welche unlimitiert mit Glucose kultiviert waren. In allen vier Versuchen zeigte sich ein Rückgang der Genexpression der Acetolactat-Synthase AlsS und der Acetolactat-Decarboxylase AlsD für die Synthese von Acetoin. Weiterhin ist auffällig, dass die Expression einiger Gene, deren Genprodukte in die Purin- und Pyrimidin Biosynthese involviert sind, bei Glucosemangel reprimiert wird. Diese Beobachtung deutet darauf hin, dass *B. licheniformis* bei unlimitierter Zugabe von Glucose einen *overflow* Metabolismus betreibt.

Weiterhin konnte über die genomweite Transkriptionsanalyse eine zusätzliche Bestätigung für die Existenz von 38 hypothetischen und konserviert hypothetischen ORFs erbracht werden.

# 4. Diskussion

# 4.1. Allgemeine und komparative Analyse des Genoms von B. licheniformis

# 4.1.1. Generelle Eigenschaften des Genoms von B. licheniformis

B. licheniformis wird der B. subtilis Gruppe (Gruppe II) der Bacilli zugeordnet. In der Literatur finden sich vielfältige Ansätze für eine Klassifizierung des Genus Bacillus. Diese beruhen auf phänotypischen Eigenschaften und rRNA Sequenzvergleichen (Ash et al., 1991; Rossler et al., 1991). Für die moderne Klassifikation wurde oft der phylogenetische Vergleich der 16S rRNA genutzt, welche jedoch bei sehr eng verwandten Taxa nur wenig variiert (Fox et al., 1992). Für Proteine codierende Gene hingegen zeigen eine deutlich höhere Variabilität und werden daher verstärkt für phylogenetische Analysen verwendet (Fox et al., 1992; Kim et al., 1999; Yamamoto et al., 1999). So zeigten zum Beispiel Untersuchungen der Nukleotidsequenz der Gyrase A unterschiedlicher Vertreter der B. subtilis Gruppe die tiefste phylogenetische Verzweigung zwischen B. subtilis und B. licheniformis, wobei sich Bacillus amyloliquefaciens und Bacillus vallismortis erst später von B. subtilis abspalten (Chun and Bae, 2000). Eine weitere aktuelle Untersuchung der phylogenetischen Verwandtschaft zwischen Bacillus Spezies bezieht sich neben der 16S rDNA auf die Sequenz der 16S-23S internal transcribed spacer (ITS) Region (Xu and Cote, 2003). Dieser Klassifizierung zufolge sind B. licheniformis und B. subtilis einer anderen Gruppe als B. anthracis und B. cereus zugeordnet. Die Einordnung von B. halodurans wiederum erfolgte in eine weitere Gruppe. Umfangreiche Analysen mittels DNA/DNA Hybridisierung des phänotypisch sehr homogenen B. licheniformis ermöglichten die Erkennung genotypischer Varianzen und darauf basierend eine Einteilung von B. licheniformis Isolaten in drei unterschiedliche Genomovars (Manachini et al., 1998). Die Klassifizierungsansätze sind vielfältig und liefern fortlaufend neue und tiefere Einblicke in die phylogenetischen Verzweigungen. Die zunehmende Anzahl vollständiger Genomsequenzen erweitert das Feld für taxonomische Vergleiche. So lassen sich bereits durch die Analyse allgemeiner Genomdaten verwandtschaftliche Beziehungen, z.B. durch den Blick auf gemeinsame Kernbereiche der Genome (Kerngenome) und zusätzlich erworbene Fremdgene, ableiten.

B. licheniformis, B. subtilis, B. halodurans, B. cereus ATCC 10987, B. anthracis Ames. Die genetischen Fähigkeiten des Wildtypstammes Bacillus licheniformis DSM13 werden von

insgesamt 87,9% aller 4.222.645 Basenpaare seines Genoms verschlüsselt. Ein Vergleich der allgemeinen Eigenschaften des Genoms von B. licheniformis mit den Genomen von vier weiteren Repräsentanten der B. subtilis Gruppe (Gruppe II) der Bacilli (3.5.2.) zeigte eine deutliche Zweiteilung in eine B. licheniformis/ subtilis/ halodurans- und eine Bacillus anthracis/ cereus- ,Untergruppe'. Deutlich wird dies auch durch das Vorhandensein von großen Plasmiden in den Genomen von B. anthracis Ames und B. cereus ATCC 10987, auf welchen im Falle von B. anthracis Ames die Virulenzgene lokalisiert sind (Okinaka et al., 1999a; Okinaka et al., 1999b). B. licheniformis, B. subtilis 168 und B. halodurans hingegen weisen jeweils ein einzelnes Chromosom auf. Diese Chromosomen sind durchschnittlich etwa eine Megabase kleiner als die von B. anthracis Ames und B. cereus ATCC 10987, was sich in einer geringeren Anzahl an identifizierten open reading frames widerspiegelt. An dieser Stelle gilt es einschränkend zu bemerken, dass die Genomgröße einzelner B. cereus Stämme in einem Bereich von 2,4-5,3 Megabasen schwankt (Carlson and Kolsto, 1994). Eine komparative Genomanalyse im Zusammenhang mit der Sequenzierung von B. anthracis Ames konnte jedoch bereits eine hohe phylogenetische Verwandtschaft zu einigen B. cereus und Bacillus thuringiensis Stämmen zeigen (Read et al., 2003). Gegenüber den vier anderen Vertretern der B. subtilis Gruppe (Gruppe II) weist das Genom von B. licheniformis die geringste Anzahl an identifizierten rRNA Operons und tRNAs auf. Eine klare Divergenz der beiden genannten "Untergruppen" wird besonders bei der Betrachtung des durchschnittlichen G+C Gehaltes offensichtlich, welcher für B. licheniformis, B. subtilis und B. halodurans zwischen 43,5 und 46,2% liegt. Das Chromosom und die Plasmide von B. anthracis Ames und B. cereus ATCC 10987 hingegen weisen lediglich einen G+C Anteil von 32,5 bis 35,6% auf.

#### B. licheniformis, B. subtilis, B. halodurans.

Eine weiterführende Analyse der Genome von *B. licheniformis*, *B. subtilis* und *B. halodurans* auf dem ORF-Level (3.5.3.) zeigt, dass sie einen gemeinsamen genomischen Bereich (Kerngenom) besitzen, der für 2.323 orthologe Proteine codiert (s. Abbildung 3.9.). Die Anzahl orthologer ORFs von *B. halodurans* und *B. subtilis*, bzw. von *B. halodurans* und *B. licheniformis* jeweils ohne eine Entsprechung im dritten Organismus liegt gerade bei 146, bzw. 189. *B. subtilis* und *B. licheniformis* hingegen verfügen über weitere 872 gemeinsame orthologe Proteine, was zusätzlich auf einen höheren verwandtschaftlichen Grad hindeutet. Der große Anteil gemeinsamer funktioneller Eigenschaften von *B. licheniformis* und *B. subtilis* steht in einem sinnvollen Zusammenhang im Hinblick auf die Anpassung an ähnliche ökologische Nischen, von denen sich die alkaliphile Lebensweise von *B. halodurans* deutlich abhebt.

#### B. licheniformis und B. subtilis.

Der Anteil aller codierten Proteine von B. licheniformis, welche auf Ebene der Aminosäuresequenz eine Identität von mindestens 30% zu B. subtilis Proteinen zeigten, liegt bei erstaunlichen 74,6%. 14,8% aller ORFs weisen sogar eine Übereinstimmung von über 85% auf. Eine vergleichende Genomanalyse von B. anthracis Ames mit 19 Vertretern der B. cereus Gruppe anhand einer komparativen Genom-Hybridisierung zeigte, dass die Vertreter der B. cereus Gruppe 66 - 92% ihrer chromosomalen Gene mit B. anthracis Ames gemeinsam haben (Read et al., 2003). Damit wird eine hohe phylogenetische Verwandtschaft dieser Organismen belegt. Im Rahmen einer komparativen Genomanalyse von drei Escherichia coli Spezies wurde lediglich ein Kerngenom von 39,2% identifiziert (Welch et al., 2002). Ein Vergleich dieser Daten weist wiederholt auf die unmittelbare phylogenetische Nähe von B. licheniformis und B. subtilis hin. Eine gemeinsame Darstellung aller B. licheniformis ORFs mit denen, die ein Ortholog in B. subtilis besitzen (Abbildung 3.3.), zeigt große Bereiche von B. licheniformis Genen, welche ebenfalls in B. subtilis präsent sind, unterbrochen von Bereichen mit spezifischen Genen von B. licheniformis. Diesen Daten wurden ORFs gegenübergestellt, welche Aufgrund ihrer ungewöhnlichen Codon usage als alien genes identifiziert wurden (3.3.3.). Daraus ergab sich eine überraschende Übereinstimmung (s. Abbildung 3.3.). Ein Großteil der Gene, deren Produkte keine oder nur eine geringe Ähnlichkeit zu B. subtilis Proteinen besitzen, wurden als Fremdgene (alien genes) eingestuft, begleitet von starken Abweichungen des durchschnittlichen G+C Gehaltes in diesen Bereichen.

# Bakteriophagen.

Besonders offensichtlich ist die Kongruenz dieser drei Faktoren in den Regionen, in welchen drei der vier Prophagen von *B. licheniformis* lokalisiert sind. Ausgehend vom Replikationsursprung im Uhrzeigersinn sind das der zweite, dritte und vierte Phage. Der erste Phage hingegen, welcher dem Bakteriophagen PBSX von *B. subtilis* entspricht, ist weder durch eine abweichende *Codon usage*, noch durch einen überdurchschnittlich divergenten G+C Gehalt auffällig. Daher kann angenommen werden, dass der Phage schon seit sehr langer Zeit an der entsprechenden Stelle inseriert ist. Des Weiteren ist in der näheren Umgebung des ersten Phagen keines der insgesamt 21 Insertionselemente identifiziert worden. Diese Beobachtungen legen die Vermutung nahe, dass die Integration dieses Phagen sehr viel länger zurückliegt, als die der drei anderen, möglicherweise sogar vor dem evolutionären Split von *B. licheniformis* und *B. subtilis*. Für den Großteil der als *alien genes* identifizierten *B. licheniformis* Proteine ohne eine Homologie zu *B. subtilis*, die drei weiteren Prophagen eingeschlossen, kann ange-

nommen werden, dass ihre Integration erst nach diesem Ereignis durch horizontalen Gentransfer erfolgte.

#### Insertionselemente.

Wie bereits angesprochen, konnten insgesamt 21 Insertionselemente im *B. licheniformis* Genom gefunden werden. Darunter befinden sich neun identische Kopien eines IS-Elementes, welche der IS3 Familie zuzuordnen sind (Rey et al., 2004). Charakteristisch für Insertionselemente der IS3 Familie sind eine Länge von 1,2 bis 1,55 Kilobasen, *inverted terminal repeats* und zwei, teilweise überlappende ORFs (Mahillon and Chandler, 1998), welche, aufgrund einer programmierten Verschiebung des Leserahmens, ein einzelnes Genprodukt ergeben (Farabaugh, 1996). Acht dieser IS-Elemente sind in nicht codierenden Bereichen lokalisiert, während sich das neunte innerhalb des *comP* Gens befindet (Lapidus et al., 2002; Mahillon and Chandler, 1998). ComP ist die Sensor-Histidin-Kinase eines Zweikomponentensystems, involviert in die frühe Kompetenzentwicklung, welche in *B. subtilis* jedoch nicht zu den essentiellen Genen gehört (Kobayashi et al., 2003). Im Genus *Bacillus* sind bisher nur zwei weitere Organismen mit je einem IS-Element der IS3-Familie bekannt. Hierbei handelt es sich um die *B. thuringiensis* Stämme *B. thuringiensis* subsp. *aizawai* HD229 und *B. thuringiensis* YBT-226 (Mahillon and Chandler, 1998).

### Co-Linearität der Genome von B. licheniformis und B. subtilis.

Von Huynen und Bork wurde gefunden, dass die Anordnung orthologer Gene im Laufe der Evolution nur in geringem Maße konserviert ist (Huynen and Bork, 1998). Eine weiterführende komparative Analyse der Genome von B. licheniformis und B. subtilis unter diesem Gesichtspunkt mittels der MUMmer Plot Analyse (Delcher et al., 2002; Kurtz et al., 2004) (3.4.2.) zeigt jedoch eine deutliche Co-Linearität auf Ebene der Nukleotidsequenz. Große Bereiche der Genome sind sich in ihrer Nukleotidsequenz sehr ähnlich. Eine Unterbrechung der Co-Linearität erfolgt durch einzelne Segmente, bei denen es sich vermutlich um Fragmente externer DNA handelt, welche im Verlauf der Evolution dem Kerngenom hinzugefügt wurden. Die hohe Sequenzkonservierung der beiden Bacilli ermöglichte es beispielsweise die aus B. subtilis bekannten (Helmann and Moran Jr., 2002) Bindemotive von zwei Sigmafaktoren zu identifizieren, welche exemplarisch für die Untersuchung der Konservierung von DNA-Bindemotiven ausgewählt wurden. Die Funktion des Sigmafaktors SigD liegt in der Kontrolle eines Regulons, welches die späten Flagellar-Gene (Helmann et al., 1988), einige in die Chemotaxis involvierte Gene und die major vegetativen Autolysine (Lazarevic et al., 1992; Margot et al., 1994; Margot et al., 1999; Marquez et al., 1990) einschließt. Der zweite untersuchte Sigmafaktor SigL kontrolliert ein Regulons, welches unter anderem den Katabolismus einiger Aminosäuren, wie z.B. Arginin, Ornithin, Leucin und Valin (Debarbouille et al., 1999; Gardan et al., 1997), des Acetoins (Huang et al., 1999) und die Induktion des Levanase Gens (Debarbouille et al., 1991) beinhaltet. Vor den entsprechenden Genen konnte das jeweilige Erkennungsmotiv der Sigmafaktoren identifiziert werden (3.4.2.2.). Daraus lässt sich, zusätzlich zu einer Konservierung der Funktionalität und der Co-Linearität, auf eine hohe funktionelle Konservierung von regulatorischen Mechanismen im Bereich des Kerngenoms schließen.

Weitere Indizien der engen phylogenetischen Verwandtschaft von *B. licheniformis* und *B. subtilis* lassen sich in der überwiegenden Genanordnung auf dem *leading strand*, der Lokalisation ihrer rRNA Operons und der Struktur des Replikationsursprungs finden.

#### Co-Direktionalität von Replikation und Transkription.

Den Genera *Bacillus* und *Clostridum* ist eine co-direktionale Replikation und Transkription zueigen, welche sich in einer typischen ungleichmäßigen Verteilung der *open reading frames* auf dem *leading* und dem *lagging strand* äußert. Der Großteil, d.h. 74,3%, aller 4286 *B. licheniformis* ORFs sind, in Übereinstimmung mit 75% aller *Bacillus subtilis* ORFs (Kunst et al., 1997), auf dem *leading strand* lokalisiert. Dieser Prozentsatz ist jedoch geringer als 82% bei den vollständig sequenziert vorliegenden Genomen von *Clostridium tetani* (Brüggemann et al., 2003) oder *Clostridium perfringens* (Shimizu et al., 2002). Entscheidend für die Anordnung der Gene auf dem *leading* bzw. dem *lagging strand* ist offensichtlich ihr Genprodukt und nicht ihre Expressionsrate (Rocha and Danchin, 2003). In diesem Zusammenhang zeigten Untersuchungen, dass die Mehrzahl der als putativ essentiell identifizierter Gene in den bisher untersuchten Genomen auf dem *leading* strand lokalisiert sind (Rocha and Danchin, 2003). Allgemein wird als Hintergrund für eine Co-Direktionalität der Replikation und Transkription eine deutlich geringere Anzahl an Transkriptionsabrüchen als bei einer entgegengesetzten Replikation und Transkription angenommen (French, 1992; Rocha and Danchin, 2003).

#### Lokalisation der rRNA Operone.

Die sieben rRNA Operone, welche *B. licheniformis* aufweist, sind analog zu den 10 rRNA Operonen von *Bacillus subtilis* (Kunst et al., 1997) hauptsächlich in der Nähe des Replikationsursprungs geclustert. Bei beiden Organismen ist auffällig, dass sich die Mehrzahl der rRNA Cluster, sechs von sieben bei *B. licheniformis* und neun von zehn bei *B. subtilis*, auf dem *leading strand* der im Uhrzeigersinn verlaufenden Replikationsgabel befindet. Im Chromosom von *Clostridium perfringens* kann ebenfalls eine entsprechende Anordnung der rRNA-Cluster beobachtet werden .(Shimizu et al., 2002). Im Genom von *C. tetani* hingegen, ist die Mehrzahl der rRNA-Gencluster auf der gegen den Uhrzeigersinn verlaufenden Repli-

kationsgabel zu finden. Fraglich ist, ob es sich hier um ein Indiz für eine präferierte Replikationsrichtung bei der Replikation des Chromosoms handelt.

# Replikationsurprung.

Die Anordnung der Gene um den Replikationsurprung (*ori*) ist in vielen Organismen hoch konserviert (Fujita et al., 1989; Ogasawara et al., 1985; Ogasawara and Yoshikawa, 1986; Yoshikawa and Ogasawara, 1991). Ihre Funktion liegt im DNA- oder RNA-Metabolismus (*dnaA*, *dnaN*, *recF*, *gyrAB*, *rnpA*), der Sporulation (*spoJIII*, *jag*), der Proteinsynthese (*rmpH*) und in der Zellteilung (*gidA*, *gidB*). Das Enzym DnaA, essentiell für die Initiation der Replikation, bindet an eine konservierte Erkennungssequenz (TTATCCACA), welche in Form multipler *repeats* im nicht-codierenden Sequenzbereich das Gen *dnaA* flankiert. Von diesen so genannten DnaA-Bindeboxen sind insgesamt drei in *B. subtilis* vorhanden (Yoshikawa and Ogasawara, 1991). In *Escherichia coli* und *Pseudomonas putida* wurden hingegen nur eine bzw. zwei DnaA-Bindeboxen identifiziert (Fujita et al., 1989; Ogasawara et al., 1985; Ogasawara and Yoshikawa, 1986). Eine genaue Betrachtung des Replikationsursprungs von *B. li-cheniformis* führte zur Identifizierung von drei DnaA-Bindeboxen, deren Anordnung den Bindeboxen in *B. subtilis* entspricht.

Untersuchungen der Genome unterschiedlicher Spezies im Vergleich ergaben, dass die Genanordnung, der Gengehalt und die regulatorischen Mechanismen auch in eng verwandten Spezies sehr divergent sein können (Lathe et al., 2000). Entgegen dieser Beobachtungen zeigte die komparative Genomanalyse von *B. licheniformis* und *B. subtilis* eine beachtliche Konservierung der Funktionalität, eine Co-Linearität im Bereich des Kerngenoms und Hinweise auf konservierte regulatorische Mechanismen.

# 4.2. Das genetische Potential von *B. licheniformis* und *B. subtilis* im Vergleich

Nach den vorangegangenen Ausführungen liegt nun das Augenmerk auf den metabolischen Fähigkeiten von *B. licheniformis* und *B. subtilis* im Vergleich und *B. licheniformis* im speziellen. Diese werden im Folgenden hinsichtlich des Kohlenstoff- und Stickstoff-Stoffwechsels, des Sekretionsapparates, der Sporulation, der Restriktionssysteme, des anaeroben Wachstums und der Synthese von Exoenzymen näher beleuchtet. Eine vollständige Liste der 1091 Gene von *B. licheniformis*, welche kein Homolog in *B. subtilis* besitzen, ist im Anhang der Veröffentlichung des Genoms von *B. licheniformis* DSM13 zu finden (Veith et al., 2004).

#### 4.2.1. C-Stoffwechsel

Sowohl *B. licheniformis* als auch *B. subtilis* weisen die vollständige Enzymausstattung für die zentralen Stoffwechselwege der Glykolyse, des Penthosephosphat-Wegs und des Tricarbonsäurezyklus auf.

#### Glyoxylatzyklus.

Von großem Interesse ist die Identifizierung zweier Gene in B. licheniformis, welche eine starke Homologie zur Isocitrat-Lyase (BLi04207) (Sandeman et al., 1991) und der Malat-Synthase (BLi04208) (Sharma et al., 2000) zeigen. Gemeinsam mit einigen anderen Enzymen des Tricarbonsäurezyklus bilden diese beiden Enzyme den anaplerotischen Glyoxylatzyklus. Entscheidend sind die Aktivitäten der Isocitrat-Lyase, welche die Spaltung von Isocitrat zu Glyoxylat und Succinat katalysiert und der Malat-Synthase, die aus Glyoxylat und Acetyl-CoA Malat synthetisiert (Abbildung 4.1.). Der Glyoxylatzyklus befähigt den Organismus mit C2-Substraten als Haupt-C-Quelle zu wachsen. Seine Funktion liegt in der Versorgung des Tricarbonsäurezyklus mit dem Acetyl-CoA Akzeptor Oxalacetat, welches diesem durch biosynthetische Reaktionen entzogen wird. Alle Organismen, die über diesen Zyklus verfügen, sind in der Lage, Speicherlipide wieder als Kohlenhydrate zugänglich zu machen. B. licheniformis ist in der Lage sowohl mit Acetat, als auch mit 2,3-Butandiol als jeweils einziger C-Quelle zu wachsen. Bei beiden Substraten handelt es sich um, für Bacilli typische, Endprodukte der unvollständigen Oxidation oder Gärung. Durch die Verwertung derselben erschließt sich für B. licheniformis eine zusätzliche Möglichkeit des Energiegewinns während des Wachstums unter unzulänglicher Nährstoffversorgung.

**Abbildung 4.1.:** Die Spaltung von Isocitrat durch die Isocitrat-Lyase in Succinat und Glyoxylat (A) und die Synthese von Malat aus Glyoxylat und Acetyl-CoA durch die Malat-Synthase (B).

B. subtilis verfügt nicht über den Glyoxylatzyklus. Es konnten weder eine Malat-Synthaseoder eine Isocitrat-Lyase-Aktivität, noch ein Homolog zu bekannten Malat-Synthasen in der
Genomsequenz nachgewiesen werden (Sonenshein, 2002). Die Genomanalysen von
B. halodurans und B. anthracis hingegen führten zur Identifizierung homologer ORFs beider
Enzyme (Sonenshein, 2002). Die Synthese einer Malat-Synthase und einer Isocitrat-Lyase
konnte ebenfalls für Bacillus stearothermophilus belegt werden (Chell and Sundaram, 1975).
Ein interessanter Zusammenhang wurde bezüglich der Pathogenität von Mycobacterium tuberculosis, dem Erreger der Tuberkulose, und dessen Isocitrat-Lyase festgestellt. Es wird angenommen, dass in chronisch infiziertem Lungengewebe Fettsäuren die hauptsächliche Kohlenstoff- und Energiequelle für M. tuberculosis sind (Segal, 1984). Durch eine Inaktivierung
der Isocitrat-Lyase konnte tatsächlich ein Rückgang der bakteriellen Virulenz in immunkompetenten Mäusen gezeigt werden (McKinney et al., 2000).

Das Vorhandensein der Malat-Synthase und der Isocitrat-Lyase sowie deren funktionelle Bestätigung durch das Wachstum von *B. licheniformis* auf Acetat und 2,3-Butandiol erweitern die Vorstellungen von der Flexibilität dieses Organismus. Die Existenz der Gene des Glyoxylatzyklus und die Fähigkeit zur Verwertung der entsprechenden Substrate bedeutet für *B. licheniformis* unter entsprechenden Bedingungen einen klaren Selektionsvorteil gegenüber *B. subtilis*.

#### 4.2.2. Anaerobes Wachstum

Unter anaeroben Bedingungen ist die Denitrifikation, die dissimilatorische Nitratreduktion unter Freisetzung von molekularem Stickstoff, eine wichtige Form der Energiegewinnung für Bakterien. Anstelle von Sauerstoff dient Nitrat als terminaler Elektronenakzeptor der Atmungskette. Im Gegensatz zur assimilatorischen Nitratreduktion werden die Endprodukte der Denitrifikation, molekularer Stickstoff oder Stickoxid, nach außen abgegeben und nicht für die Synthese von Aminosäuren verwendet. In vielen Lehrbüchern werden als klassische Denitrifikanten *Pseudomonas* Spezies und *Bacillus* Arten, unter anderem *B. licheniformis*, genannt (Sneath et al., 1986). Weiterhin beschrieb im Jahre 1978 eine französische Arbeitsgruppe um M. Durand die Denitrifikation einiger *B. licheniformis* Stämme (Pichinoty et al., 1978). Entgegen den Erwartungen konnte im Genom von *B. licheniformis* DSM13 nicht die vollständige Enzymausstattung für die Denitrifikation gefunden werden. Lediglich ein Gen, welches eventuell für eine einzelne Untereinheit der NO-Reduktase codieren könnte und die Gene der dissimilatorischen Nitrat-Reduktase (*narGHIJ*) wurden innerhalb eines Clusters mit dem Gen des Transkriptionsregulators der anaeroben Gene (*finr*) identifiziert. Die Existenz der

dissimilatorischen Nitrat-Reduktase ist jedoch nicht unbedingt als ein Indiz für die Denitrifikation zu werten, da sie ebenso für die Nitratammonifikation notwendig ist. Experimentell ließ sich für *B. licheniformis* DSM13 und einige andere *B. licheniformis* Stämme (DSM14, DSM1913, DSM1969, DSM12369 und DSM12370) ebenfalls keine Denitrifikation nachweisen. Möglicherweise handelt es sich bei der Untereinheit der potentiellen NO-Reduktase um die nicht mehr funktionsfähigen Überreste eines Genclusters der Denitrifikation.

#### Anaerobe Ribonukleotid-Reduktase.

Ein interessanter Aspekt eröffnete sich mit der Identifizierung der Gene einer anaeroben Ribonukleotid-Reduktase der Klasse III (BLi03824) und zwei potentiellen zugehörigen Aktivierungsenzymen (BLi03823, BLi04172) im Genom von B. licheniformis. Ribonukleotid-Reduktasen katalysieren die Reduktion von Ribonukleosid-Triphosphaten zu den entsprechenden Desoxyribonukleotiden, welche essentiell für die DNA-Synthese sind. Diese Enzyme werden in drei Klassen eingeteilt, wobei Ribonukleotid-Reduktasen der Klasse I Sauerstoff benötigen, Klasse II Enzyme unabhängig von Sauerstoff agieren und auf Enzyme der Klasse III Sauerstoff toxisch wirkt. Bisher wurden die anaeroben Ribonukleotid-Reduktasen von E. coli und von Lactococcus lactis isoliert und charakterisiert (Reichard, 1993; Torrents et al., 2000). Hier handelt es sich um einen Gram-negativen und einen Gram-positiven Organismus, für deren anaerobe Ribonukleotid-Reduktasen der Klasse III ein gemeinsames generelles Funktionsprinzip nachgewiesen wurde (Torrents et al., 2000). Entscheidend für die katalytische Aktivität der anaeroben Ribonukleotid-Reduktase ist ein Sauerstoff-sensitives Glycyl-Radikal. Für die Generierung des Glycyl-Radikals und damit für die Aktivierung der Ribonukleotid-Reduktase ist ein spezifisches Aktivierungsenzym zuständig (Sun et al., 1995). Anhand der Genomsequenz konnten in mehreren Gram-negativen Bakterien, Gram-positiven Bakterien und Archaeen vollständige Sequenzen für die anaerobe Ribonukleotid-Reduktase identifiziert werden (Torrents et al., 2000). B. cereus besitzt ebenfalls diese Gene und ist nachweislich in der Lage, anaerob zu wachsen (Sneath et al., 1986). Eine BLAST Analyse der Arbeitsgruppe um D. P. Nagle zeigte, dass in allen verfügbaren Bacillus-Sequenzen putative anaerobe Ribonukleotid-Reduktasen der Klasse III gefunden werden konnte, außer in B. subtilis (Folmsbee et al., 2004). Weiterhin konnten sie zeigen, dass B. subtilis und Bacillus mojavensis zum strikt anaeroben Wachstum die Zugabe von DNA oder Desoxyribonukleosiden benötigen. Diese Ergebnisse stehen im Einklang mit dem Fehlen einer anaeroben Ribonukleotid-Reduktase in B. subtilis, zeigen aber auch, das B. subtilis in der Lage ist, diesen Mangel durch Aufnahme externer DNA oder Desoxyribonukleoside auszugleichen. Untersuchungen zu folge lysiert B. subtilis unter Nährstoffmangelbedingungen, um andere Zellen mit Nährstoffen zu versorgen und die Sporulation herauszuzögern (Gonzalez-Pastor et al., 2003). Zu den Nährstoffen, welche durch die Lyse von Zellen zugänglich werden, gehören auch DNA und Desoxyribonukleotide. Hypothetisch wäre es *B. subtilis* auf diese Weise auch unter Sauerstoffmangel möglich, für einen begrenzten Zeitraum die DNA-Synthese aufrecht zu erhalten.

B. licheniformis ist, nach in dieser Arbeit geschilderten Experimenten (3.6.1.), in der Lage anaerob in Minimalmedium, ohne weitere Zusätze und mit Glucose als einziger C-Quelle zu wachsen. Nach den obigen Ausführungen deutet dies auf eine Aktivität der identifizierten Ribonukleotid-Reduktase in vivo hin. Die vergleichende Analyse der anaeroben Ribonukleotid-Reduktase von E. coli und Lactococcus lactis verweist auf eine hohe Konservierung des Funktionsprinzips dieses Enzyms (Torrents et al., 2000). Ob dies für die B. licheniformis Ribonukleotid-Reduktase ebenfalls zutrifft, erfordert detaillierte biochemische und proteinchemische Untersuchungen. Auch in diesem Zusammenhang lässt sich eine größere Flexibilität von B. licheniformis gegenüber B. subtilis, hinsichtlich der Anpassung an eine unzureichende Sauerstoffversorgung, erkennen.

#### 4.2.3. N-Stoffwechsel

Sowohl B. licheniformis als auch B. subtilis weisen eine hohe metabolische Diversität hinsichtlich ihrer Fähigkeiten zur Nutzung unterschiedlicher Stickstoff-Quellen auf. Im Zuge der Adaptation an ihre saprophytische Lebensweise sind sie in der Lage, mittels extrazellulärer Enzyme aus exogenen Reservoirs, wie Proteinen, Peptiden und Aminosäuren, Stickstoff zu akquirieren. Beide Organismen verfügen diesbezüglich über eine Glutamat-spezifische Protease (mpr), zwei Serinproteasen (vpr, epr), zwei weitere Proteasen (yrrN, yrrO) und die Peptidase T (pepT) (s. Tabelle 3.6.). Zusätzlich besitzt B. licheniformis eine Zink-Protease (BLi01909), welche kein Homolog in B. subtilis aufweist (s. Tabelle 3.7.). Beide Organismen verfügen über die Gene der Schlüsselenzyme für die Stickstoff-Assimilation, was auf eine ähnliche Funktionsweise der Systeme hinweist (Rey et al., 2004). Dazu gehören die Glutamin-Synthetase (GlnA), das wichtigste Enzym der NH<sub>3</sub>-Assimilation, ihr Transkriptionsrepressor (GlnR), der Transkriptionsregulator TnrA der globalen Stickstoff-Regulation, der Ammonium-Transporter (NrgA) und die Gene des Nitrat-Transporters NasA, der assimilatorische Nitrat-Reduktase NasBC und der assimilatorischen Nitrit-Reduktase NasDE. Glutamin dient als Ausgangsmolekül nahezu aller Aminogruppen und wird in erster Instanz durch die Glutamin-Synthetase bereitgestellt. Über die Nitratassimilation akquiriert die Zelle ebenfalls

NH<sub>3</sub>, dieser Prozess ist aber mit einem höheren Energieaufwand verbunden. Aus diesem Grund erfolgt eine Repression der beteiligten Enzyme durch Ammonium.

# 4.2.4. Sekretionsapparat und Exoenzyme

In Anpassung an ihre saprophytische Lebensweise verfügen *B. licheniformis* und *B. subtilis* über die Fähigkeit eine Vielzahl extrazellulärer, hoch polymerer Substrate zu nutzen. Notwendig dafür sind extrazelluläre Hydrolasen, deren Gesamtheit zusammen mit anderen in das Medium sekretierten Proteinen das so genannte Sekretom bilden. Pre-Exoenzyme verfügen am Amino-Terminus über ein Signal-Peptid, anhand dessen die Erkennung und Bindung an den Sekretionsapparat vermittelt wird. Im Anschluss an die Translokation wird dieses Signal-Peptid von Signal-Peptidasen entfernt und das Exoenzym freigegeben.

# Sec-Sekretionssystem und Signalpeptidasen.

In B. licheniformis und B. subtilis sind alle Gene des Sec Sekretionssystems secA (BLi03773), secDF (BLi02891), secE (BLi00118), secG (BLi03643), secY (BLi00153), ffh (BLi01818) und ftsY (BLi01816) vorhanden. Für das Twin-Arginine-Translocation (TAT) Exportsystem konnten vier Komponenten identifiziert werden (BLi00282, BLi00283, BLi 00619, BLi00620). Dem Sec-System kommt in diesem Zusammenhang die größte Bedeutung zu, da über diesen Weg mit Abstand die meisten Proteine transloziert werden (Tjalsma et al., 2000). Ebenso besitzen beide Organismen die Gene der vier Signal-Peptidasen sipS (BLi00675), sipT (BLi01655), sipV (BLi01122) und sipW (BLi02638) (Tjalsma et al., 2000). B. subtilis verfügt über eine weitere Signal-Peptidase SipU. Ein entsprechendes Ortholog für dieses Enzym konnte jedoch in B. licheniformis nicht identifiziert werden. Das Fehlen von SipU ist allerdings nicht außergewöhnlich, da einigen B. subtilis Stämmen bis zu vier der fünf Signal-Peptidasen fehlen (Tjalsma et al., 1998). Es konnte gezeigt werden, dass im Grunde das Vorhandensein von SipS oder SipT ausreichend für eine Prozessierung von Exoenzymen ist. Zellen mit keiner dieser beiden Signal-Peptidasen sind nicht überlebensfähig (Tjalsma et al., 1998). Die genannten Signal-Peptidasen weisen scheinbar ähnliche Substratspezifitäten auf, bevorzugen jedoch unterschiedliche Pre-Proteine, wie Studien an B. amyloliquefaciens zeigten (Tjalsma et al., 1997). Das Zusammenwirken der Enzyme des Sekretionsapparates und das Vorhandensein von bis zu fünf Signalpeptidasen sind möglicherweise entscheidende Faktoren, welche den Bacilli ihre natürliche Kapazität einer erstaunlich vielfältigen Sekretion an Exoenzymen, in zum Teil sehr hohen Konzentrationen, verleihen. Die Exoenzymproduktion in diesem Maßstab durch eine Vielzahl an Bacillus Spezies findet eine breite Anwendung in

der Industrie. In Abbildung 4.2. sind die Komponenten des Sec-abhängigen Proteintransportes in *B. subtilis* dargestellt.

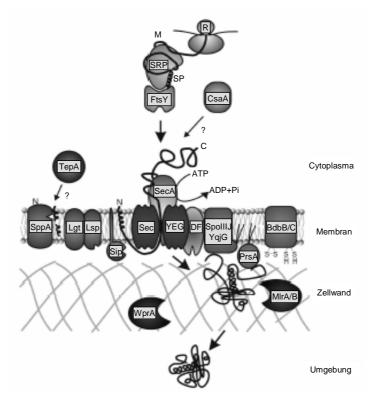


Abbildung 4.2.: Komponenten des Sec-abhängigen Proteintransportes in *B. subtilis* (Tjalsma et al., 2004). Sekretorische Proteine werden ribosomal (R) als ein Vorläuferprotein synthetisiert, welches aus dem Hauptprotein (M) und einem N-terminalen Signalpeptid (SP) besteht. Cytoplasmatische Chaperone (SRP/FtsY; CsaA) halten das Vorläuferprotein in einem translozierbaren Zustand und vermitteln die Bindung an die Komponenten der Translokase (SecA, SecE, SecY, SecG, SecDF). Während oder kurz nach der Translokation wird das Signalpeptid entweder von Typ I Signal-Peptidasen (SipS-W) entfernt oder durch die Diacylglyceryl-Transferase (Lgt) modifiziert und anschließend von der Lipoprotein-spezifischen Signal-Peptidase (Lsp) abgespalten. SspA und TepA sind möglicherweise in den Abbau des Signalpeptids involviert. Für die Faltung des sekretorischen Proteins sind PrsA, BdbBC und/oder SpollIJ/YqjG zuständig. In die Qualitätskontrolle des Exoenzyms sind HtrA, HtrB und WprA involviert.

#### Exoenzyme.

Zu den industriell relevanten Enzymen, welche sowohl von *B. subtilis*, als auch von *B. licheniformis* synthetisiert werden, gehören Polysaccharid-, Protein- (s. 4.2.3.) und Lipidabbauende Enzyme. Insgesamt 38 Enzyme dieser Art konnten in *B. licheniformis* identifiziert werden (s. Tabellen 3.6. und 3.7.), 17 davon ohne eine homologe Entsprechung in *B. subtilis*. Besonders erwähnt seien hier das Subtilisin Carlsberg (Jacobs et al., 1985; Smith et al., 1968; Syed et al., 1993), die Glutamat-spezifische Protease (Kakudo et al., 1992; Svendsen and Breddam, 1992), die Hitze- und pH-stabile α-Amylase (Gray et al., 1986; Kandra et al., 2002; Stephens et al., 1984; Yuuki et al., 1985), die maltogene α-Amylase (Kim et al., 1992), Cellulose- und Chitinabbauende Enzyme.

#### Subtilisin.

Subtilisine sind sehr gut untersuchte alkalische Serin-Proteasen, die in die zwei Typen Subtilisin Carlsberg und Subtilisin BPN' unterschieden werden. Analysen der Aminosäuresequenz der Enzyme zeigten eine Homologie von etwa 70% (Smith et al., 1968). Subtilisine werden von verschiedenen *Bacillus* Spezies synthetisiert und finden ihre primäre kommerzielle Anwendung in Haushaltsdetergenzien mit einer jährlichen Produktion von bis zu 500 Tonnen (Rao et al., 1998). *B. licheniformis* (BLi01109), *B. subtilis* und *B. pumilis* produzieren den Enzym Typ Subtilisin Carlsberg (Zukowski, 1989), wobei *B. subtilis* 168 kein Homolog dieses Proteins besitzt.

### Glutamat-spezifische Protease.

Über die Glutamat-spezifische Protease (Mpr) (BLi00340) verfügen sowohl *B. licheniformis* DSM13, als auch *B. subtilis* 168. Dieses Enzym hydrolysiert spezifisch die Peptidbindungen Glu-Xaa und Asp-Xaa, wobei Glu-Xaa bevorzugt wird (Svendsen and Breddam, 1992). Anwendungen solch spezifischer Proteasen finden sich in der Enzymmodifikation, z.B. der Produktion biologisch aktiver Peptide (Kakudo et al., 1992).

#### α-Amylasen.

α-Amylasen im Allgemeinen finden vielfältigen Einsatz in der Lebensmittel-, Papier-, Fermentations- und Textilindustrie, aufgrund ihrer hydrolytischen Aktivität gegenüber α-1,4-Bindungen in Stärke. Die Substratspezifität dieser Enzyme geht jedoch über den Abbau von Stärke hinaus und die Produkte variieren enzymspezifisch. B. licheniformis besitzt zwei α-Amylasen, die maltogene α-Amylase (BLi00658) (Kim et al., 1992) und die hitze- und pHstabile α-Amylase (BLi00656), deren Gene jeweils kein Homolog in B. subtilis aufweisen. Die maltogene α-Amylase ist nicht thermostabil und besitzt die Fähigkeit sowohl Stärke, als auch Pullulan und Cyclodextrine zu hydrolysieren. Der Abbau von Stärke findet zudem vorwiegend unter Abspaltung von Maltose statt. Dieses Enzym besitzt des Weiteren eine Transferaseaktiviät und katalysiert die Bildung von α-1,6-glycosidischen Bindungen (Kim et al., 1992). Im Vergleich dazu hydrolysiert die α-Amylase von Bacillus stearothermophilus KP 1064 lösliche Stärke zu den Hauptprodukten Maltose und Glucose, besitzt ebenfalls eine Cyclodextrinaseaktivität, aber keine Transferaseaktivität (Suzuki and Imai, 1985). Aus Bacillus megaterium konnte ein amylolytisches Enzym isoliert werden, welches Stärke und Pullulan hydrolysiert und die Bildung glycosidischer α-1,4-Bindungen katalysiert (David et al., 1987). Die thermo- und pH-stabile α-Amylase von B. licheniformis produziert überwiegend Maltohexaose und Maltoheptaose bei der Hydrolyse von Stärke und kann weder Pullulan noch Cyclodextrine hydrolysieren (Kim, 1991). Das B. licheniformis Enzym ist in seiner DNA- und

seiner Proteinsequenz homolog zu der thermophilen  $\alpha$ -Amylase von *B. stearothermophilus* und gehört zur selben Enzymfamilie wie die  $\alpha$ -Amylasen von *B. coagulans* und *B. amyloliquefaciens* (Gray et al., 1986). Die  $\alpha$ -Amylase von *B. subtilis* wird nicht zu dieser Enzymfamilie gezählt (Gray et al., 1986).

#### Celluloseabbau.

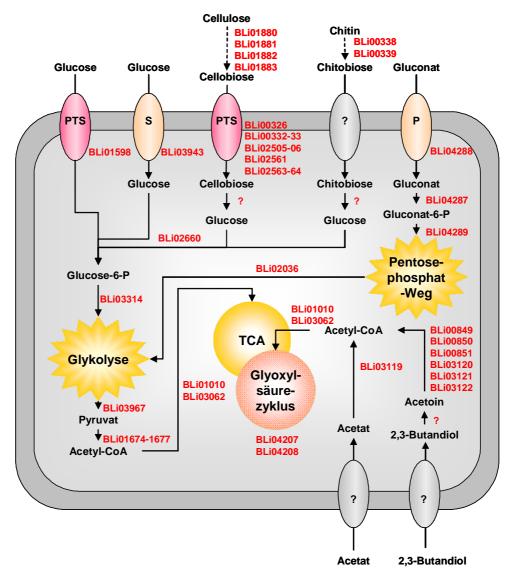
Cellulose ist mengenmäßig weltweit eines der bedeutendsten Strukturpolysaccharide und der Hauptbestandteil pflanzlicher Zellwände. In diesem Zusammenhang ergibt sich aus der Identifizierung von Genen Celluloseabbauender Exoenzyme in der Genomsequenz, welche kein Homolog in B. subtilis 168 aufweisen, ein neuer Aspekt hinsichtlich der Nutzung unterschiedlicher C-Quellen durch B. licheniformis. Dazu gehören eine Cellulose-1,4-β-Cellobiosidase (BLi01881), eine Cellulase (BLi01882), eine Endo-1,4-β-Mannosidase (BLi01883) und eine Endo-1,4-Glucanase (BLi01880), deren Gene in einem Cluster vorliegen. Diese Enzyme katalysieren die Hydrolyse von Cellulose in ihre Untereinheiten, die Cellobiose. Es konnten weitere putative Gene identifiziert werden, die für ein Cellobiose spezifisches Phosphotransferasesystem (PTS) (Bli00332, Bli00333, BLi00326, BLI02505, Bli02506, Bli02561, BLi02563, BLi02564) und β-Glucosidasen codieren. Damit verfügt B. licheniformis theoretisch über das vollständige genetische Potential zur Hydrolyse von Cellulose in die Cellobioseuntereinheiten, deren Transport in das Cytoplasma und die Hydrolyse in die Glucose Bausteine. In diesem Kontext konnte gezeigt werden, dass B. licheniformis in der Lage ist, Carboxymethylcellulose zu verwerten (Rey et al., 2004). Für B. subtilis wurde zwar die Verwertung von Cellobiose beschrieben, entsprechende Gene oder Proteine für die Aufnahme und Spaltung dieses Substrates konnten bisher jedoch nicht gefunden werden (Lai et al., 1997; Steinmetz, 1993; Stülke and Hillen, 2000). B. stearothermophilus hingegen verfügt über ein Operon, welches für die Cellobiosenutzung codiert (Lai and Ingram, 1993).

#### Chitinabbau.

Im Genom von *B. licheniformis* DSM13 konnten zwei Gene in einem Cluster identifiziert werden, die potentiell für zwei Chitinasen (BLi00338, Bli00339) codieren. Diese Gene besitzen keine homologe Entsprechung im Genom von *B. subtilis*. Chitin ist in der Natur ein Äquivalent der Cellulose sowie Bestandteil des Exoskeletts von Invertebraten und der Zellwände von Pilzen. Die Untereinheiten des Chitins, Diacetylchitobiose, bestehen aus zwei β-1,4-glycosidisch verknüpften N-Acetyl-Glucosaminen. Chitinasen können die β-1,4-Bindungen endo- oder exogen hydrolysieren, die Spaltung von Chitobiose erfolgt in der Regel durch Chitobiasen. In der Genomsequenz von *B. licheniformis* DSM13 konnte kein Gen mit einer Homologie zu bekannten Chitobiasen gefunden werden, aber ein putatives Chitin bindendes Pro-

tein (BLi00521). Bereits gut untersucht ist das Chitinase-System von *Bacillus circulans* (Watanabe et al., 1990). Es gilt die Funktionalität der Chitinabbauenden Gene in *B. licheniformis* DSM13 experimentell durch den Abbau von Chitin zu belegen.

Die Einblicke in die Fähigkeiten von *B. licheniformis* hinsichtlich seines vielschichtigen Potentials in der Nutzung unterschiedlichster C-Quellen vergrößern die Vorstellung von seiner Anpassungsfähigkeit an ökologische Nischen und eröffnen eventuell neue kommerzielle Einsatzmöglichkeiten. In Abbildung 4.3. sind die Abbauwege der erwähnten Substrate in *B. licheniformis* schematisch dargestellt.



**Abbildung 4.3.:** Aufnahme und Verwertung einzelner Substrate durch *B. licheniformis*. Für die jeweilig zuständigen Enzyme ist in rot die ORF-Nummer angegeben.

PTS - Phosphotransferasesystem

S - Symporter

P - Permease

- Transport in die Zelle unbekannt / zuständiges Enzym unbekannt

TCA - Tricarbonsäurezyklus

# 4.2.5. Antibiotika, Sekundärmetabolite und Siderophore

#### Extra-ribosomale Lipopeptide.

Von vielen Bacillus Stämmen ist bekannt, das sie extra-ribosomale, oberflächen-aktive Lipopeptide synthetisieren (Fiechter, 1992). Zu den bekanntesten Lipoproteinen zählt das von B. subtilis produzierte Surfactin, das selbst in geringen Konzentrationen die Oberflächenspannung von Wasser reduziert und sowohl antibakterielle als auch fungizide Eigenschaften besitzt (Arima et al., 1968). Neben B. subtilis gehört B. licheniformis zu den Surfactin-Produzenten (Fox and Bala, 2000; Moran et al., 2000; Peypoux et al., 1999). Das Antibiotikum Bacitracin, welches die Biosynthese der Zellwand inhibiert, wird ebenfalls von einigen B. licheniformis Stämmen extra-ribosomal synthetisiert (Katz and Demain, 1977). Überraschenderweise konnte in B. licheniformis DSM13 kein Operon für die Bacitracin-Synthese gefunden werden, stattdessen wurden drei Gene eines extra-ribosomalen Peptidsynthetase Komplexes mit dem zugehörigen Gen einer putative Thioesterase identifiziert, welche für den Lichenysin-Synthetase Komplex codieren (BLi00401, BLi0402, BLi00403, BLi00404) (Konz et al., 1999). Das Lichenysin Operon und das Bacitracin Operon weisen eine hohe Homologie auf, die sich in einer ähnlichen Struktur der Syntheseprodukte widerspiegelt (Grangemard et al., 2001). Ein großer Vorteil von Lichenysin gegenüber Surfactin ist, dass seine Synthese auch unter strikt anaeroben Bedingungen möglich ist (Javaheri et al., 1985; Yakimov et al., 1995). Das ist von Bedeutung beim Einsatz entsprechender B. licheniformis Stämme im Rahmen der Ölförderung (Yakimov et al., 1995).

#### Polyketid-Synthasen.

Die Synthese einer Vielzahl mikrobieller Metabolite ähnelt der Fettsäuresynthese und wird von so genannten Polyketid-Synthasen katalysiert. *B. subtilis* verfügt über solche Enzyme, wie z.B. die Polyketid-Synthase Pks und die Pliplastin-Synthase Pps. Ein homologes Gegenstück zu einem dieser Enzyme konnte in *B. licheniformis* DSM13 nicht identifiziert werden. Untersuchungen hinsichtlich einer minimalen Genausstattung von *B. subtilis* zeigten, dass diese ORFs nicht zu den essentiellen Genen gehören (Kobayashi et al., 2003).

#### Siderophore.

Es wurde mehrfach eine rotbraune Färbung des Mediums während des Wachstums von *B. licheniformis* beobachtet (Sneath et al., 1986). Im Zusammenhang mit dieser Pigmentbildung könnten vier Gene stehen, die im Genom von *B. licheniformis* identifiziert wurden (BLi01185, BLi01186, BLi01187, BLi01188). Diese ORFs weisen Homologien zu den Genen *iucABCD* auf, welche für das Siderophor Aerobactin auf dem enterobakteriellen Virulenzplasmid pColV-K30 codieren (Martinez et al., 1994). Siderophore sind Chelatkomplexe,

welche mit hoher Affinität unlösliches Eisen komplexieren und in die Zelle transportieren. Eisen ist eine essentielle Komponente von Cytochromen und Eisen-Schwefel-Proteinen, die in den Elektronentransport involviert sind. Für *B. subtilis* wurde die Synthese eines Siderophores nachgewiesen, das ebenfalls homolog zu einem enterobakteriellen Protein ist, in diesem Falle zu Enterobactin von *E. coli* (Grossman et al., 1993).

# 4.2.6. Sporulation

Bis auf sechs Gene weist *B. licheniformis* DSM13 eine homologe Entsprechung zu allen in *B. subtilis* 168 identifizierten Sporulationsgenen (Kunst et al., 1997) auf. Bei den fehlenden ORFs handelt es sich um Gene, deren Produkte in die Synthese eines Polysaccharids der Sporenhülle involviert sind (Rey et al., 2004). Es scheint sich jedoch nicht, um für die Sporulation essentielle Gene zu handeln, da *B. licheniformis* DSM13 am Ende der logarithmischen Wachstumsphase hitzeresistente Sporen ausbildet.

# 4.2.7. Restriktionssysteme

Das industrielle Interesse an der Expression rekombinanter Enzyme durch B. licheniformis nimmt aufgrund höherer Enzymerträge im Vergleich zu B. subtilis stetig zu. Es stellte sich jedoch heraus, dass etablierte Methoden zur Transformation von B. subtilis mit heterologer DNA (Ferrari and Hoch, 1989; Mountain, 1989; Rao et al., 1998; Spizizen, 1958) nicht dieselbe Effizienz bei B. licheniformis zeigten (Schallmey et al., 2004). Eine mögliche Erklärung bietet die Identifizierung von zwei Operonstrukturen, die für Typ I Restriktionsenzyme codieren (BLi04315, BLi04316, BLi04318 und BLi00743, BLi00744, BLi00745, BLi00746). Die beiden Operone zeigen eine hohe Ähnlichkeit zu dem Hsd-System (Host specificity of DNA). Hierbei handelt es sich zumeist um Heteropentamere, welche DNA ausgehend von einer asymmetrischen Erkennungssequenz zufällig, aber mindestens 1000 bp entfernt, schneiden (Wilson and Murray, 1991). Dieser Enzymkomplex weist neben einer Restriktionsaktivität auch die Fähigkeit der Modifikation unvollständig methylierter DNA auf (Abbildung 4.4.). Soweit bekannt, verfügt B. subtilis über keine Typ I Restriktionssysteme. Die Analyse der Codon usage des B. licheniformis Genoms zeigte, dass es sich bei den genannten ORFs mit hoher Wahrscheinlichkeit um alien genes handelt, welche durch horizontalen Gentransfer in das Chromosom gelangt sind. Eine abweichender G+C Gehalt dieser Gene unterstützt diese These (Rey et al., 2004).

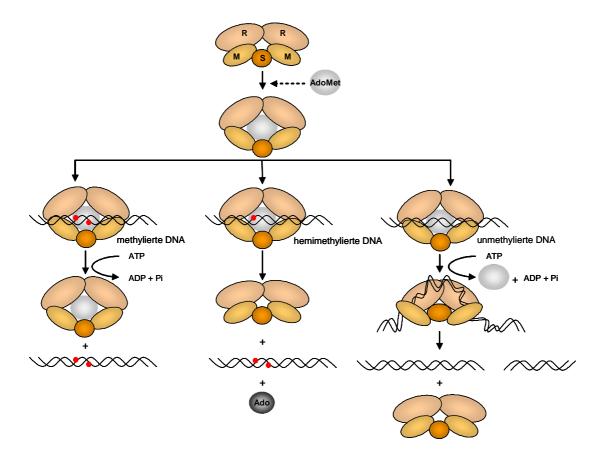


Abbildung 4.4.: Reaktionsschema des Typ I Restriktionssystems

Der Enzymkomplex besteht aus zwei HsdR-Untereinheiten (R), zwei HsdM-Untereinheiten (M) und einer HsdS-Untereinheit (S). Eine Aktivierung des Multienzyms mit S-Adenosylmethionin (AdoMet) bewirkt eine Konformationsänderung, wodurch die Bindung an DNA ermöglicht wird. Handelt es sich um methylierte DNA, wird diese unter Verbrauch von ATP wieder freigegeben. Eine Bindung an hemimethylierte DNA hingegen hat eine Aktivierung der Methylase-Funktion des Enzymkomplexes und eine vollständige Methylierung der DNA zur folge, wobei S-Adenosylhomocystein (Ado) freigesetzt wird. Bindet das Restriktionssystem an unmethylierte DNA, transloziert der Komplex unter Freigabe des S-Adenosyl-Methionins und ATP-Verbrauch entlang der DNA, wobei die Bindung an die Erkennungssequenz erhalten bleibt. Auf diese Weise entsteht ein DNA-Loop und der Doppelstrang wird mindestens 1000 bp entfernt von der Enzymbindesequenz geschnitten.

# 4.3. Verifizierung des genetischen Potentials von *B. licheniformis* mittels genomweiter Transkriptionsanalyse

Im Rahmen der Sequenzierung und Analyse des Genoms von *B. licheniformis* DSM13 konnten insgesamt 4.286 offene Leserahmen (ORFs) identifiziert werden. Darunter befindet sich eine Vielzahl von Genen, die das Wissen um das genetische Potential von *B. licheniformis* erweitern. Aufbauend darauf gilt es nun, die physiologische Umsetzung dieses Potentials zu untersuchen. Einen ersten Schritt in diese Richtung stellt die transkriptionsanalytische Untersuchung des Genoms von *B. licheniformis* dar. Diesbezüglich wurde ein DNA-Microarray erstellt, der 95% aller ORFs, größer als 300 Basenpaare, des Genoms von *B. licheniformis* 

abdeckt. Die Einsatzmöglichkeiten dieses Genomchips sind vielseitig und ermöglichen eine genomweite, qualitative Untersuchung der Genexpression unter unterschiedlichsten Bedingungen. In dieser Arbeit wurde der Schwerpunkt auf die Verwertung verschiedener Kohlenstoffquellen gelegt und die daraus resultierenden Transkriptionsänderungen analysiert. Die Expressionsänderungen der Gene, auf welche im Folgenden genauer eingegangen wird, sind für die jeweiligen Versuchsbedingungen in Tabelle 4.1. zusammengefasst.

# 4.3.1. Aspekte des C-Stoffwechsels

# 4.3.1.1. Genexpression unter Verwertung von C2-Substraten

Wie bereits beschrieben, konnten im Zuge der Annotation des Genoms von *B. licheniformis* DSM13 unter anderem zwei Gene mit hoher Homologie zu der Isocitrat-Lyase und der Malat-Synthase, den essentiellen Enzymen des Glyoxylatzyklus, identifiziert werden. Wachstumsversuche mit Acetat und 2,3-Butandiol, als jeweils einzige C-Quelle, bestätigten die Verwertung dieser Substrate durch *B. licheniformis*. Die transkriptionsanalytischen Untersuchungen von *B. licheniformis* Zellen, welche mit Acetat, bzw. 2,3-Butandiol, als einziger C-Quelle gewachsen sind, zeigten jeweils eine ähnliche Änderung des Expressionsmusters im Vergleich zum Wachstum mit Glucose (s. Abbildung 3.11. und Abbildung 3.12.). Diese Beobachtung deutet auf ein einheitliches Regulon, das von diesen beiden Substraten angesprochen wird.

Die Gene der Isocitrat-Lyase und der Malat-Synthase werden im sinnvollen Einklang mit dem Gen der Acetyl-CoA-Synthetase (acsA), für die Aktivierung von Acetat, und den Genen der Gluconeogenese, der Phosphoenolpyruvat-Carboxykinase (pckA) und der Glyceraldehyd-3-Phosphat-Dehydrogenase B (gapB), höher transkribiert. Die Transkription der glykolytischen Gene, der Triosephosphat-Isomerase (tpi) der Glyceraldehyd-3-Phosphat-Dehydrogenase A (gapA), der Phosphoglycerat-Kinase (pgk), der Phosphoglycerat-Mutase (pgm) und der Enolase (eno), hingegen wird, im Vergleich zur Glucoseverwertung, reprimiert. Die Regulation der Isocitrat-Lyase und der Malat-Synthase wurde in dem Gram-negativen Bakterium E. coli bereits eingehend untersucht. In diesem Organismus sind die Gene der Isocitrat-Lyase (aceA) und der Malat-Synthase (aceB) mit einem dritten Gen geclustert, welches für die Isocitrat-

**Tabelle 4.1.:** Transkriptionsänderungen einzelner Gene bei Verwertung von Acetat, 2,3-Butandiol und Gluconat jeweils im Vergleich zur Verwertung von Glucose und bei Wachstum unter unlimitierter N-Zugabe und Glucose-Limitierung im Vergleich zu Glucoseüberschuss und N-Limitierung. (↑ - min. 2 - 3fache Transkriptionserhöhung; ↓ - min. 2 - 3fache Transkriptionsabnahme; ↔ keine auffällige Transkriptionsänderung; n.d. - nicht bestimmt)

Gen-ID	Produkt	Funktion	Acetat/	2,3-Butan-	Gnt/	NH <sub>4</sub> CI/
			Glc	diol/ Glc	Glc	Glc
BLi00199	SigW	RNA Polymerase ECF-Type Sigmafaktor	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi00540	YdbM	unbekannt; ähnlich zu Butyryl-CoA-	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
		Dehydrogenase				
BLi00849	AcoA	Acetoin-Dehydrogenase E1 α-UE	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi00850	AcoB	Acetoin-Dehydrogenase E1 β-UE	<b>↑</b>	<b>↑</b>	n.d.	<b>↑</b>
BLi00851	AcoC	Acetoin-Dehydrogenase E2	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
		(dihydrolipoamide acetyltransferase)				
BLi01109	Subtilisin Carlsberg	(EC 3.4.21.62)	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi01122	SipV	Signalpeptidase I	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi01294	PbpE	Penicillin-bindendes Protein	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi01598	PtsG	Glucose-specifische Enzym II (PTS)	$\leftrightarrow$	$\leftrightarrow$	$\downarrow$	$\leftrightarrow$
BLi02106	putatives Protein	putative Butyryl-CoA-Dehydrogenase	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi03052	GapB	Glyceraldehyde-3-Phosphate-Dehydrogenase	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi03071	YtsJ	unbekannt; ähnlich zu Malate-Dehydrogenase	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03086	AckA	Acetate-Kinase	n.d.	n.d.	n.d.	n.d.
BLi03119	AcsA	Acetyl-CoA-Synthetase	<b>↑</b>	<b>↑</b>	<b>↑</b>	<b>↑</b>
BLi03120	AcuA	Acetoin-Dehydrogenase	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi03121	AcuB	Acetoin-Dehydrogenase	n.d.	n.d.	n.d.	n.d.
BLi03122	AcuC	Acetoin-Dehydrogenase	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
BLi03139	MalS	Malate-Dehydrogenase	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03197	PckA	Phosphoenolpyruvate-Carboxykinase	$\uparrow$	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi03464	YusJ	unbekannt; ähnlich zu Butyryl-CoA-	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
		Dehydrogenase				
BLi03466	YusL	unbekannt; ähnlich zu 3-Hydroxyacyl-CoA-	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	<b>↑</b>
		Dehydrogenase				
BLi03661	Eno	Enolase	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03662	Pgm	Phosphoglycerate Mutase	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03663	Трі	Triosephosphate Isomerase	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03664	Pgk	Phosphoglycerate Kinase	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03665	GapA	Glyceraldehyde-3-Phosphate-Dehydrogenase	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03847	AlsD	α-Acetolactat-Decarboxylase	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
BLi03848	AlsS	α-Acetolactat-Synthase	$\downarrow$	n.d.	$\downarrow$	n.d.
BLi03918	putatives Protein	putative Format-Dehydrogenase	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi03953	YwkA	unbekannt; ähnlich zu Malate-Dehydrogenase	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi03960	FbaA	Fructose-1,6-bisphosphate-Aldolase	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	<b>↑</b>
BLi03997	Pta	Pphosphotransacetylase	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi04137	MleA	potentielles Malat-Enzym	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
BLi04207	Isocitrate-Lyase	Isocitrate-Lyase (EC 4.1.3.1)	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi04208	Malate-Synthase	Malate-Synthase (EC 4.1.3.2)	<b>↑</b>	<b>↑</b>	$\leftrightarrow$	$\leftrightarrow$
BLi04286	GntR	Transkriptionsrepressor des Gluconatoperons	$\leftrightarrow$	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$
BLi04287	GntK	Gluconat-Kinase	$\leftrightarrow$	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$
BLi04288	GntP	Gluconat-Permease	$\leftrightarrow$	$\leftrightarrow$	<b>↑</b>	$\leftrightarrow$
BLi04289	GntZ	6-Phosphogluconat-Dehydrogenase	$\leftrightarrow$	$\leftrightarrow$	$\uparrow$	$\leftrightarrow$

Dehydrogenase-Kinase/Phosphatase (*aceK*) codiert (Chung et al., 1988; LaPorte and Chung, 1985; LaPorte et al., 1985). Das Operon *aceA-aceB-aceK* unterliegt einer komplexen Kontrolle durch den Repressor IclR, der wiederum von dem positiven Regulator des Fettsäureabbaus und der Fettsäuresynthese FadR kontrolliert wird (Gui et al., 1996; Henry and Cronan, 1991; Maloy and Nunn, 1982).

#### Regulation der Isocitrat-Dehydrogenase.

Während des Wachstums mit Acetat oder Fettsäuren als einziger C-Quelle ergibt sich eine Konkurrenz der Isocitrat-Lyase und der Isocitrat-Dehydrogenase um das Substrat Isocitrat. Eine elegante Lösung dieser Rivalität findet sich in *E. coli* durch das dritte Gen des *ace*-Operons, die Isocitrat-Dehydrogenase-Kinase/Phosphatase (AceK). Dieses Enzym katalysiert die Phosphorylierung und damit Deaktivierung der Isocitrat-Dehydrogenase (LaPorte, 1993; Sunnarborg et al., 1990) bei der Induktion des Glyoxylatzyklus. *B. licheniformis* weist kein Homolog zu AceK auf. Im Gegensatz zu *E. coli* wurde für das Archaeon *Haloferax volcanii* keine Änderung der Isocitrat-Dehydrogenase-Aktivität unter den entsprechenden Bedingungen festgestellt (Serrano et al., 1998). Hinsichtlich der gegensätzlichen Beobachtungen in *E. coli* und *H. volcanii* stellt sich die Frage, ob mit einer Induktion des Glyoxylsäurezyklus unbedingt eine Regulation der Aktivität der Isocitrat-Dehydrogenase verbunden ist. Im Rahmen der DNA-Microarray Analyse konnte auf Ebene der Transkription für *B. licheniformis* keine Regulation der Isocitrat-Dehydrogenase (Icd) festgestellt werden. In diesem Zusammenhang ist eine Untersuchung der Isocitrat-Dehydrogenase-Aktivität in Abhängigkeit von einem induzierten oder reprimierten Glyoxylatzyklus in *B. licheniformis* von Interesse.

### Induktion des Glyoxylatzyklus.

Der Gram-positive Organismus *Corynebacterium glutamicum* verfügt ebenfalls über die Gene des Glyoxylatzyklus, deren Regulation von den Arbeitsgruppen um B.J. Eikmanns und H. Sahm untersucht wird. Auch hier konnte eine Induktion der Isocitrat-Lyase und der Malat-Synthase durch Acetat als einziger C-Quelle, aber auch durch Acetat im Gemisch mit anderen C-Quellen gezeigt werden (Reinscheid et al., 1994a; Reinscheid et al., 1994b). Diese Ergebnisse stellen einen Gegensatz zu den Beobachtungen in Gram-negativen Bakterien dar. Hier wird, z.B. für das phototrophe Purpurbakterium *Rhodobacter capsulatus*, *Yersinia pestis* und *E. coli*, von einer exklusiven Induktion des Glyoxylatzyklus durch Acetat oder langkettige Fettsäuren als einziger C-Quelle berichtet (Blasco et al., 1991; Hillier and Charnetzky, 1981; Kornberg, 1966; Vanderwinkel et al., 1963). Ob für die Isocitrat-Lyase und die Malat-Synthase von *B. licheniformis* eine Induktion durch Acetat oder andere C2-Komponenten im Gemisch mit präferierten C-Quellen erfolgt, gilt es zu untersuchen.

#### Aktivierung von Acetat.

Unter Verwertung von Acetat kann die Synthese von Acetyl-CoA von der Acetat-Kinase (AckA) und der Phophotransacetylase (Pta) über das Zwischenprodukt Acetylphosphat katalysiert werden, oder auf direktem Wege durch die Acetyl-CoA-Synthetase (AcsA) erfolgen (Abbildung 4.5.). In *C. glutamicum* werden *ackA* und *pta* gemeinsam mit den Genen des Glyoxylatzyklus (*aceA*, *aceB*) induziert (Wendisch et al., 1997). Für die Regulation der Gene des Glyoxylatzyklus in *C. glutamicum* wurde ein neuer negativer Transkriptionsregulator (RamB) identifiziert, dessen hoch konserviertes 13-bp Bindemotive unter anderem vor den Genen *ackA*, *pta*, *aceA* und *aceB* lokalisiert ist (Gerstmeir et al., 2004). *B. licheniformis* weist weder ein homologes Protein zu RamB auf, noch deuten die durchgeführten Transkriptionsanalysen

**Abbildung 4.5.:** Aktivierung von Acetat durch die Acetat-Kinase (Ack) und die Phosphotransacetylase (Pta) (**A**) oder die Acetyl-CoA-Synthetase (**B**)

auf eine Aktivierung von Acetat durch AckA und Pta, sondern durch die Acetyl-CoA-Synthetase (AcsA) hin. In *E. coli* erfolgt die Acetataktivierung ebenfalls hauptsächlich durch die Acetyl-CoA-Synthetase (Brown et al., 1977; Oh et al., 2002). *B. subtilis* verfügt nicht über die Gene des Glyoxylatzyklus und ist nicht in der Lage, mit Acetat als einziger C-Quelle zu wachsen. Für eine Aktivierung von Acetat zu Acetyl-CoA ist hier aber offensichtlich auch die Acetyl-CoA-Synthetase, anstelle der Acetat-Kinase und der Phosphotransacetylase, zuständig (Grundy et al., 1993). Basierend auf den ersten Transkriptionsanalysen von *B. licheniformis* DSM13 sind Gemeinsamkeiten aber auch Unterschiede zu den bisher bekannten Regulationsmechanismen der Acetatverwertung zu finden. Hier bietet sich ein neuer Ansatz zur Aufklärung der diesbezüglichen regulatorischen Vorgänge in *B. licheniformis*.

#### Acetoinverwertung.

Während des Wachstums von B. licheniformis mit Acetat bzw. 2,3-Butandiol konnte im Zusammenhang mit einer Nutzung von Acetoin für die Gluconeogenese, eine erhöhte Transkription der Gene acoA, acoB und acoC, als auch der Gene acuA und acuC in B. licheniformis detektiert werden. Die beiden von den Genclustern aco und acu codierten Acetoin-Dehydrogenasen weisen in B. subtilis offensichtlich keinerlei Ähnlichkeit zueinander auf (Grundy et al., 1994; Grundy et al., 1993). Charakterisierungen des Acetoin-Stoffwechsels in B. subtilis zeigten, dass es sich bei der von den aco-Genen codierte Acetoin-Dehydrogenase um das essentielle Enzym des Acetoin-Abbaus handelt (Huang et al., 1999). Da eine defekte Acu-Acetoin-Dehydrogenase die Acetoinnutzung zwar einschränkt (Grundy et al., 1994; Grundy et al., 1993), eine nicht aktive Aco-Acetoin-Dehydrogenase die Verwertung von Acetoin jedoch zum Erliegen bringt (Huang et al., 1999). Der genaue Reaktionsmechanismus für die Synthese von Acetyl-CoA aus Acetoin ist bisher nicht bekannt. Es wird jedoch eine CoA abhängige Oxidation und Spaltung des Acetoins angenommen, an welcher die Reduktionsäquivalente FAD und NAD beteiligt sein könnten (Oppermann et al., 1991; Priefert et al., 1991) (Abbildung 4.6.). Essentiell für das Wachstum mit 2,3-Butandiol ist die 2,3-Butandiol-Dehydrogenase. Dieses Enzym katalysiert die Oxidation von 2,3-Butandiol zu Acetoin (Abbildung 4.6.). Im Rahmen der DNA-Microarrayanalyse des Wachstums mit 2,3-Butandiol konnte für fünf Gene mit einer Ähnlichkeit zu Dehydrogenasen eine erhöhte Transkription detektiert werden. Dabei handelt es sich um ydbM (BLi00540), yusJ (BLi03464), yusL (BLi03466), und zwei putative Gene (Bli02106 und BLi03918). Für diese Gene wurde während des Wachstums mit Acetat ebenfalls eine Erhöhung der Transkription beobachtet. Unter Verwertung von Gluconat erfolgt offenbar keine Regulation dieser Gene, bei Wachstum mit limitierter Glucosezugabe hingegen wurde die Transkription von yusJ, yusL, und dem putativen Gen Bli02106 ebenfalls induziert. Lediglich für YdbM und das putative Protein BLi03918 konnte in dieser Arbeit eine exklusive Induktion durch 2,3-Butandiol und Acetat beobachtete werden, womit es sich bei diesen beiden Genen um potentielle Kandidaten für die 2,3-Butandiol-Dehydrogenase handeln könnte. Diese Vermutung bedarf allerdings weiterer Untersuchungen der 2,3-Butandiolverwertung von B. licheniformis.

Abbildung 4.6.: Verwertung von 2,3-Butandiol und die Synthese von Acetyl-CoA.

#### Malat-Enzym.

Die Analyse der Genomsequenz von *B. subtilis* zeigte das Vorhandensein von vier paralogen Genen, welche für die potentiellen Malat-Enzyme *ywkA*, *malS*, *ytsJ* und *mleA* codieren (Doan et al., 2003; Kunst et al., 1997). In der Sequenz von *B. licheniformis*, ist für jedes dieser Gene ein Homolog zu finden. Transkriptionsanalysen ergaben für *B. subtilis* eine vergleichbare Genexpression von *malS* und *ytsJ* mit Malat oder Glucose, was die Autoren eine Regulation, unabhängig von Malat, als auch von dem glykolytischen und gluconeogenetischen Kohlenstoff-Stoffwechsel, vermuten lässt (Doan et al., 2003). In der vorliegenden Arbeit hingegen konnte eine erhöhte Transkription für das potentielle Malat-Enzym MalS bei Wachstum mit Acetat, aber nicht mit 2,3-Butandiol, gezeigt werden. Wobei es sich durchaus um einen Hinweis auf eine Rolle des MalS für die Synthese von Pyruvat aus Malat (Abbildung 4.7.) im Rahmen der Gluconeogenese handeln kann. Eine Bestätigung dieser Vermutung und auch eventuell einer Abhängigkeit der *malS* Transkription von Acetat bedarf jedoch weiterer Analysen.

Abbildung 4.7.: Katalyse der Pyruvatsynthese aus Malat durch das Malat-Enzym.

#### Subtilisin Carlsberg.

Gene, die unter Nutzung von 2,3-Butandiol, jedoch nicht unter Verwertung von Acetat, höher transkribiert werden, sind die des Subtilisins Carlsberg, des Sigmafaktors  $\sigma^W$ , des Penicillin-Bindeproteins (PbpE) und der Signalpeptidase SipV. Möglicherweise besteht ein Zusammenhang zwischen der gemeinsamen Höherexpression dieser Gene. Für das *B. subtilis* Subtilisin (AprE) konnte *in vitro* eine Kontrolle durch den Haupt-Sigmafaktor  $\sigma^A$  gezeigt werden (Park

et al., 1989). Der Sigmafaktor  $\sigma^W$  ist unter anderem in die Regulation einiger Gene involviert, deren Produkte der Detoxifikation dienen (Helmann and Moran Jr., 2002). Dazu gehören z.B. das Fosfomycin-Resistenzprotein (FosB), das Penicillin-Bindeprotein (PbpE), eine putative Erythromycin-Esterase (YbfO), eine Bromoperoxidase (YdjP), eine Epoxid-Hydrolase (YfhM) und zwei Protease IV Homologa (YteI, YqeZ) (Cao et al., 2002; Huang et al., 1999). Hypothetisch wäre es möglich, dass in *B. licheniformis* die Transkription des Subtilisin Carlsberg von  $\sigma^W$  vermittelt wird, oder aber gleichzeitig mit der Induktion der Serin-Protease ein Resistenzgen, wie z.B. PbpE, mit Hilfe von  $\sigma^W$  für den Selbstschutz induziert wird. Eine erhöhte Expression der Signalpeptidase SipV könnte wiederum mit der Exkretion des Subtilisin Carlsberg in Verbindung stehen. Um eine Verbindung dieser Gene herzustellen oder auszuschließen, sind weitere Untersuchungen von Interesse.

### 4.3.1.2. Vergleichende Transkriptionsanalyse bei C- und N-Limitierung

Eine weiterführende Transkriptionsanalyse wurde hinsichtlich der Genexpression von *B. licheniformis* DSM13 bei Wachstum unter Glucoselimitierung und unlimitierter Glucosezugabe durchgeführt. Für die Gene der Glycolyse zeigte sich hier keine bemerkenswerte Expressionsänderung (Abbildung 3.16.). Lediglich für das Gen der Fructose-1,6-Bisphosphat-Aldolase (*fbaA*) konnte eine Erhöhung der Expression unter Glucoselimitierung detektiert werden. Die Hintergründe einer Expressionsänderung von *fbaA* sind vorerst fraglich.

#### Acetyl-CoA-Synthetase.

Des Weiteren ließ sich eine deutliche Steigerung der Transkription des Acetyl-CoA-Synthetase Gens (*acsA*), im Vergleich zum unlimitierten Wachstum mit Glc, detektieren. Diese Beobachtung geht konform mit *B. subtilis* Studien, in welchen eine CcpA vermittelte Repression der Acetyl-CoA-Synthetase bei Glucoseüberschuss gezeigt wurde (Blencke et al., 2003; Grundy et al., 1994; Grundy et al., 1993; Hueck et al., 1994; Weickert and Chambliss, 1990).

#### Acetoin-Synthese.

Allgemein konnte unter unlimitierter Zugabe von Glucose, nicht nur im Vergleich zu Glucoselimitiertem Wachstum, sondern auch zum Wachstum mit Gluconat, Acetat bzw. 2,3-Butandiol, eine erhöhte Expression der Acetolactat-Synthase (alsS) und der Acetolactat-Decarboxylase (alsD) beobachtet werden. Die Funktion dieser Enzyme liegt in der Synthese von Acetoin aus Pyruvat über das Zwischenprodukt Acetolactat (Abbildung 4.8.). Für unter C-Überschuss angezogene Kulturen wird allgemein eine Entkopplung von Anabolismus und Katabolismus beschrieben, welche sich in Form eines so genannten Overflow-Metabolismus

äußert (Russell and Cook, 1995; Teixeira de Mattos and Neijssel, 1997). Eine besonders hohe Produktion von Metaboliten wird hier für B. subtilis vor allem unter Phosphat-Limitierung beschrieben (Dauner et al., 2001). Aber auch unter Stickstoff-Limitierung konnte im Vergleich zu einer Glucose-Limitierung eine erhöhte Ausbeute an Acetat, Acetoin, Diacyl und Riboflavin detektiert werden (Dauner et al., 2001). Für eine Induktion der Acetolactat-Synthase (AlsS) und der Acetolactat-Decarboxylase (AlsD) wird in B. licheniformis eine pH-Abhängigkeit vermutet, um auf diese Weise, durch eine erhöhte Acetoinsynthese aus Pyruvat eine Übersäuerung des Mediums durch Acetat zu vermeiden (Hornbaek et al., 2004). Eine gesteigerte Produktion von Acetoin und eine Aktivierung der alsD Transkription bei niedrigen pH-Werten wurden ebenfalls bei B. subtilis beobachtet (Amanullah et al., 2001; Presecan-Siedel et al., 1999; Turinsky et al., 2000). Die Induktion der Acetoin-Synthese-Gene, die in dieser Arbeit gemessen wurde, steht nicht in einem Zusammenhang mit einem absinkendem pH-Wert. Eine Transkriptionserhöhung für alsS und alsD wurde ausschließlich in Zellen, die in kontinuierlicher Kultur gewachsen sind, detektiert. Eine Übersäuerung des Mediums ist unter diesen Bedingungen unwahrscheinlich. Möglicherweise unterliegt die Regulation der Acetoin-Synthese mehreren komplexen Faktoren.

Abbildung 4.8.: Synthese von Acetoin und Butandiol aus Pyruvat.

#### Subtilisin Carlsberg.

Analog zu der Transkriptionsanalyse von Zellen, welche mit 2,3-Butandiol gewachsen sind, konnte eine erhöhte Expression der Gene des Subtilisin Carlsberg und der Signalpeptidase SipV unter Glucoselimitiertem Wachstum ermittelt werden (s. 4.2.1.1.). Eine wiederholte gemeinsame Expressionsänderung dieser Gene weist verstärkt auf einen funktionellen Zusammenhang hin.

#### Acetoinverwertung.

Wie bereits für das Wachstum mit Acetat, bzw. 2,3-Butandiol beschrieben (s. 4.2.1.1.), konnte unter Glucoselimitiertem Wachstum von *B. licheniformis* ebenfalls eine Expressionserhöhung für die Gene *acoA*, *acoB* und *acoC*, sowie für *acuA* und *acuC* ermittelt werden, welche jeweils für die Untereinheiten der zwei Acetoin-Dehydrogenasen Aco und Acu codieren. Im

Einklang mit diesen Ergebnissen zeigten transkriptionsanalytische Untersuchungen von *B. subtilis* unter Glucosemangelbedingungen für die Gene *acoA* und *acoC* ebenfalls eine erhöhte Expression (Yoshida et al., 2001). Des Weiteren ergaben diese Untersuchungen eine CcpA unabhängige Glucoserepression dieser Gene, welche offensichtlich eine wichtige Rolle in der Gluconeogenese einnehmen. Für *B. subtilis* wird eine Nutzung von Acetoin hauptsächlich während der stationären Phase im Zusammenhang mit der Sporulation angenommen (Lopez and Thoms, 1976). Im Gegensatz dazu stehen die Ergebnisse der Transkriptionsanalysen dieser Arbeit. Die verwendete RNA stammt jeweils aus Zellen, welche im Chemostat kontinuierlich in der exponentiellen Wachstumsphase gehalten wurden, oder aus Zellen der mittleren logarithmischen Phase einer Batch-Kultur. Insgesamt stehen eine Acetoin-Synthese unter Glucoseüberschuss und eine Nutzung von Acetoin unter Glucosemangel durchaus in einem sinnvollen Zusammenhang mit dem Bestreben des Organismus, ein optimales Wachstum unter bestmöglicher Ausnutzung aller zur Verfügung stehenden Ressourcen zu betreiben.

### 4.3.1.3. Vergleichende Transkriptionsanalyse bei Wachstum mit Gluconat bzw. Glucose als C-Quelle

B. licheniformis DSM13 weist in seinem Genom ein vollständiges Operon für die Nutzung von Gluconat auf, das eine auffallende Homologie zu dem korrespondierenden gntRKPZ Operon in B. subtilis besitzt (Yoshida et al., 1994). In B. subtilis unterliegt die Regulation dieses Operons der Katabolitrepression und wird von Gluconat induziert (Dowds et al., 1978; Nihashi and Fujita, 1984). In einer Arbeit von Y. Fujita et al. (Fujita et al., 1986) wird das B. subtilis gnt-Operon als das erste polycistronisch transkribierte Operon in Gram-positiven Bakterien vorgestellt. Die transkriptionsanalytische Untersuchung des Wachstums von B. licheniformis mit Gluconat als einziger C-Quelle im Vergleich zum Wachstum mit Glucose, ließ eine erhöhte Transkription der Gene der Gluconat-Permease (gntP), der Gluconat-Kinase (gntK) und des negativen Repressors des Gluconat-Operons (gntR) erkennen. Anhand einer quantitativen Expressionsanalyse mittels Real-time PCR konnte für die Gene gntK, gntP und gntZ eine 20fache Erhöhung der Transkription detektiert werden, für gntR lediglich eine 6,6fache. Offensichtlich erfolgt in B. licheniformis, analog zu den oben genannten Untersuchungen von B. subtilis, eine Induktion des Gluconat-Operons durch Gluconat in Abwesenheit von Glucose. Zusätzlich zu einer Induktion der Gluconat-Gene wurde eine starke Repression des Glucose-spezifischen Enzyms II des Phosphotransferasesystems, in Ermangelung des Induktors Glucose, vermerkt. Die Gene der Glycolyse und des Tricarbonsäurezyklus bleiben von dem Substratwechsel weitestgehend unbeeinflusst. Wie bereits für das Wachstum mit Acetat bzw. 2,3-Butandiol und die Glucose-Limitierung beschrieben (4.2.1.1. und 4.2.1.2.), werden während des Wachstums mit Gluconate im Vergleich zum Wachstum mit Glucose die Gene der Acetoin-Synthese (alsS, alsD) reprimiert und das Gen der Acetyl-CoA-Synthetase (acsA) aktiviert. Womit auch in diesem Fall auf einen Overflow-Metabolismus unter Glucose-überschuss hingedeutet wird.

# 4.3.2. Bestätigung hypothetischer und konserviert hypothetischer ORF's aufgrund transkriptionsanalytisch ermittelter Regulation

Insgesamt 636 Gene, d.h. 15% aller ORFs des B. licheniformis Genoms, wurden als hypothetisch oder konserviert hypothetisch eingestuft. Als einzige Belege für die Existenz dieser Gene können bisher ein Start- und Stoppcodon, eine potentielle Ribosomenbindestelle und eine sinnvolle Codon usage in dem eventuell codierenden Bereich gewertet werden. Eine funktionelle Bestätigung der hypothetischen Annotation kann nur mit Hilfe der Methoden der Molekularbiologie, Biochemie und Proteinchemie erfolgen. Einen ersten Ansatz hierfür und für eine funktionelle Einordnung der betreffenden ORFs bietet sich im Rahmen der Transkriptionsanalyse. Unter den in dieser Arbeit untersuchten Bedingungen konnte für 38 der hypothetischen oder konserviert hypothetischen Gene eine Regulation auf Ebene der Transkription bestimmt werden (s. Tabelle 3.17.). Der überwiegende Teil dieser Gene wurde unter Verwertung von 2,3-Butandiol bzw. Acetat im Vergleich zu Glucose induziert oder reprimiert. Diese Substrate induzierten, unter den in dieser Arbeit gegebenen Bedingungen, genomweit die umfangreichsten Expressionsänderungen, wie die Auflistungen in den Tabellen A.1. und A.2. im Anhang belegen und was sich auch bezüglich der regulierten hypothetischen ORFs widerspiegelt. Einschränkend muss jedoch in diesem Zusammenhang erwähnt werden, dass B. licheniformis für Acetat, bzw. 2,3-Butandiol und Glucose sehr stark abweichende Wachstumsraten aufweist (3.6.1.). Für die Untersuchungen mit Gluconat und limitierter, bzw. unlimitierter Zugabe von Glucose hingegen, wurde während der Anzucht in kontinuierlicher Kultur eine annähernd gleiche Wachstumsrate eingestellt. Starke Unterschiede in der Wachstumsrate stehen möglicherweise in einem Zusammenhang mit der Anzahl der regulierten Gene. Einige der, unter diesen Bedingungen regulierten, hypothetischen ORFs sind im Genom unmittelbar nebeneinander lokalisiert, so dass hier die Vermutung nahe liegt, dass es sich um potentielle Operone handelt. Eine Analyse der angrenzenden ORFs hinsichtlich eines funktionellen Hinweises blieb jedoch erfolglos (s. 3.7.6.). Für diese hypothetischen ORFs kann eine allosterische Regulation durch 2,3-Butandiol, bzw. Acetat angenommen und ihre Funktion in einen möglichen Zusammenhang mit deren Verwertung gestellt werden. Weitere Untersuchungen, z.B. auf transkriptioneller Ebene unter Verwertung unterschiedlichster Substrate, können weitere Indizien für die Aufgabe der entsprechenden Genprodukte liefern und hypothetische ORFs in ihrer Existenz bestätigen.

#### 4.4. Ausblick

Die vollständige Sequenzierung des Genoms von B. licheniformis DSM13 stellt eine nahezu unerschöpfliche Quelle für weitere Untersuchungen bezüglich der komplexen Funktionsweise dieses vielseitigen Organismus dar. Dies gilt insbesondere im Bezug zu dem bereits sehr gut untersuchten B. subtilis. Hier sind vor allem die Unterschiede der phylogenetisch nah verwandten Bacilli von großem Interesse. Zumal sich bereits nach einem "kurzen Blick" eine enorme Flexibilität von B. licheniformis hinsichtlich seiner Fähigkeiten unterschiedlichste Substrate, unter variierender Sauerstoffverfügbarkeit zu nutzen, zeigt. Die hier dargestellten Untersuchungen, lassen eine außergewöhnlich gute Anpassung an ökologische Nischen im Laufe der Evolution dieses Organismus erahnen, welche es näher zu beleuchten gilt. Eine weitere faszinierende Möglichkeit ergibt sich aus dem möglichen Vergleich des Wildtypstammes mit industriell genutzten, modifizierten Stämmen. Wo liegen Schwächen und Stärken der einzelnen Stämme? Wie lassen sich Produktionsabläufe und -bedingungen optimieren? Die optimale Voraussetzung für einen genomweiten Vergleich ist die vollständige Nukleotidsequenz kommerziell genutzter B. licheniformis Stämme. Aber bereits Hybridisierungen des B. licheniformis Genomchips mit Gesamt-DNA würden umfangreiche Informationen über das Vorhandensein oder Fehlen essentieller Gene liefern.

### 5. Zusammenfassung

### 5.1. Das Genomprojekt Bacillus licheniformis

- 1. Das vollständige Genom von *B. licheniformis* DSM13 wurde sequenziert. Bei einer Gesamtlänge von 4.222.748 bp stellen 87,9% die codierende Sequenz dar. Der durchschnittliche G+C Gehalt liegt bei 46,2%.
- 2. Insgesamt wurden 4.286 offene Leserahmen identifiziert. 32% davon konnte eine Funktion zugewiesen werden, 39,8% wurden analog zur Annotation von *Bacillus subtilis* als ,Y-Gene' annotiert. Weiteren 25,3% der Gene konnte keine eindeutige Annotation zugewiesen werden, diese wurden daher als putativ (10,4% aller ORFs), hypothetisch (11,3% aller ORFs) oder konserviert hypothetisch (3,6% aller ORFs) eingestuft. *B. licheniformis* besitzt des Weiteren sieben Cluster ribosomaler RNA und 72 Gene für transfer RNAs.
- 3. Das Chromosom von *B. licheniformis* DSM13 weist eine deutliche Co-Direktionalität der Replikation und Transkription auf. Deutlich wird dies, durch eine Lokalisierung von 74,3% aller ORFs auf dem *leading strand*.
- 4. Der G+C Skew des Genoms von *B. licheniformis* zeigt einen für Bacilli klassischen Verlauf. Der Replikationsursprung liegt vor dem Gen *dna*A, wobei die Umgebung des *ori*C die konservierte Anordnung der relevanten Gene und drei DnaA Bindeboxen analog zu *B. subtilis* aufweist.
- 5. Das Genom von *B. licheniformis* DSM13 enthält vier potentielle Prophagen. Gemeinsam mit insgesamt 263 ORFs wurden die Gene von drei der vier Phagen als Fremdgene identifiziert, welche durch horizontalen Gentransfer in das Chromosom gelangt sind. Der vierte Phage, welcher Homologien zu dem PBSX-Phagen von *B. subtilis* aufweist, scheint sehr viel länger in das Genom von *B. licheniformis* inseriert zu sein, möglicherweise bereits vor dem evolutionären Split von *B. subtilis* und *B. licheniformis*.

### 5.2. Komparative Genomanalyse von *B. licheniformis* DSM13 mit *B. subtilis*

- 1. 74,6% aller *B. licheniformis* ORFs zeigen auf Ebene der Aminosäuresequenz eine Identität von mindestens 30% zu *B. subtilis* ORFs, 43,9% sind zu mehr als 70% identisch und 14,8% weisen eine mehr als 85%ige Identität auf.
- 2. Auf Ebene der Nukleotidsequenz, ist eine hohe Co-Linearität der Genome von *B. licheniformis* und *B. subtilis* zu vermerken.
- 3. Eine Vielzahl an Exoenzymen wird im Genom von *B. licheniformis* codiert. 21 davon besitzen ein Homolog in *B. subtilis*. Insgesamt 17 Exoenzyme sind im Genom von *B. licheniformis*, aber nicht im Genom von *B. subtilis* zu finden. Dazu gehören das Subtilisin Carlsberg, die Glutamat-spezifische Protease, die maltogene α-Amylase, die Hitze- und pH-stabile α-Amylase und Chitin- und Celluloseabbauende Enzyme.
- 4. Die aus *B. subtilis* bekannten Bindemotive für die Sigmafaktoren SigD und SigL wurden vor den entsprechenden Genen im Genom von *B. licheniformis* identifiziert und weisen auf eine Konservierung regulatorischer Mechanismen hin.
- 5. 25,4% aller in *B. licheniformis* identifizierter Gene besitzen kein Homolog in *B. subtilis*. Bei 52% davon handelt es sich um hypothetische Gene. Die Genprodukte von 11%, der für *B. licheniformis* einzigartigen Gene im Vergleich zu *B. subtilis*, sind in Transportfunktionen, 8% in regulatorische Prozesse und 17% in verschiedenste enzymatische Vorgänge involviert.
- 6. Unter den ORFs, welche keine homologe Entsprechung in *B. subtilis* aufweisen, sind der nicht-ribosomale Peptidsynthetasekomplex der Lichenysin-Biosynthese zu finden, mehrere Gene, die potentiell in die Synthese eines Siderophores involviert sind, zwei Operonstrukturen, die für Typ I Restriktionsenzyme codieren, die Gene einer Isocitrat-Lyase und einer Malat-Synthase und Gene, die für eine anaerobe Ribonukleotidreduktase und ihr zugehöriges Aktivierungsenzym codieren.

### 5.3. Genomvergleich von *B. licheniformis* DSM13 mit anderen Bacilli der *B. subtilis* Gruppe II

- 1. Die Genome von *B. licheniformis*, *B. subtilis* und *B. halodurans* weisen keine Plasmide und jeweils ein einzelnes, im Durchschnitt eine Megabase kleineres Chromosom als *B. anthracis* Ames und *B. cereus* ATCC 10987 auf.
- 2. Der codierende Anteil der Sequenz der fünf Organismen liegt in einem Bereich von 84,3 bis 87,9%. Auf den Chromosomen von *B. licheniformis*, *B. subtilis* und *B. halodurans* sind durchschnittlich 1000 ORFs weniger als in *B. anthracis* Ames und *B. cereus* ATCC 10987 codiert.
- 3. Der durchschnittliche G+C Gehalt der Genome von *B. licheniformis*, *B. subtilis* und *B. halodurans* ist etwa 9% höher, als der von *B. anthracis* Ames und *B. cereus* ATCC 10987.
- 4. In *B. licheniformis* ist der Anteil an ribosomalen RNA Clustern und Genen für transfer RNAs im Vergleich zu den anderen vier Organismen am geringsten.
- 5. Basierend auf den allgemeinen Genomdaten lässt sich eine Unterteilung der fünf Organismen in die *B. subtilis/ B. licheniformis/ B. halodurans* und die *B. anthracis/ B. cereus*-Gruppe vornehmen.

# 5.4. Vergleich der Genome von *B. licheniformis*, *B. subtilis* und *B. halodurans* auf der Ebene der Aminosäuresequenz

1. Alle drei Organismen besitzen ein gemeinsames Kerngenom von 2.323 ORFs. 189 weitere ORFs haben *B. licheniformis* und *B. halodurans* gemeinsam, ohne dass diese eine Entsprechung in *B. subtilis* aufweisen. 146 ORFs von *B. subtilis* sind homolog zu ORFs von *B. halodurans* und besitzen kein Homologon in *B. licheniformis*. *B. subtilis* und *B. licheniformis* weisen 872 homologe ORFs zusätzlich zum Kerngenom auf. Insgesamt sind 902 ORFs für *B. licheniformis*, 1.408 ORFs für *B. halodurans* und 771 ORFs für *B. subtilis unique*.

### 5.5. Untersuchung des Wachstums von B. licheniformis

- 1. *B. licheniformis* DSM13 wächst anaerob mit Glucose und kann aerob sowohl Acetat, als auch 2,3-Butandiol als jeweils einzige C-Quelle verwerten, wofür eine anaerobe Ribonukleotidreduktase bzw. die Enzyme des Glyoxylatzyklus notwendig sind. Die dafür erforderlichen Gene konnten auf dem Genom von *B. licheniformis* detektiert werden.
- 2. Die Wachstumsraten von *B. licheniformis* wurden aerob in Minimalmedium unter Zugabe unterschiedlicher C-Quellen ermittelt. Unter Verwertung von Glucose liegt die Wachstumsrate bei 0,95/h, für Gluconat bei 0,68/h, für Xylose bei 0,3/h, für 2,3-Butandiol bei 0,15/h und für Acetat bei 0,13/h. Bei anaerobem Wachstum mit Glucose wurde eine Wachstumsrate von 0,045/h ermittelt.
- 3. Die Schwellenwerte der Konzentration von Glucose, Gluconat, NH<sub>4</sub>Cl und Xylose für limitiertes, bzw. unlimitiertes Wachstum im Rahmen der Anzucht von *B. licheniformis* in kontinuierlicher Kultur liegen bei 20 mM, 35 mM, 30 mM, bzw. 50 mM.

### 5.6. Transkriptionsanalytische Untersuchung von B. licheniformis

- 1. Basierend auf der Genomsequenz wurde ein DNA-Microarray hergestellt, welcher 95% aller *B. licheniformis* ORFs größer 300 bp abdeckt.
- 2. Die vergleichende Expressionsanalyse des Wachstums von *B. licheniformis* mit Acetat, bzw. 2,3-Butandiol im Vergleich zum Wachstum mit Glucose zeigte eine Induktion der Gene des Glyoxylatzyklus unter Anpassung an die Verwertung von C2-Substraten. Eine Quantifizierung der Expressionsänderungen bei Verwertung von Acetat im Vergleich zu Glucose ergab für die Isocitrat-Lyase einen Faktor 305,2 und für die Malat-Synthase einen Faktor 265,7.
- 3. Die Transkriptionsanalyse von *B. licheniformis* bei Wachstum auf Gluconat im Vergleich zu Glucose zeigte eine erhöhte Expression des Gluconat-Operons, wobei die Expression der Gene der Glykolyse und des Tricarbonsäurezyklus weitestgehend unbeeinflusst blieb. Die Quantifizierung der Expressionsänderung ergab für die Gluconat-Permease, die Gluconat-Kinase und die 6-Phosphogluconat-Dehydrogenase einen Faktor 20 und für den negativen Repressor des Gluconat-Operons einen Faktor 6,6.

4. Im Allgemeinen konnte unter Glucosemangelbedingungen eine Induktion der Acetyl-CoA-Synthetase und eine Repression der Acetolactat-Synthase und der Acetolactat-Decarboxylase festgestellt werden. Eine limitierte Zugabe von Glucose und die Verwertung von Acetat, bzw. 2,3-Butandiol hatten des Weiteren eine erhöhte Transkription der Aco- und der Acu-Acetoin-Dehydrogenase zur Folge. Diese Beobachtungen deuten auf eine erhöhte Acetoin-Synthese bei Glucose-Überschuss und eine verstärkte Nutzung von Acetat und Acetoin unter Glucosemangel hin.

### Literaturverzeichnis

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R., and Wahl, G. M. (1979). Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Methods Enzymol *68*, 220-242.

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridizationwith DNA probes. Proc Natl Acad Sci U S A 74, 5350-5354.

Amanullah, A., McFarlane, C. M., Emery, A. N., and Nienow, A. W. (2001). Scale-down model to simulate spatial pH variations in large scale bioreactors. Biotechnol Bioeng *73*, 390-399.

Arima, K., Kakinuma, A., and Tamura, G. (1968). Surfactin, a crystalline peptidelipid surfactant produced by Bacillus subtilis: isolation, characterization and its inhibition of fibrin clot formation. Biochem Biophys Res Commun *31*, 488-494.

Ash, C., Farrow, J. A. E., Wallbanks, S., and Collins, M. D. (1991). Phylogenetic heterogeneity of the genus bacillus revealed by comparative analysis of small subunit ribosomal RNA sequences. Lett Appl Microbiol *13*, 202-206.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., *et al.* (2004). The Pfam protein families database. Nucleic Acids Res *32 Database issue*, D138-141.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2004). GenBank: update. Nucleic Acids Res *32 Database issue*, D23-26.

Blasco, R., Cardenas, J., and Castillo, F. (1991). Regulation of isocitrate lyase in *Rhodobacter capsulatus*. Curr Microbiol 22, 73-76.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997). The complete genome sequence of Escherichia coli K-12. Science *277*, 1453-1474.

Blencke, H.-M., Homuth, G., Ludwig, H., Mäder, U., Hecker, M., and Stülke, J. (2003). Transcriptional profiling of Gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. Metab Eng *5*, 133-149.

Bonfield, J. K., Smith, K. F., and Staden, R. (1995). A new DNA sequence assembly program. Nucleic Acids Research 23, 4992–4999.

Brown, T. D., Jones-Mortimer, M. C., and Kornberg, H. L. (1977). The enzymic interconversion of acetate and acetyl-coenzyme A in Escherichia coli. J Gen Microbiol *102*, 327-336.

Brüggemann, H., Bäumer, S., Fricke, W. F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A., and Gottschalk, G. (2003). The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. Proc Natl Acad Sci USA *100*, 1316-1321.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., *et al.* (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science *273*, 1058-1073.

Cao, M., Kobel, P. A., Morshedi, M. M., Wu, M. F., Paddon, C., and Helmann, J. D. (2002). Defining the Bacillus subtilis sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches. J Mol Biol *316*, 443-457.

Carlson, C. R., and Kolsto, A. B. (1994). A small (2.4 Mb) Bacillus cereus chromosome corresponds to a conserved region of a larger (5.3 Mb) Bacillus cereus chromosome. Mol Microbiol *13*, 161-169.

Chell, R. M., and Sundaram, T. K. (1975). Isolation and characterization of isocitrate lyase and malate synthase from Bacillus stearothermophilus. Biochem Soc Trans 3, 303-306.

Chun, J., and Bae, K. S. (2000). Phylogenetic analysis of Bacillus subtilis and related taxa based on partial gyrA gene sequences. Antonie Van Leeuwenhoek 78, 123-127.

Chung, T., Klumpp, D. J., and LaPorte, D. C. (1988). Glyoxylate bypass operon of *Escherichia coli*: cloning and determination of the functional map. J Bacteriol *170*, 386-392.

Dauner, M., Storni, T., and Sauer, U. (2001). Bacillus subtilis metabolism and energetics in carbon-limited and excess-carbon chemostat culture. J Bacteriol *183*, 7308-7317.

David, M. H., Günther, H., and Röper (1987). Starch 39, 436.

Debarbouille, M., Gardan, R., Arnaud, M., and Rapoport, G. (1999). Role of bkdR, a transcriptional activator of the sigL-dependent isoleucine and valine degradation pathway in Bacillus subtilis. J Bacteriol *181*, 2059-2066.

Debarbouille, M., Martin-Verstraete, I., Kunst, F., and Rapoport, G. (1991). The Bacillus subtilis sigL gene encodes an equivalent of sigma 54 from gram-negative bacteria. Proc Natl Acad Sci U S A 88, 9092-9096.

Declerck, N., Machius, M., Wiegand, G., Huber, R., and Gaillardin, C. (2000). Probing structural determinants specifying high thermostability in *Bacillus licheniformis* alpha-amylase. J Mol Biol *301*, 1041-1057.

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res *30*, 2478-2483.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). Use of a cDNA microarray to analys gene expression patterns in human cancer. Nature Genet *14*, 457-460.

Doan, T., Servant, P., Tojo, S., Yamaguchi, H., Lerondel, G., Yoshida, K., Fujita, Y., and Aymerich, S. (2003). The Bacillus subtilis ywkA gene encodes a malic enzyme and its transcription is activated by the YufL/YufM two-component system in response to malate. Microbiology *149*, 2331-2343.

Dowds, B., Baxter, L., and McKillen, M. (1978). Biochim Biophys Acta 541, 18-34.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res *8*, 186–194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res *8*, 175-185.

Farabaugh, P. J. (1996). Programmed translational frameshifting. Annu Rev Genet 30, 507-528.

Ferrari, E., and Hoch, J. A. (1989). Genetics. In Biotechnology Handbooks, C. R. Harwood, ed. (New York, Plenum Publishing Corporation), pp. 57-72.

Fiechter, A. (1992). Biosurfactants: moving towards industrial applications. Trends Biotechnol *10*, 208-217.

Fiedler, S., and Wirth, R. (1988). Transformation of bacteria with plasmid DNA by electroporation. Anal Biochem *170*, 38–44.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., and et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science *269*, 496-512.

Folmsbee, M. J., McInerney, M. J., and Nagle, D. P. (2004). Anaerobic growth of Bacillus mojavensis and Bacillus subtilis requires deoxyribonucleosides or DNA. Appl Environ Microbiol *70*, 5252-5257.

Fox, G. E., Wisotzkey, J. D., and Jurtshuk, P. J. (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol *42*, 166–170.

Fox, S. L., and Bala, G. A. (2000). Production of surfactant from Bacillus subtilis ATCC 12332 using potato substrates. Bioresour Technol *75*, 235-240.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., and et al. (1995). The minimal gene complement of Mycoplasma genitalium. Science *270*, 397-403.

French, S. (1992). Consequences of replication fork movement through transcription units in vivo. Science *258*, 1362-1365.

Froyshov, O., and Laland, S. G. (1974). On the biosynthesis of bacitracin by a soluble enzyme complex from *Bacillus licheniformis*. Eur J Biochem *46*, 235-242.

Fujita, M. Q., Yoshikawa, H., and Ogasawara, N. (1989). Structure of the dnaA region of Pseudomonas putida: conservation among three bacteria, Bacillus subtilis, Escherichia coli and P. putida. Mol Gen Genet *215*, 381-387.

Fujita, Y., Fujita, T., Miwa, Y., Nihashi, J., and Aratani, Y. (1986). Organization and transcription of the gluconate operon, gnt, of Bacillus subtilis. J Biol Chem *261*, 13744-13753.

Gardan, R., Rapoport, G., and Debarbouille, M. (1997). Role of the transcriptional activator RocR in the arginine-degradation pathway of Bacillus subtilis. Mol Microbiol *24*, 825-837.

Gerstmeir, R., Cramer, A., Dangel, P., Schaffer, S., and Eikmanns, B. J. (2004). RamB, a novel transcriptional regulator of genes involved in acetate metabolism of Corynebacterium glutamicum. J Bacteriol *186*, 2798-2809.

Gonzalez-Pastor, J. E., Hobbs, E. C., and Losick, R. (2003). Cannibalism by sporulating bacteria. Science *301*, 510-513.

Gottschalk, G. (1986). Bacterial Metabolism, Second Edition edn (New York, Springer-Verlag).

Grangemard, I., Wallach, J., Maget-Dana, R., and Peypoux, F. (2001). Lichenysin: a more efficient cation chelator than surfactin. Appl Biochem Biotechnol *90*, 199-210.

Gray, G. L., Mainzer, S. E., Rey, M. W., Lamsa, M. H., Kindle, K. L., Carmona, C., and Requadt, C. (1986). Structural genes encoding the thermophilic alpha-amylases of *Bacillus stearothermophilus* and *Bacillus licheniformis*. J Bacteriol *166*, 635-643.

Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 26, 2286–2290.

Grossman, T. H., Tuckman, M., Ellestad, S., and Osburne, M. S. (1993). Isolation and characterization of Bacillus subtilis genes involved in siderophore biosynthesis: relationship between B. subtilis sfpo and Escherichia coli entD genes. J Bacteriol *175*, 6203-6211.

Grundy, F. J., Turinsky, A. J., and Henkin, T. M. (1994). Catabolite regulation of Bacillus subtilis acetate and acetoin utilization genes by CcpA. J Bacteriol *176*, 4527-4533.

Grundy, F. J., Waters, D. A., Takova, T. Y., and Henkin, T. M. (1993). Identification of genes involved in utilization of acetate and acetoin in Bacillus subtilis. Mol Microbiol *10*, 259-271.

Gui, L., Sunnarborg, A., and LaPorte, D. C. (1996). Regulated expression of a repressor protein: FadR activates *icIR*. J Bacteriol *178*, 4704-4709.

Hacker, J., and Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol *54*, 641-679.

Hall, M. (1996). Combinatorial theory, 2nd edn (New York, Wiley-Intersience).

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166, 557–580.

Helmann, J. D., Marquez, L. M., and Chamberlin, M. J. (1988). Cloning, sequencing, and disruption of the Bacillus subtilis sigma 28 gene. J Bacteriol *170*, 1568-1574.

Helmann, J. D., and Moran Jr., C. P. (2002). RNA Polymerase and sigma factors. In Bacillus subtilis and Ist Closest Relatives, A. L. Sonenshein, J. A. Hoch, and R. Losick, eds. (Washington, DC, ASM Press), pp. 289-312.

Henry, M. F., and Cronan, J. E., Jr. (1991). Escherichia coli transcription factor that both activates fatty acid synthesis and represses fatty acid degradation. J Mol Biol 222, 843-849.

Hillier, S., and Charnetzky, W. T. (1981). Glyoxylate bypass enzymes in Yersinia species and multiple forms of isocitrate lyase in Yersinia pestis. J Bacteriol *145*, 452-458.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C., and Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. Nucleic Acids Res *24*, 4420-4449.

Hori, H., and Osawa, S. (1987). Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol *4*, 445-472.

Hornbaek, T., Jakobsen, M., Dynesen, J., and Nielsen, A. K. (2004). Global transcription profiles and intracellular pH regulation measured in Bacillus licheniformis upon external pH upshifts. Arch Microbiol.

Huang, M., Oppermann-Sanio, F. B., and Steinbuchel, A. (1999). Biochemical and molecular characterization of the Bacillus subtilis acetoin catabolic pathway. J Bacteriol *181*, 3837-3841.

Hueck, C. J., Hillen, W., and Saier, M. H., Jr. (1994). Analysis of a cis-active sequence mediating catabolite repression in gram-positive bacteria. Res Microbiol *145*, 503-518.

Hulo, N., Sigrist, C. J., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004). Recent improvements to the PROSITE database. Nucleic Acids Res *32 Database issue*, D134-137.

Huynen, M. A., and Bork, P. (1998). Measuring genome evolution. Proc Natl Acad Sci USA 95, 5849-5856.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. Gene *96*.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., et al. (2003). Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature *423*, 87-91.

Jacobs, M., Eliasson, M., Uhlen, M., and Flock, J. I. (1985). Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res *13*, 8913-8926.

Javaheri, M., Jennemann, G. E., McInerney, M. J., and Knapp, R. M. (1985). Anaerobic production of a biosurfactant by *Bacillus licheniformis* JF-2. Appl Environ Microbiol *50*, 698-700.

Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H., and et al. (1992). Purification, characterization, cloning, and expression of a glutamic acid-specific protease from *Bacillus licheniformis* ATCC 14580. J Biol Chem *267*, 23782-23788.

Kandra, L., Gyemant, G., Remenyik, J., Hovanszki, G., and Liptak, A. (2002). Action pattern and subsite mapping of *Bacillus licheniformis* alpha-amylase (BLA) with modified maltooligosaccharide substrates. FEBS Lett *518*, 79-82.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., and Kawashima, K. e. a. (2000). Complete genome structure of the nitrogenfixing symbiotic bacterium Mesorhizobium loti. DNA Res *7*, 331-338.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., *et al.* (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). DNA Res *3*, 185-209.

Katz, E., and Demain, A. L. (1977). The peptide antibiotics of Bacillus: chemistry, biogenesis, and possible functions. Bacteriol Rev *41*, 449-474.

Kilstrup, M., and Kristiansen, K. N. (2000). Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer. Nucleic Acids Res 28, E55.

Kim, B. J., Lee, S. H., Lyu, M. A., Kim, S. J., Bai, G. H., Chae, G. T., Kim, E. C., Cha, C. Y., and Kook, Y. H. (1999). Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). J Clin Microbiol *37*, 1714–1720.

Kim, I. C. (1991), PhD thesis, Seoul.

Kim, I. C., Cha, J. H., Kim, J. R., Jang, S. Y., Seo, B. C., Cheong, T. K., Lee, D. S., Choi, Y. D., and Park, K. H. (1992). Catalytic properties of the cloned amylase from *Bacillus licheniformis*. J Biol Chem *267*, 22108-22114.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., *et al.* (2003). Essential Bacillus subtilis genes. Proc Natl Acad Sci U S A *100*, 4678-4683.

Konz, D., Doekel, S., and Marahiel, M. A. (1999). Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. J Bacteriol *181*, 133-140.

Koonin, E. V., Makarova, K. S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol, *55*, 709–742.

Kornberg, H. L. (1966). Anaplerotic sequences and their role in metabolism. In Essays in biochemestry, P. N. Campbell, and G. P. Greville, eds. (New York, Academic Press), pp. 1-31.

Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Bates, K., Browne, P., van den Broek, A., Cochrane, G., Duggan, K., Eberhardt, R., et al. (2004). The EMBL Nucleotide Sequence Database. Nucleic Acids Res *32 Database issue*, D27-30.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249-256.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome Biol *5*, R12.

Lai, X., Davis, F. C., Hespell, R. B., and Ingram, L. O. (1997). Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant Escherichia coli and identification of a putative binding region for disaccharides. Appl Environ Microbiol *63*, 355-363.

Lai, X., and Ingram, L. O. (1993). Cloning and sequencing of a cellobiose phosphotransferase operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. J Bacteriol *175*, 6441-6450.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lander, E. S., and Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics 2, 231–239.

Lapidus, A., Galleron, N., Andersen, J. T., Jorgensen, P. L., Ehrlich, S. D., and Sorokin, A. (2002). Colinear scaffold of the Bacillus licheniformis and Bacillus subtilis genomes and its use to compare their competence genes. FEMS Microbiol Lett *209*, 23-30.

LaPorte, D. C. (1993). The isocitrate dehydrogenase phosphorylation cycle: regulation and enzymology. J Cell Biochem *51*, 14-18.

LaPorte, D. C., and Chung, T. (1985). A single gene codes for the kinase and phosphatase which regulate isocitrate dehydrogenase. J Biol Chem *260*, 15291-15297.

LaPorte, D. C., Thorsness, P. E., and Koshland, D. E., Jr. (1985). Compensatory phosphorylation of isocitrate dehydrogenase. A mechanism for adaptation to the intracellular environment. J Biol Chem *260*, 10563-10568.

Lathe, W. C., 3rd, Snel, B., and Bork, P. (2000). Gene context conservation of a higher order than operons. Trends Biochem Sci 25, 474-479.

Lawrence, J. G., and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol *44*, 383-397.

Lazarevic, V., Margot, P., Soldo, B., and Karamata, D. (1992). Sequencing and analysis of the Bacillus subtilis lytRABC divergon: a regulatory unit encompassing the structural genes of the N-acetylmuramoyl-L-alanine amidase and its modifier. J Gen Microbiol *138 ( Pt 9)*, 1949-1961.

Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol 13, 660-665.

Lopez, J. M., and Thoms, B. (1976). [Relations between catabolite repression and sporulation in Bacillus subtilis (author's transl)]. Arch Microbiol *109*, 181-186.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research *25*, 955–964.

Lucchini, S., Thompson, A., and Hinton, J. C. D. (2001). Microarrays for microbiologists. Microbiology, 1403-1414.

Mahillon, J., and Chandler, M. (1998). Insertion sequences. Microbiol Mol Biol Rev 62, 725-774.

Maloy, S. R., and Nunn, W. D. (1982). Genetic regulation of the glyoxylate shunt in Escherichia coli K-12. J Bacteriol *149*, 173-180.

Manachini, P. L., Fortina, M. G., Levati, L., and Parini, C. (1998). Contribution to phenotypic and genotypic characterization of Bacillus licheniformis and description of new genomovars. Syst Appl Microbiol *21*, 520-529.

Margot, P., Mauel, C., and Karamata, D. (1994). The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168 cell wall hydrolase not involved in vegetative cell autolysis. Mol Microbiol *12*, 535-545.

Margot, P., Pagni, M., and Karamata, D. (1999). Bacillus subtilis 168 gene lytF encodes a gamma-D-glutamate-meso-diaminopimelate muropeptidase expressed by the alternative vegetative sigma factor, sigmaD. Microbiology *145* (*Pt 1*), 57-65.

Marquez, L. M., Helmann, J. D., Ferrari, E., Parker, H. M., Ordal, G. W., and Chamberlin, M. J. (1990). Studies of sigma D-dependent functions in Bacillus subtilis. J Bacteriol *172*, 3435-3443.

Martinez, J. L., Herrero, M., and de Lorenzo, V. (1994). The organization of intercistronic regions of the aerobactin operon of pCoIV-K30 may account for the differential expression of the *iucABCD iutA* genes. J Mol Biol *238*, 288-293.

McKinney, J. D., Höner zu Bentrup, K., Munzo-Elias, E. J., Miczak, A., Chen, B., W.-T., C., Swenson, D., Sacchettini, J. C., Jacobs, W. R., and Russell Jr., D. G. (2000). Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature 406, 735-738.

Merkl, R. (2004). Score-based identification of genomic islands. BMC Bioinformatics 5, 22.

Moran, A. C., Olivera, N., Commendatore, M., Esteves, J. L., and Sineriz, F. (2000). Enhancement of hydrocarbon waste biodegradation by addition of a biosurfactant from Bacillus subtilis O9. Biodegradation *11*, 65-71.

Moss, J. E., Cardozo, T. J., Zychlinsky, A., and Groisman, E. A. (1999). The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. Mol Microbiol *33*, 74-83.

Mountain, A. (1989). Gene expression systems in for *Bacillus subtilis*. In Biotechnology handbooks 2:Bacillus, C. R. Harwood, ed. (New York, Plenum Puplishing Corporation), pp. 73-114.

Mrazek, J., and Karlin, S. (1998). Strand compositional asymmetry in bacterial and large viral genomes. Proc Natl Acad Sci U S A *95*, 3720-3725.

Nihashi, J., and Fujita, Y. (1984). Catabolite repression of inositol dehydrogenase and gluconate kinase syntheses in Bacillus subtilis. Biochim Biophys Acta *798*, 88-95.

Ogasawara, N., Moriya, S., Mazza, P. G., and Yoshikawa, H. (1985). Conservation of genes and their organization in the chromosomal replication origin region of *Bacillus subtilis* and *Escherichia coli*. EMBO J *4*, 3345-3350.

Ogasawara, N., and Yoshikawa, H. (1986). Genes and their organization in the replication origin region of the Bacillus subtilis chromosome: comparison with the Escherichia coli chromosome. Adv Biophys *21*, 19-33.

Oh, M. K., Rohlin, L., Kao, K. C., and Liao, J. C. (2002). Global expression profiling of acetate-grown Escherichia coli. J Biol Chem *277*, 13175-13183.

Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Manter, D., *et al.* (1999a). Sequence, assembly and analysis of pX01 and pX02. J Appl Microbiol *87*, 261-262.

Okinaka, R. T., Cloud, K., Hampton, O., Hoffmaster, A. R., Hill, K. K., Keim, P., Koehler, T. M., Lamke, G., Kumano, S., Mahillon, J., *et al.* (1999b). Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. J Bacteriol *181*, 6509-6515.

Oppermann, F. B., Schmidt, B., and Steinbuchel, A. (1991). Purification and characterization of acetoin:2,6-dichlorophenolindophenol oxidoreductase, dihydrolipoamide dehydrogenase, and dihydrolipoamide acetyltransferase of the Pelobacter carbinolicus acetoin dehydrogenase enzyme system. J Bacteriol *173*, 757-767.

O'Sullivan, D., Twomey, D. P., Coffey, A., Hill, C., Fitzgerald, G. F., and Ross, R. P. (2000). Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. Mol Microbiol 36, 866-875.

Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., *et al.* (2003). The ERGO genome analysis and discovery system. Nucleic Acids Res *31*, 164-171.

Park, S. S., Wong, S. L., Wang, L. F., and Doi, R. H. (1989). Bacillus subtilis subtilisin gene (aprE) is expressed from a sigma A (sigma 43) promoter in vitro and in vivo. J Bacteriol *171*, 2657-2665.

Pearson, W. R. (1994). Using the FASTA program to search protein and DNA sequence databases. Methods Mol Biol *25*, 365-389.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A 85, 2444–2448.

Peypoux, F., Bonmatin, J. M., and Wallach, J. (1999). Recent trends in the biochemistry of surfactin. Appl Microbiol Biotechnol *51*, 553-563.

Pichinoty, F., Garcia, J. L., Job, C., and Durand, M. (1978). [Denitrification by Bacillus licheniformis]. Can J Microbiol *24*, 45-49.

Presecan-Siedel, E., Galinier, A., Longin, R., Deutscher, J., Danchin, A., Glaser, P., and Martin-Verstraete, I. (1999). Catabolite regulation of the *pta* gene as part of carbon flow pathways in *Bacillus subtilis*. J Bacteriol *181*, 6889-6897.

Priefert, H., Hein, S., Kruger, N., Zeh, K., Schmidt, B., and Steinbuchel, A. (1991). Identification and molecular characterization of the Alcaligenes eutrophus H16 aco operon genes involved in acetoin catabolism. J Bacteriol *173*, 4056-4071.

Priest, F. G. (1993). Systematics and Ecology of *Bacillus*. In *Bacillus subtilis* and other Gram-Positive Bacteria, J. A. SonensHoch, and R. h. Losick, A.L., eds. (Washington, DC, ASM Press), pp. 3-15.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. Microbiol Mol Biol Rev *62*, 597-635.

Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E., Tettelin, H., Fouts, D. E., Eisen, J. A., Gill, S. R., et al. (2003). The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature *423*, 81-86.

Reichard, P. (1993). The anaerobic ribonucleotide reductase from Escherichia coli. J Biol Chem 268, 8383-8386.

Reinscheid, D. J., Eikmanns, B. J., and Sahm, H. (1994a). Characterization of the isocitrate lyase gene from Corynebacterium glutamicum and biochemical analysis of the enzyme. J Bacteriol *176*, 3474-3483.

Reinscheid, D. J., Eikmanns, B. J., and Sahm, H. (1994b). Malate synthase from Corynebacterium glutamicum: sequence analysis of the gene and biochemical characterization of the enzyme. Microbiology *140 (Pt 11)*, 3099-3108.

Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., Lopez de Leon, A., Xiang, H., Gusti, V., Clausen, I. G., *et al.* (2004). Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species. Genome Biol *5*, R77.

Rocha, E. P., and Danchin, A. (2003). Gene essentiality determines chromosome organisation in bacteria. Nucleic Acids Res *31*, 6570-6577.

Rossler, D., Ludwig, W., Schleifer, K. H., Lin, C., McGill, T. J., Wisotzkey, J. D., Jurtshuk, P., Jr., and Fox, G. E. (1991). Phylogenetic diversity in the genus Bacillus as seen by 16S rRNA sequencing studies. Syst Appl Microbiol *14*, 266-269.

Russell, J. B., and Cook, G. M. (1995). Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiol Rev *59*, 48-62.

Salkinoja-Salonen, M. S., Vuorio, R., Andersson, M. A., Kämpfer, P., Andersson, M. C., Honkanen-Buzalski, T., and Scoging, A. C. (1999). Toxigenic Strains of *Bacillus licheniformis* Related To Food Poisoning. Applied and Environmental Microbiology, 4637-4645.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory).

Sandeman, R. A., Hynes, M. J., Fincham, J. R., and Connerton, I. F. (1991). Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and *Neurospora crassa*. Mol Gen Genet *228*, 445-452.

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins *9*, 56–68.

Sanger, F., Nicklen, S., and Coulson, A. R. (1992). DNA sequencing with chainterminating inhibitors. Biotechnology *24*, 104–108.

Schallmey, M., Singh, A., and Ward, O. P. (2004). Developments in the use of *Bacillus* species for industrial production. Can J Microbiol *50*, 1-17.

Segal, W. (1984). In The Mycobacteria: A Sourcebook, G. P. Kubica, and L. G. Wayne, eds. (New York, Dekker), pp. 547-573.

Serrano, J. A., Camacho, M., and Bonete, M. J. (1998). Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in Haloferax volcanii. FEBS Lett *434*, 13-16.

Shalon, D., Smith, S. J., and Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescentprobe hybridization. Genome Res *6*, 639-645.

Sharma, V., Sharma, S., Hoener zu Bentrup, K., McKinney, J. D., Russell, D. G., Jacobs, W. R., Jr., and Sacchettini, J. C. (2000). Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. Nat Struct Biol *7*, 663-668.

Shena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., and Hayashi, H. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. Proc Natl Acad Sci USA *99*, 996-1001.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., *et al.* (1997). Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics. J Bacteriol *179*, 7135-7155.

Smith, E. L., DeLange, R. J., Evans, W. H., Landon, M., and Markland, F. S. (1968). Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'; evolutionary relationships. J Biol Chem *243*, 2184-2191.

Sneath, P. H. A., Mair, N. S., Sharpe, M. E., and Holt, J. G. (1986). Bergey 's manual of sytematic bacteriology, Vol 2 (Baltimore, Williams & Wilkins).

Sonenshein, A. L. (2002). The Krebs Citric Acid Cycle. In *Bacillus subtilis* and Its Closest Relatives, A. L. Sonenshein, J. A. Hoch, and R. Losick, eds. (Washington, DC, American Society for Microbiology), pp. 151-161.

Sonenshein, A. L., Hoch, J. A., and Losick, R. (1993). Bacillus subtilis and other Gram-Positive Bacteria. In (american Society for Mikrobiology).

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol *98*, 503-517.

Spizizen, J. (1958). Transformation of biochemically deficient strains of *Bacillus subtilis* in deoxyribonucleate. Proc Natl Acad Sci U S A *44*, 1072-1078.

Staden, R., Beal, K. F., and Bonfield, J. K. (2000). The Staden package, 1998. Methods Mol Biol 132, 115-130.

Steinmetz, M. (1993). Carbohydrate catabolism: pathways, enzyms, genetic regulation and evolution. In *Bacillus subtilis* and Other Gram-Positive Bacteria, A. L. Sonenshein, J. A. Hoch, and R. Losick, eds. (Washington, DC, Am. Soc. Microbiol.), pp. 157-170.

Stephens, M. A., Ortlepp, S. A., Ollington, J. F., and McConnell, D. J. (1984). Nucleotide sequence of the 5' region of the *Bacillus licheniformis* alpha-amylase gene: comparison with the *B. amyloliquefaciens* gene. J Bacteriol *158*, 369-372.

Stülke, J., and Hillen, W. (2000). Regulation of Carbon Catabolism in *Bacillus* Spezies. Annu Rev Microbiol *54*, 849-880.

Sun, X., Eliasson, R., Pontis, E., Andersson, J., Buist, G., Sjoberg, B. M., and Reichard, P. (1995). Generation of the glycyl radical of the anaerobic Escherichia coli ribonucleotide reductase requires a specific activating enzyme. J Biol Chem *270*, 2443-2446.

Sunnarborg, A., D., K., Chung, T., and LaPorte, D. C. (1990). Regulation of the glyoxylate bypass operon: cloning and characterization of *icIR*. J Bacteriol *172*, 2642-2649.

Suzuki, Y., and Imai, T. (1985). Appl Microbiol Biotechnol 21, 20.

Svendsen, I., and Breddam, K. (1992). Isolation and amino acid sequence of a glutamic acid specific endopeptidase from *Bacillus licheniformis*. Eur J Biochem *204*, 165-171.

Syed, R., Wu, Z. P., Hogle, J. M., and Hilvert, D. (1993). Crystal structure of selenosubtilisin at 2.0-A resolution. Biochemistry *32*, 6157-6164.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., et al. (2000). Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res 28, 4317-4331.

Talaat, A. M., Howard, S. T., Hale VI, W., Lyons, R., Garner, H., and Johnston, S. A. (2002). Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. Nucleic Acids Research *30*, 1-9.

Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997). A genomic perspective on protein families. Science *278*, 631–637.

Tech, M., and Merkl, R. (2003). YACOP: Enhanced gene prediction obtained by a combination of existing methods. In Silico Biol 3, 441-451.

Teixeira de Mattos, M. J., and Neijssel, O. M. (1997). Bioenergetic consequences of microbial adaptation to low-nutrient environments. J Biotechnol *59*, 117-126.

Tettelin, H., Radune, D., Kasif, S., Khouri, H., and Salzberg, S. L. (1999). Optimized Multiplex PCR: Efficiently Closing a Whole-Genome Shotgun Sequencing Project. Genomics *62*, 500-507.

Tjalsma, H., Antelmann, H., Jongbloed, J. D., Braun, P. G., Darmon, E., Dorenbos, R., Dubois, J. Y., Westers, H., Zanen, G., Quax, W. J., *et al.* (2004). Proteomics of protein secretion by Bacillus subtilis: separating the "secrets" of the secretome. Microbiol Mol Biol Rev *68*, 207-233.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S., and van Dijl, J. M. (2000). Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. Microbiol Mol Biol Rev *64*, 515-547.

Tjalsma, H., Bolhuis, A., van Roosmalen, M. L., Wiegert, T., Schumann, W., Broekhuizen, C. P., Quax, W. J., Venema, G., Bron, S., and van Dijl, J. M. (1998). Functional analysis of the secretory precursor processing machinery of Bacillus subtilis: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases. Genes Dev *12*, 2318-2331.

Tjalsma, H., Noback, M. A., Bron, S., Venema, G., Yamane, K., and van Dijl, J. M. (1997). Bacillus subtilis contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different sip genes. J Biol Chem 272, 25983-25992.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., *et al.* (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. Nature *388*, 539-547.

Torrents, E., Buist, G., Liu, A., Eliasson, R., Kok, J., Gibert, I., Graslund, A., and Reichard, P. (2000). The anaerobic (class III) ribonucleotide reductase from *Lactococcus lactis*. Catalytic properties and allosteric regulation of the pure enzyme system. J Biol Chem *275*, 2463-2471.

Tschech, A., and Pfennig, N. (1984). Growth yield increase linked to caffeate reduction in *Acetobacte-rium woodii*. Arch Microbiol *137*, 163-167.

Turinsky, A. J., Moir-Blais, T. R., Grundy, F. J., and Henkin, T. M. (2000). *Bacillus subtilis ccpA* gene mutants specifically defective in activation of acetoin biosynthesis. J Bacteriol *182*, 5611-5614.

Vanderwinkel, E., Liard, P., Ramos, F., and Wiame, J. M. (1963). Genetic control of the regulation of isocitritase and malate synthase in Escherichia coli K 12. Biochem Biophys Res Commun *12*, 157-162.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K. H., Ehrenreich, P., Baumer, S., Henne, A., Liesegang, H., Merkl, R., et al. (2004). The complete genome sequence of Bacillus licheniformis DSM13, an organism with great industrial potential. J Mol Microbiol Biotechnol *7*, 204-211.

Velculescu, V. E., Zhang, L., Zhou, W., Vogelszein, J., Basrai, M. A., Basset, D. E. J., Hieter, B., Vogelstein, B., and Kinzler, K. W. (1997). Characterization of the yeast transcriptome. Cell 88, 243-251.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001). The sequence of the human genome. Science 291, 1304-1351.

Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19, 259-268.

Watanabe, T., Oyanagi, W., Suzuki, K., and Tanaka, H. (1990). Chitinase system of Bacillus circulans WL-12 and importance of chitinase A1 in chitin degradation. J Bacteriol *172*, 4017-4022.

Waterman, M. S. (1984). Efficient sequence alignment algorithms. J Theor Biol 108, 333–337.

Weickert, M. J., and Chambliss, G. H. (1990). Site-directed mutagenesis of a catabolite repression operator sequence in Bacillus subtilis. Proc Natl Acad Sci U S A 87, 6238-6242.

Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., *et al.* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proc Natl Acad Sci U S A *99*, 17020-17024.

Wendisch, V. F., Spies, M., Reinscheid, D. J., Schnicke, S., Sahm, H., and Eikmanns, B. J. (1997). Regulation of acetate metabolism in Corynebacterium glutamicum: transcriptional control of the isocitrate lyase and malate synthase genes. Arch Microbiol *168*, 262-269.

Wilson, G. G., and Murray, N. E. (1991). Restriction and modification systems. Annu Rev Genet *25*, 585-627.

Xu, D., and Cote, J. C. (2003). Phylogenetic relationships between Bacillus species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. Int J Syst Evol Microbiol *53*, 695-704.

Yakimov, M. M., Timmis, K. N., Wray, V., and Fredrickson, H. L. (1995). Characterization of a new lipopeptide surfactant produced by thermotolerant and halotolerant subsurface Bacillus licheniformis BAS50. Appl Environ Microbiol *61*, 1706-1713.

Yamamoto, S., Bouvet, P. J., and Harayama, S. (1999). Phylogenetic structures of the genus Acineto-bacter based on gyrB sequences: comparison with the grouping by DNA-DNA hybridization. Int J Syst Bacteriol *49*, 87–95.

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C. M., Matsunaga, M., Yamaguchi, H., Tojo, S., Yamamoto, M., Nishi, R., Ogasawara, N., *et al.* (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in Bacillus subtilis. Nucleic Acids Res *29*, 683-692.

Yoshida, K., Seki, S., and Fujita, Y. (1994). Nucleotide sequence and features of the Bacillus licheniformis gnt operon. DNA Res *1*, 157-162.

Yoshikawa, H., and Ogasawara, N. (1991). Structure and function of DnaA and the DnaA-box in eubacteria: evolutionary relationships of bacterial replication origin. Mol Microbiol *5*, 2589-2597.

Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N., and Udaka, S. (1985). Complete nucleotide sequence of a gene coding for heat- and pH-stable alpha-amylase of *Bacillus licheniformis*: comparison of the amino acid sequences of three bacterial liquefying alpha-amylases deduced from the DNA sequences. J Biochem (Tokyo) *98*, 1147-1156.

Zukowski, M. (1989). Production of commercially valuable products. In Biotechnology handbook, C. R. Harwood, ed. (New York, Plenum Publishing Corporation), pp. 311-337.

### A. Anhang

# A.1. Expressionsdaten ausgewählter *B. licheniformis* Gene der Transkriptionsanalyse Acetat gegen Glucose

**Tabelle A.1.:** Expressionsdaten weiterer Gene, welche mindestens um einen Faktor 3 beim Vergleich von Wachstum mit Acetat und Glucose reguliert werden. Die Expressionsfaktoren sind jeweils als Verhältnis der Expressionswerte bei Wachstum mit Acetat zu den Expressionsdaten bei Wachstum mit Glucose angegeben (Acetat/Glucose). Gene deren Expression unter Verwertung von Acetat erhöht ist, sind grau unterlegt.

Gen-ID	Produkt	Proteinbeschreibung	Acetat/
			Glucose
BLi04291	AhpC	alkyl hydroperoxide reductase (small subunit)	0,02
BLi04292	AhpF	alkyl hydroperoxide reductase (large subunit) / NADH	0,03
		dehydrogenase	
BLi01141	AmyD	sugar transport	10,13
BLi01224	AppD	oligopeptide ABC transporter (ATP-binding protein)	0,26
BLi02185	AprX	intracellular alkaline serine protease	7,68
BLi01270	CotX	spore coat protein (insoluble fraction)	151,98
BLi01269	CotY	spore coat protein (insoluble fraction)	265,26
BLi01268	CotZ	spore coat protein (insoluble fraction)	40,03
BLi02910	CoxA	spore cortex protein	10,49
BLi01974	CwIC	N-acetylmuramoyl-L-alanine amidase cwlL precursor (EC 3.5.1.28) (Cell	69,58
		wall hydrolase) (Autolysin)	
BLi04133	CydB	cytochrome bd ubiquinol oxidase (subunit II)	6,72
BLi04132	CydC	ABC transporter required for expression of cytochrome bd (ATP-binding	18,25
		protein)	
BLi04131	CydD	ABC transporter required for expression of cytochrome bd	16,31
		(ATP-binding protein)	
BLi02807	Des	membrane phospholipid desaturase	21,76
BLi02319	DfrA	dihydrofolate reductase	0,28
BLi03902	DhbA	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	3,63
BLi02999	EtfA	electron transfer flavoprotein (alpha subunit)	32,72
BLi03000	EtfB	electron transfer flavoprotein (beta subunit)	20,72
BLi03519	FhuC	ferrichrome ABC transporter (ATP-binding protein)	13,59
BLi02658	FhuD	ferrichrome ABC transporter (ferrichrome-binding protein)	5,65
BLi01163	GerPC	probable spore germination protein	85,87
BLi01161	GerPE	probable spore germination protein	86,13
BLi01992	GlnR	transcriptional repressor of the glutamine synthetase gene	0,07
BLi02740	GrpE	heat-shock protein (activation of DnaK)	0,31
BLi04256	HtpG	class III heat-shock protein (molecular chaperone)	0,24
BLi04113	KatA	vegetative catalase 1	0,32
BLi04196	KatE1	catalase 2	0,23
BLi01924	Kbl	2-amino-3-ketobutyrate CoA ligase	12,14

BLi01143	MelA	alpha-D-galactoside galactohydrolase	12,11
BLi01143	MmgA		9,40
		acetyl-CoA acetyltransferase	
BLi03966 BLi04095	MmgC	acyl-CoA dehydrogenase	15,91
	MmgE	function unknown	6,32
BLi03545	MntB	manganese ABC transporter (ATP-binding protein)	0,03
BLi03546	MntD	manganese ABC transporter	0,02
BLi00526	MntH	manganese transporter	0,13
BLi03480	MrgA	metalloregulation DNA-binding stress protein	0,16
BLi01140	MsmE	multiple sugar-binding protein	3,51
BLi00485	NasE	assimilatory nitrite reductase (subunit)	0,21
BLi02408	Ndk	nucleoside diphosphate kinase	4,88
BLi00900	PerR	transcriptional repressor of the peroxide regulon	0,22
BLi01072	PrsA	protein secretion (post-translocation molecular chaperone)	0,31
BLi04341	PurA	adenylosuccinate synthetase	0,09
BLi00695	PurB	adenylosuccinate lyase	0,09
BLi00696	PurC	phosphoribosylaminoimidazole succinocarboxamide synthetase	0,14
BLi00694	PurK	phosphoribosylaminoimidazole carboxylase II	0,22
BLi00698	PurQ	phosphoribosylformylglycinamidine synthetase I	0,29
BLi00206	putative protein	putative transporter	6,86
BLi00723	putative protein	putative phosphotriesterase protein	11,00
BLi01175	putative protein	putative ammonium transporter	0,12
BLi01176	putative protein	putative nitrogen regulatory protein P-II	0,07
BLi01465	putative protein	putative portal protein	0,16
BLi02102	putative protein	putative enoyl(3-hydroxyisobutyryl)-coenzyme A hydratase protein (EC	10,78
		4.2.1.17)	
BLi02103	putative protein	putative enoyl-CoA hydratase (EC 4.2.1.17)	4,00
BLi02106	putative protein	putative butyryl-CoA dehydrogenase	17,58
BLi02978	putative protein	putative response regulator aspartate phosphatase	0,29
BLi0340	putative protein	putative response regulator	0,30
BLi03918	putative protein	putative formate dehydrogenase alpha chain (EC 1.2.1.2)	8,69
BLi03993	putative protein	putative decarboxylase/dehydratase	0,09
BLi04117	putative protein	putative ABC transporter ATP-binding protein	10,60
BLi01771	PyrAA	carbamoyl-phosphate synthetase (glutaminase subunit)	0,24
BLi01768	PyrP	uracil permease	0,28
BLi02446	SerA	phosphoglycerate dehydrogenase	0,23
BLi02594	SpolVB	intercompartmental signalling of pro-sigma-K processing/activation in the	22,18
		mother-cell	
BLi02348	Xpt	xanthine phosphoribosyltransferase	0,20
BLi00045	YaaT	unknown; similar to unknown proteins	0,33
BLi00061	YabJ	unknown; similar to translation initiation inhibitor	0,30
BLi00446	YbfA	unknown	0,23
BLi00479	Ycnl	unknown; similar to unknown proteins	4,32
BLi00480	YcnJ	unknown; similar to copper export protein	4,21
BLi00540	YdbM	unknown; similar to butyryl-CoA dehydrogenase	3,78
	YdfO	unknown; similar to unknown proteins	0,13

BLi00978	YhcN	unknown; similar to unknown proteins	24,06
BLi01019	YhdL	unknown	0,24
BLi01058	YheA	unknown; similar to unknown proteins	6,57
BLi03946	YhjL		0,32
		unknown; similar to sensory transduction pleiotropic regulatory protein	
BLi01497	YkoX	unknown; similar to alkaline phosphatase	0,09
BLi04260	YkrX	unknown; similar to unknown proteins	0,22
BLi01608	YkuA	unknown; similar to penicillin-binding protein	0,22
BLi02213	YocH	unknown; similar to cell wall-binding protein	4,27
BLi01308	YoeB	unknown	0,32
BLi02264	YojL	unknown; similar to cell wall-binding protein	3,18
BLi02270	YozR	unknown	75,55
BLi02430	YpeB	unknown; similar to unknown proteins	10,92
BLi02424	YphA	unknown; similar to unknown proteins	0,30
BLi02359	YpqE	unknown; similar to phosphotransferase system enzyme II	0,27
BLi02551	YqjM	unknown; similar to NADH-dependent flavin oxidoreductase	0,29
BLi03002	YsiA	unknown; similar to transcriptional regulator (TetR/AcrR family)	14,10
BLi03098	Ytcl	unknown; similar to acetate-CoA ligase	3,53
BLi03177	YttB	unknown; similar to multidrug resistance protein	9,92
BLi03124	YtxD	unknown; similar to flagellar motor apparatus	4,20
BLi03123	YtxE	unknown; similar to motility protein	10,77
BLi03999	YuaB	unknown; similar to unknown proteins from B. subtilis	0,18
BLi03451	YurY	unknown; similar to ABC transporter (ATP-binding protein)	24,36
BLi03464	YusJ	unknown; similar to butyryl-CoA dehydrogenase	52,38
BLi03465	YusK	unknown; similar to acetyl-CoA C-acyltransferase	7,74
BLi03466	YusL	unknown; similar to 3-hydroxyacyl-CoA dehydrogenase	25,12
BLi03472	YusV	unknown; similar to iron(III) dicitrate transport permease	11,73
BLi03741	YvpA	unknown; similar to pectate lyase	23,74
BLi03495	YvqG	unknown; similar to unknown proteins	0,21
BLi03496	YvqH	unknown; similar to unknown proteins from B. subtilis	0,13
BLi03969	YwjF	unknown; similar to iron-sulphur-binding reductase	19,41
BLi03696	YwrO	unknown; similar to NAD(P)H oxidoreductase	0,19
BLi04262	YxeB	unknown; similar to ABC transporter (binding protein)	29,39
BLi04217	Yxel	unknown; similar to penicillin amidase	25,17
BLi00935	YxiD	unknown; similar to unknown proteins	0,33

# A.2. Expressionsdaten ausgewählter *B. licheniformis* Gene der Transkriptionsanalyse 2,3-Butandiol gegen Glucose

**Tabelle A.2.:** Expressionsdaten weiterer Gene, welche mindestens um einen Faktor 3 beim Vergleich von Wachstum mit 2,3-Butandiol und Glucose reguliert werden. Die Expressionsfaktoren sind jeweils als Verhältnis der Expressionswerte bei Wachstum mit 2,3-Butandiol zu den Expressionsdaten bei Wachstum mit Glucose angegeben (2,3-Butandiol/Glucose). Gene deren Expression unter Verwertung von 2,3-Butandiol erhöht ist, sind grau unterlegt.

Gen-ID	Produkt	Produktbeschreibung	Butandiol/
			Glucose
BLi03965	AcdA	acyl-CoA dehydrogenase	11,05
BLi00852	AcoL	acetoin dehydrogenase E3 component (dihydrolipoamide	19,1
		dehydrogenase)	
BLi00853	AcoR	transcriptional activator of the acetoin dehydrogenase operon	7,58
BLi04291	AhpC	alkyl hydroperoxide reductase (small subunit)	0,05
BLi04292	AhpF	alkyl hydroperoxide reductase (large subunit)	0,09
		/ NADH dehydrogenase	
BLi01141	AmyD	sugar transport	5,03
BLi01208	ArgB	N-acetylglutamate 5-phosphotransferase	7,55
BLi01206	ArgC	N-acetylglutamate gamma-semialdehyde dehydrogenase	8,62
BLi01209	ArgD	N-acetylornithine aminotransferase	4,69
BLi03084	ArgG	argininosuccinate synthase	5,29
BLi03083	ArgH	argininosuccinate lyase	5,07
BLi01207	ArgJ	ornithine acetyltransferase / amino-acid acetyltransferase	5,56
BLi02585	Bcd	leucine dehydrogenase	6,63
BLi02582	BkdAA	branched-chain alpha-keto acid dehydrogenase E1 subunit (2-	7,03
		oxoisovalerate dehydrogenase alpha subunit)	
BLi02581	BkdAB	branched-chain alpha-keto acid dehydrogenase E1 subunit (2-	14,26
		oxoisovalerate dehydrogenase beta subunit)	
BLi02580	BkdB	branched-chain alpha-keto acid dehydrogenase E2 subunit (lipoamide	10,3
		acyltransferase)	
BLi02584	Buk	probable branched-chain fatty-acid kinase (butyrate kinase)	5,22
BLi01210	CarA	carbamoyl-phosphate transferase-arginine (subunit A)	4,07
BLi01270	CotX	spore coat protein (insoluble fraction)	36,9
BLi01269	CotY	spore coat protein (insoluble fraction)	113,56
BLi01268	CotZ	spore coat protein (insoluble fraction)	20,77
BLi01709	CtaF	cytochrome caa3 oxidase (subunit IV)	0,295
BLi01710	CtaG	function unknown	3,21
BLi01974	CwlC	N-acetylmuramoyl-L-alanine amidase cwlL precursor	15,96
		(EC 3.5.1.28) (Cell wall hydrolase) (Autolysin)	
BLi04132	CydC	ABC transporter required for expression of cytochrome bd	9,14
		(ATP-binding protein)	
BLi03567	СурХ	cytochrome P450-like enzyme	0,06
BLi02807	Des	membrane phospholipid desaturase	52,72
BLi02999	EtfA	electron transfer flavoprotein (alpha subunit)	24,14

BLi03000	EtfB	electron transfer flavoprotein (beta subunit)	20,06
BLi01811	FabD	malonyl CoA-acyl carrier protein transacylase	3,28
BLi01812	FabG	beta-ketoacyl-acyl carrier protein reductase	3,11
BLi01097	FabHB	beta-ketoacyl-acyl carrier protein synthase III	7,84
BLi03519	FhuC	ferrichrome ABC transporter (ATP-binding protein)	16,32
BLi02658	FhuD	ferrichrome ABC transporter (ferrichrome-binding protein)	4,7
BLi02602	FolD	methylenetetrahydrofolate dehydrogenase / methenyltetrahydrofolate	0,32
		cyclohydrolase	
BLi01163	GerPC	probable spore germination protein	36,69
BLi01161	GerPE	probable spore germination protein	33,13
BLi01992	GlnR	transcriptional repressor of the glutamine synthetase gene	0,15
BLi00996	GlpD	glycerol-3-phosphate dehydrogenase	6,73
BLi00994	GlpF	glycerol uptake facilitator	8,55
BLi04254	GlpQ	glycerophosphoryl diester phosphodiesterase	27,53
BLi02162	GltA	glutamate synthase (large subunit)	0,26
BLi02161	GltB	glutamate synthase (small subunit)	0,33
BLi03732	HisF	HisF cyclase-like protein	0,25
BLi03737	HisG	ATP phosphoribosyltransferase	0,29
BLi03734	HisH	amidotransferase	0,23
BLi02806	HxIR	positive regulator of hxIAB expression	4,78
BLi01423	IspA	major intracellular serine protease	3,89
BLi04113	KatA	vegetative catalase 1	0,3
BLi01924	Kbl	2-amino-3-ketobutyrate CoA ligase	8,96
BLi04201	LicT	transcriptional antiterminator required for substrate-dependent induc-	3,07
		tion and catabolite repression of bgIPH	
BLi02583	LpdV	probable branched-chain alpha-keto acid dehydrogenase E3 subunit	6,6
		(dihydrolipoamide dehydrogenase)	
BLi01143	MelA	alpha-D-galactoside galactohydrolase	4,52
BLi04094	MmgD	citrate synthase III	6,1
BLi02104	MmsA	methylmalonate-semialdehyde dehydrogenase	14,6
BLi03547	MntA	manganese ABC transporter (membrane protein)	0,04
BLi03545	MntB	manganese ABC transporter (ATP-binding protein)	0,11
BLi03546	MntD	manganese ABC transporter	0,05
BLi00526	MntH	manganese transporter	0,15
BLi00483	NasC	assimilatory nitrate reductase (catalytic subunit)	0,22
BLi00485	NasE	assimilatory nitrite reductase (subunit)	0,14
BLi03889	NrgB	nitrogen-regulated PII-like protein	0,08
BLi03651	OpuCA		0,27
		glycine betaine/carnitine/choline ABC transporter (ATP-binding protein)	,
BLi03649	OpuCC	glycine betaine/carnitine/choline ABC transporter (osmoprotectant-	0,26
		binding protein)	, -
BLi01294	PbpE	penicillin-binding protein 4* (spore cortex)	4,34
BLi00900	PerR	transcriptional repressor of the peroxide regulon	0,29
BLi03822	Pmi	mannose-6-phosphate isomerase	5,08
BLi02586	Ptb	probable phosphate butyryltransferase	6,86
0_00		p. 1.1. 2.10 p. 1.00 p	5,00

BLi03436	PucR	transcriptional regulator of puc genes	0,14
BLi04341	PurA	adenylosuccinate synthetase	0,11
BLi00695	PurB	adenylosuccinate lyase	0,1
BLi00696	PurC	phosphoribosylaminoimidazole succinocarboxamide	0,11
		synthetase	
BLi00704	PurD	phosphoribosylglycinamide synthetase	0,26
BLi00693	PurE	phosphoribosylaminoimidazole carboxylase I	0,14
BLi00700	PurF	glutamine phosphoribosylpyrophosphate amidotransferase	0,19
BLi00703	PurH	phosphoribosylaminoimidazole carboxy formyl formyltransferase / ino-	0,21
		sine-monophosphate cyclohydrolase	
BLi00694	PurK	phosphoribosylaminoimidazole carboxylase II	0,08
BLi00699	PurL	phosphoribosylformylglycinamidine synthetase II	0,15
BLi00701	PurM	phosphoribosylaminoimidazole synthetase	0,24
BLi00702	PurN	phosphoribosylglycinamide formyltransferase	0,21
BLi00698	PurQ	phosphoribosylformylglycinamidine synthetase I	0,33
BLi00206	putative protein	putative transporter	6,07
BLi00488	putative protein	putative cotransporter	3,97
BLi00647	putative protein	putative transcriptional regulator	9,9
BLi00807	putative protein	putative lipopolysaccharide biosynthesis glycosyltransferase	10,2
BLi01038	putative protein	putative acetyltransferase (EC 2.3.1)	0,28
BLi01063	putative protein	putative response regulator aspartate phosphatase A(EC 3.1)	5,59
BLi01127	putative protein	putative permease	4,27
BLi01175	putative protein	putative ammonium transporter	0,07
BLi01176	putative protein	putative nitrogen regulatory protein P-II	0,04
BLi01188	putative protein	putative siderophore biosynthesis protein lucC	3,61
BLi01470	putative protein	putative phage capsid protein	0,07
BLi01531	putative protein	putative integrase/recombinase	0,3
BLi02102	putative protein	putative enoyl(3-hydroxyisobutyryl)-coenzyme A hydratase protein (EC	15,2
		4.2.1.17)	
BLi02103	putative protein	putative enoyl-CoA hydratase (EC 4.2.1.17)	5,11
BLi02105	putative protein	putative 2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60)	7,43
BLi02106	putative protein	putative butyryl-CoA dehydrogenase	18,06
BLi02819	putative protein	putative oxidoreductase protein (EC 1)	0,1
BLi02978	putative protein	putative response regulator aspartate phosphatase	0,33
BLi03498	putative protein	putative pectin methylesterase	8,86
BLi03622	putative protein	putative phage protein	3,79
BLi03657	putative protein	putative iron(III) transporter binding protein	14,1
BLi03918	putative protein	putative formate dehydrogenase alpha chain (EC 1.2.1.2)	5,04
BLi04117	putative protein	putative ABC transporter ATP-binding protein	9,16
BLi04351	putative protein	putative 3-oxoacyl- acyl-carrier protein reductase	9,66
BLi01771	PyrAA	carbamoyl-phosphate synthetase (glutaminase subunit)	0,19
BLi01772	PyrAB	carbamoyl-phosphate synthetase (catalytic subunit)	0,26
BLi01773	PyrK	dihydroorotate dehydrogenase (electron transfer subunit)	0,27
BLi01768	PyrP	uracil permease	0,32
BLi04037	QoxD	cytochrome aa3 quinol oxidase (subunit IV)	0,26

BLi02446	SerA	phosphoglycerate dehydrogenase	0,14
BLi00199	SigW	RNA polymerase ECF-type sigma factor	3,36
BLi01122	SipV	signal peptidase I	4,55
BLi01109	subtilisin	(EC 3.4.21.62)	5,34
	carlsberg		
BLi02348	Xpt	xanthine phosphoribosyltransferase	0,17
BLi00061	YabJ	unknown; similar to translation initiation inhibitor	0,29
BLi03730	YbbJ	unknown; similar to unknown proteins	0,27
BLi00200	YbbM	unknown; similar to unknown proteins	4,21
BLi00446	YbfA	unknown	0,12
BLi00275	YcbA		0,25
		unknown; similar to two-component sensor histidine kinase [YcbB]	
BLi02810	YcgT	unknown; similar to thioredoxin reductase	7,57
BLi00464	YclO	unknown; similar to ferrichrome ABC transporter	5,16
		(permease)	
BLi00465	YclP	unknown; similar to ferrichrome ABC transporter	6,66
		(ATP-binding protein)	
BLi00466	YclQ	unknown; similar to ferrichrome ABC transporter	4,75
		(binding protein)	
BLi00513	YdaH	unknown	0,06
BLi00540	YdbM	unknown; similar to butyryl-CoA dehydrogenase	4,33
BLi02024	YdfO	unknown; similar to unknown proteins	0,1
BLi02514	YdgC	unknown; similar to unknown proteins	0,33
BLi01365	YesL	unknown; similar to unknown proteins from B. subtilis	5,84
BLi03475	YfiY	unknown; similar to iron(III) dicitrate transport permease	7,67
BLi00945	YgaO	unknown	5,09
BLi01062	YhaR	unknown; similar to enoyl CoA hydratase	4,84
BLi00978	YhcN	unknown; similar to unknown proteins	12,1
BLi01290	YjcJ	unknown; similar to cystathionine beta-lyase	0,33
BLi01948	YjiC	unknown; similar to macrolide glycosyltransferase	0,3
BLi02090	YkkB	unknown; similar to N-acetyltransferase	3,02
BLi01648	YknW	unknown; similar to unknown proteins	4,05
BLi01493	YkoM	unknown; similar to transcriptional regulator (MarR family)	3,32
BLi01497	YkoX	unknown; similar to alkaline phosphatase	0,07
BLi01617	YkuD	unknown; similar to unknown proteins	2,96
BLi01714	YlbD	unknown; similar to unknown proteins	3,47
BLi01727	YlbP	unknown; similar to unknown proteins	4,16
BLi02485	YndL	unknown; similar to phage-related replication protein	48,45
BLi02809	YocG	unknown; similar to two-component response regulator [YocF]	6,91
BLi02291	YodN	unknown	4,08
BLi00529	YojK	unknown; similar to macrolide glycosyltransferase	3,12
BLi02270	YozR	unknown	49,46
BLi02317	YplQ	unknown; similar to hemolysin III homolog	3,03
BLi02312	YpmQ	unknown; similar to unknown proteins	3,28
BLi02684	YqfX	unknown; similar to unknown proteins	11,89

BLi02578	YqiW	unknown; similar to unknown proteins from B. subtilis	3,59
BLi02577	YqiX	unknown; similar to amino acid ABC transporter	5,1
		(binding protein)	
BLi03002	YsiA	unknown; similar to transcriptional regulator (TetR/AcrR family)	17,11
BLi03093	Ytdl	unknown; similar to unknown proteins	3,64
BLi03081	YtkK	unknown; similar to 3-oxoacyl- acyl-carrier protein reductase	9,71
BLi03123	YtxE	unknown; similar to motility protein	9,22
BLi03999	YuaB	unknown; similar to unknown proteins from B. subtilis	0,33
BLi03451	YurY	unknown; similar to ABC transporter (ATP-binding protein)	13,97
BLi03464	YusJ	unknown; similar to butyryl-CoA dehydrogenase	48,84
BLi03466	YusL	unknown; similar to 3-hydroxyacyl-CoA dehydrogenase	58,08
BLi03472	YusV	unknown; similar to iron(III) dicitrate transport permease	11,23
BLi00213	YvcC	unknown; similar to ABC transporter (ATP-binding protein)	6,3
BLi03715	YvdA	unknown; similar to carbonic anhydrase	10,2
BLi04298	YvfR	unknown; similar to ABC transporter (ATP-binding protein)	41,63
BLi04299	YvfS1	unknown; similar to ABC transporter transmembrane subunit	26,03
BLi04303	YvfT		11,66
		unknown; similar to two-component sensor histidine kinase [YvfU]	
BLi03752	YvIB	unknown; similar to unknown proteins	3,29
BLi03569	YvnA	unknown; similar to unknown proteins from B. subtilis	0,04
BLi04350	YvoA	unknown; similar to transcriptional regulator (GntR family)	32,55
BLi03741	YvpA	unknown; similar to pectate lyase	12,1
BLi03495	YvqG	unknown; similar to unknown proteins	0,3
BLi03497	YvqI	unknown	0,24
BLi03969	YwjF	unknown; similar to iron-sulphur-binding reductase	14,4
BLi04262	YxeB	unknown; similar to ABC transporter (binding protein)	50,24
BLi04217	Yxel	unknown; similar to penicillin amidase	27,55
BLi04187	YxjG	unknown; similar to unknown proteins from B. subtilis	0,32
BLi04092	YxlJ	unknown; similar to DNA-3-methyladenine glycosidase	0,21

# A.3. Expressionsdaten ausgewählter *B. licheniformis* Gene der Transkriptionsanalyse N-unlimitiert/Glucose-limitiert gegen Glucose-unlimitiert/N-limitiert

**Tabelle A.4.:** Expressionsdaten weiterer Gene, welche mindestens um einen Faktor 3 beim Vergleich von Stickstoff-unlimitiertem Wachstum und Glucose-unlimitiertem Wachstum reguliert werden. Die Expressionswerte werden jeweils als Verhältnis der Expressionswerte bei Stickstoff-unlimitiertem Wachstum zu den Expressionsdaten bei Glucose-unlimitiertem Wachstum angegeben (NH<sub>4</sub>Cl/Glucose). Gene deren Expression bei Stickstoff-unlimitiertem Wachstum erhöht ist, sind grau unterlegt.

Gen-ID	Produkt	Proteinbeschreibung	NH <sub>4</sub> CI/
			Glucose
BLi03965	AcdA	acyl-CoA dehydrogenase	5,7
BLi03849	AlsR	transcriptional regulator of the alpha-acetolactate operon	0,31
BLi01396	DppE	dipeptide ABC transporter (dipeptide-binding protein)	3,09
		(sporulation)	
BLi02999	EtfA	electron transfer flavoprotein (alpha subunit)	3,63
BLi03000	EtfB	electron transfer flavoprotein (beta subunit)	3,62
BLi01992	GlnR	transcriptional repressor of the glutamine synthetase gene	0,22
BLi01423	IspA	major intracellular serine protease	3,05
BLi01924	Kbl	2-amino-3-ketobutyrate CoA ligase	6,13
BLi00855	MalA	6-phospho-alpha-glucosidase	4,69
BLi03968	MmgA	acetyl-CoA acetyltransferase	3,54
BLi04094	MmgD	citrate synthase III	8,43
BLi04095	MmgE	function unknown	5,56
BLi02104	MmsA	methylmalonate-semialdehyde dehydrogenase	10,75
BLi00485	NasE	assimilatory nitrite reductase (subunit)	0,19
BLi00962	PrkA	serine protein kinase	4,39
BLi03436	PucR	transcriptional regulator of puc genes	0,15
BLi04341	PurA	adenylosuccinate synthetase	0,28
BLi00694	PurK	phosphoribosylaminoimidazole carboxylase II	0,33
BLi00488	putative protein	putative cotransporter	7,89
BLi01063	putative protein	putative response regulator aspartate phosphatase A	3,75
		(EC 3.1)	
BLi01175	putative protein	putative ammonium transporter	0,09
BLi01176	putative protein	putative nitrogen regulatory protein P-II	0,09
BLi01309	putative protein	putative cell wall-binding protein	0,24
BLi01470	putative protein	putative phage capsid protein	0,08
BLi02012	putative protein	putative acetyltransferase	0,26
BLi02102	putative protein	putative enoyl(3-hydroxyisobutyryl)-coenzyme A hydratase protein	7,19
		(EC 4.2.1.17)	
BLi02105	putative protein	putative 2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60)	7,84
BLi02106	putative protein	putative butyryl-CoA dehydrogenase	16,59
BLi03498	putative protein	putative pectin methylesterase	8,53

BLi03918	putative protein	putative formate dehydrogenase alpha chain (EC 1.2.1.2)	3,31
BLi03009	RnhC	ribonuclease HIII	0,33
BLi04018	SacT	transcriptional antiterminator involved in positive regulation of sacA	5,36
		and sacP	
BLi02446	SerA	phosphoglycerate dehydrogenase	0,3
BLi01122	SipV	signal peptidase I	5,47
BLi02607	SpollIAH	mutants block sporulation after engulfment	31,38
BLi01109	subtilisin	(EC 3.4.21.62)	11,42
	carlsberg		
BLi02637	TasA	translocation-dependent antimicrobial spore component	3,71
BLi02348	Xpt	xanthine phosphoribosyltransferase	0,32
BLi02024	YdfO	unknown; similar to unknown proteins	0,11
BLi00856	YfiA	unknown; similar to unknown proteins	4,27
BLi00629	YfmQ	unknown	0,24
BLi03945	YhjK	unknown; similar to unknown proteins	12,63
BLi03946	YhjL	unknown; similar to sensory transduction pleiotropic regulatory pro-	4,59
		tein	
BLi01493	YkoM	unknown; similar to transcriptional regulator (MarR family)	5,22
BLi04260	YkrX	unknown; similar to unknown proteins	0,27
BLi03123	YtxE	unknown; similar to motility protein	6,11
BLi03464	YusJ	unknown; similar to butyryl-CoA dehydrogenase	4,86
BLi03466	YusL	unknown; similar to 3-hydroxyacyl-CoA dehydrogenase	8,67
BLi04280	YvfK	unknown; similar to maltose/maltodextrin-binding protein	4,49
BLi04135	YwcJ	unknown; similar to nitrite transporter	0,19
BLi03969	YwjF	unknown; similar to iron-sulphur-binding reductase	5,46

### A.4. Hochexprimierte Gene in *B. licheniformis*

	·
ORF-Nr.	Annotation
BLi00029	YaaK unknown; similar to unknown proteins
BLi00050	AbrB: transcriptional pleiotropic regulator of transition state genes
BLi00062	SpoVG required for spore cortex synthesis
BLi00120	RpIK ribosomal protein L11 (BL11)
BLi00121	RpIA ribosomal protein L1 (BL1)
BLi00122	RpIJ ribosomal protein L10 (BL5)
BLi00123	RpIL ribosomal protein L12 (BL9)
BLi00128	RpsL ribosomal protein S12 (BS12)
BLi00129	RpsG ribosomal protein S7 (BS7)
BLi00130	FusA: elongation factor G
BLi00131	TufA elongation factor Tu
BLi00132	RpsJ ribosomal protein S10 (BS13)
BLi00133	RpIC ribosomal protein L3 (BL3)
BLi00134	RpID ribosomal protein L4
BLi00135	RpIW ribosomal protein L23
BLi00136	RpIB ribosomal protein L2 (BL2)
BLi00137	RpsS ribosomal protein S19 (BS19)
BLi00138	RpIV ribosomal protein L22 (BL17)
BLi00139	RpsC ribosomal protein S3 (BS3)
BLi00140	RpIP ribosomal protein L16
BLi00141	RpmC ribosomal protein L29
BLi00142	RpsQ ribosomal protein S17 (BS16)
BLi00143	RpIN ribosomal protein L14
BLi00144	RpIX ribosomal protein L24 (BL23) (histone-like protein HPB12)
BLi00145	RpIE ribosomal protein L5 (BL6)
BLi00146	RpsN ribosomal protein S14
BLi00147	RpsH ribosomal protein S8 (BS8)
BLi00148	RpIF ribosomal protein L6 (BL8)
BLi00149	RpIR ribosomal protein L18
BLi00150	RpsE ribosomal protein S5
BLi00151	RpmD ribosomal protein L30 (BL27)
BLi00152	RpIO ribosomal protein L15
BLi00157	InfA: initiation factor IF-1
BLi00159	RpsM ribosomal protein S13
BLi00160	RpsK ribosomal protein S11 (BS11)
BLi00161	RpoA RNA polymerase (alpha subunit)
BLi00162	RpIQ ribosomal protein L17 (BL15)
BLi00167	RpIM ribosomal protein L13
BLi00168	Rpsl ribosomal protein S9
BLi00534	GsiB: general stress protein
BLi00573	putative permease

BLi00578	CspC: cold-shock protein			
BLi00623	GroES: class I heat-shock protein (chaperonin)			
BLi00624	GroEL: class I heat-shock protein (chaperonin)			
BLi00681	hypothetical			
BLi00888	putative small acid-soluble spore protein			
BLi00932	hypothetical			
BLi00971	CspB: major cold-shock protein			
BLi01050	SspB small acid-soluble spore protein (major beta-type SASP)			
BLi01073	conserved hypothetical			
BLi01074	hypothetical			
BLi01128	conserved hypothetical			
BLi01131	conserved hypothetical			
BLi01132	conserved hypothetical			
BLi01389	hypothetical			
BLi01430	conserved hypothetical			
BLi01431	hypothetical			
BLi01470	putative phage capsid protein			
BLi01477	putative phage tail protein			
BLi01487 YrkC unknown; similar to unknown proteins				
BLi01536	1536 conserved hypothetical			
BLi01551	hypothetical			
BLi01556	putative phage tail protein			
BLi01558	hypothetical			
BLi01565	YomP unknown; similar to phage-related protein			
BLi01599	PtsH histidine-containing phosphocarrier protein of the PTS (HPr protein)			
BLi01607	YkwD unknown; similar to unknown proteins from B. subtilis			
BLi01635	YkuU unknown; similar to 2-cys peroxiredoxin			
BLi01669	YkzG unknown; similar to unknown proteins			
BLi01674	PdhA pyruvate dehydrogenase (E1 alpha subunit)			
BLi01675	PdhB pyruvate dehydrogenase (E1 beta subunit)			
BLi01676	PdhC pyruvate dehydrogenase (dihydrolipoamide acetyltransferase E2 subunit)			
BLi01677	PdhD pyruvate dehydrogenase / 2-oxoglutarate dehydrogenase			
	(dihydrolipoamide dehydrogenase E3 subunit)			
BLi01701	YlaN unknown; similar to unknown proteins			
BLi01725	RpmF ribosomal protein L32			
BLi01803	RpmB ribosomal protein L28			
BLi01813	AcpA: acyl carrier protein			
BLi01819	RpsP ribosomal protein S16 (BS17)			
BLi01824	RpIS ribosomal protein L19			
BLi01870	RpsB ribosomal protein S2			
BLi01871	Tsf elongation factor Ts			
BLi01893	RpsO ribosomal protein S15 (BS18)			
BLi02026	conserved hypothetical			
BLi02063	putative L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)			
BLi02064	close homolog to AraA: L-arabinose isomerase			

BLi02066	putative acetoin reductase (EC 1.1.1.5)
BLi02134	YjlD unknown; similar to NADH dehydrogenase
BLi02230	YozQ unknown
BLi02245	YocJ unknown; similar to acyl-carrier protein phosphodiesterase
BLi02332	CspD: cold-shock protein
BLi02352	hypothetical
BLi02414	Hbs: non-specific DNA-binding protein HBsu
BLi02637	TasA translocation-dependent antimicrobial spore component
BLi02667	RpmGA possible ribosomal protein L33
BLi02679	SodA superoxide dismutase
BLi02732	RpsU ribosomal protein S21
BLi02740	GrpE: heat-shock protein (activation of DnaK)
BLi02747	RpsT ribosomal protein S20 (BS20)
BLi02791	Yrkl unknown; similar to unknown proteins
BLi02921	RpmA ribosomal protein L27 (BL24)
BLi02923	RpIU ribosomal protein L21 (BL20)
BLi02953	Tig trigger factor (prolyl isomerase)
BLi02979	hypothetical
BLi02998	TrxA thioredoxin
BLi03033	RpIT ribosomal protein L20
BLi03034	Rpml ribosomal protein L35
BLi03099	close homolog to SspB: small acid-soluble spore protein (major beta-type SASP)
BLi03111	RpsD ribosomal protein S4 (BS4)
BLi03262	hypothetical
BLi03319	Yugl unknown; similar to polyribonucleotide nucleotidyltransferase
BLi03358	ComX: competence pheromone precursor
BLi03379	conserved hypothetical
BLi03455	YusA unknown; similar to unknown proteins
BLi03613	hypothetical
BLi03661	Eno: enolase
BLi03665	GapA: glyceraldehyde-3-phosphate dehydrogenase
BLi03780	Hag: flagellin protein
BLi03836	YwtC unknown
BLi03931	AtpE: ATP synthase (subunit c)
BLi03955	RpmE ribosomal protein L31
BLi03960	FbaA: fructose-1,6-bisphosphate aldolase
BLi04225	hypothetical
BLi04291	AhpC: alkyl hydroperoxide reductase (small subunit)
BLi04361	RpsR ribosomal protein S18
BLi04363	RpsF ribosomal protein S6 (BS9)
BLi04379	RpmH ribosomal protein L34
BLi04379	RpmH ribosomal protein L34

## A.5. Fremdgene (Aliengenes) in B. licheniformis

-					
ORF-Nr.	Annotation				
BLi00234	ThrZ threonyl-tRNA synthetase (minor)				
BLi00235	hypothetical				
BLi00236	hypothetical				
BLi00237	conserved hypothetical				
BLi00238	YrkA unknown; similar to hemolysin-like				
BLi00239	YrkA unknown; similar to hemolysin-like				
BLi00247	hypothetical				
BLi00248	YbfF unknown				
BLi00249	hypothetical				
BLi00253	Ybfl unknown; similar to transcriptional regulator (AraC/XylS family) fragment2				
BLi00255	hypothetical				
BLi00256	conserved hypothetical				
BLi00574	conserved hypothetical				
BLi00576	close homolog to Des: membrane phospholipid desaturase				
BLi00577	CspC: cold-shock protein				
BLi00579	putative dehydrogenase				
BLi00580	YyaS unknown; similar to unknown proteins				
BLi00581	YybA unknown; similar to transcriptional regulator (MarR family)				
BLi00582	PaiA transcriptional repressor of sporulation, septation and degradative enzyme genes				
BLi00583	PaiB transcriptional repressor of sporulation and degradative enzyme genes				
BLi00584	putative arginase protein (EC 3.5.3.1)				
BLi00585	putative ABC transporter (ATP-binding protein)				
BLi00586	YdeO unknown; similar to unknown proteins from B. subtilis				
BLi00588	putative two-component sensor histidine kinase				
BLi00590	putative multidrug transporter				
BLi00709	putative transposase				
BLi00718	hypothetical				
BLi00719	hypothetical				
BLi00720	hypothetical				
BLi00721	conserved hypothetical				
BLi00722	putative phosphotriesterase protein				
BLi00723	putative phosphotriesterase protein				
BLi00724	putative phosphopentomutase				
BLi00725	conserved hypothetical				
BLi00726	NagB: N-acetylglucosamine-6-phosphate isomerase				
BLi00734	hypothetical				
BLi00742	YefA unknown; similar to RNA methyltransferase				
BLi00743	putative type I restriction-modification system specificity subunit				
BLi00744	putative type I restriction-modification system methylation subunit				
BLi00745	putative type I restriction enzyme specificity protein (EC 3.1.21.3)				
BLi00746	putative restriction enzyme type I helicase subunit				

BLi00747	putative 5-methylcytosine-specific restriction enzyme A (EC 3.1.21)
BLi00748	close homolog to BgIP: PTS enzyme IIBCA component
BLi00749	close homolog to BglH: beta-glucosidase
BLi00750	putative transcription antiterminator
BLi00754	Yeel unknown; similar to unknown proteins
BLi01428	putative integrase-phage associated
BLi01429	YqaB unknown; similar to phage-related protein
BLi01432	hypothetical
BLi01433	YqaE unknown; similar to transcriptional regulator (phage-related) (Xre family)
BLi01436	putative phage protein
BLi01437	putative phage protein
BLi01442	YqaL unknown
BLi01443	YqaM unknown; similar to phage-related protein
BLi01445	hypothetical
BLi01446	hypothetical
BLi01448	hypothetical
BLi01449	hypothetical
BLi01450	hypothetical
BLi01451	conserved hypothetical
BLi01452	YqaN unknown
BLi01453	putative single-stranded DNA-binding protein
BLi01454	hypothetical
BLi01456	hypothetical
BLi01458	hypothetical
BLi01460	YqaQ unknown; similar to unknown proteins from B. subtilis
BLi01461	hypothetical
BLi01462	hypothetical
BLi01471	putative phage protein
BLi01472	hypothetical
BLi01475	putative phage protein
BLi01486	putative glycosyl hydrolase/lysozyme
BLi01488	hypothetical
BLi01489	YdfS unknown; similar to unknown proteins
BLi01491	YkzB unknown
BLi01531	putative integrase/recombinase
BLi01532	hypothetical
BLi01533	hypothetical
BLi01537	hypothetical
BLi01538	LigB: DNA ligase (ATP-dependent)
BLi01539	hypothetical
BLi01540	hypothetical
BLi01543	hypothetical
BLi01544	hypothetical
BLi01546	hypothetical
BLi01547	putative terminase-small subunit

BLi01548	putative terminase
BLi01549	putative phage head portal protein
BLi01550	putative phage pi3 protein 23, ATP dependent Clp protease
BLi01552	conserved hypothetical
BLi01553	conserved hypothetical
BLi01554	putative phage protein
BLi01555	hypothetical
BLi01557	putative phage protein
BLi01559	YqbO unknown; similar to phage-related protein
BLi01560	hypothetical
BLi01561	putative phage protein
BLi01562	putative phage protein
BLi01563	PBSX prophage ORF xkdV
BLi01566	putative phosphodiesterase
BLi01567	BhlA: holin-like protein
BLi01568	XIyB N-acetylmuramoyl-L-alanine amidase (PBSX prophage-mediated lysis)
BLi01569	putative holin
BLi01570	hypothetical
BLi01572	hypothetical
BLi01573	hypothetical
BLi01574	putative integrase-phage associated
BLi01575	conserved hypothetical
BLi01577	putative response regulator aspartate phosphatase H (EC 3.1)
BLi01578	putative response regulator aspartate phosphatase H (EC 3.1)
BLi01579	YoaT unknown; similar to unknown proteins
BLi01580	YozG unknown; similar to transcriptional regulator
BLi01581	YoaS unknown
BLi01584	hypothetical
BLi01585	conserved hypothetical
BLi01586	hypothetical
BLi01588	hypothetical
BLi01589	hypothetical
BLi01977	hypothetical
BLi01978	hypothetical
BLi01980	hypothetical
BLi01986	conserved hypothetical
BLi01988	hypothetical
BLi01994	hypothetical
BLi01995	putative phage protein
BLi01996	hypothetical
BLi01997	putative phage protein
BLi01998	hypothetical
BLi01999	hypothetical
BLi02000	hypothetical
BLi02002	hypothetical

BLi02005	hypothetical
BLi02006	hypothetical
BLi02007	hypothetical
BLi02022	putative transcriptional regulator MarR family
BLi02023	YdfN unknown; similar to NAD(P)H nitroreductase
BLi02024	YdfO unknown; similar to unknown proteins
BLi02025	YwrF unknown; similar to unknown proteins
BLi02115	putative glycerophosphodiester phosphodiesterase
BLi02116	putative D-mannonate dehydratase (EC 4.2.1.8)
BLi02118	putative C4-dicarboxylate transporter, large subunit
BLi02119	putative dicarboxylate transporter small subunit
BLi02120	DctB: possible C4-dicarboxylate binding protein
BLi02121	close homolog to CcpA: transcriptional regulator mediating carbon catabolite repression
BLi02698	putative endonuclease V (EC 3.1.25.1)
BLi02699	YwqL unknown; similar to putative endonuclease
BLi02700	hypothetical
BLi02701	hypothetical
BLi02702	hypothetical
BLi02703	hypothetical
BLi02704	conserved hypothetical
BLi02705	conserved hypothetical
BLi02706	conserved hypothetical
BLi02762	Sda developmental checkpoint coupling sporulation initiation to replication initiation
BLi02770	hypothetical
BLi02771	hypothetical
BLi02772	putative RNA polymerase sigma-K factor, fragment 1
BLi02773	RNA polymerase sigma factor precursor
BLi02774	putative alanyl-tRNA synthetase (EC 6.1.1.7)
BLi02775	putative amino acid aldolase or racemase
BLi02776	putative translation initiation inhibitor
BLi02777	YccC unknown; similar to asparaginase
BLi02779	putative threonine synthase (EC 4.2.3.1)
BLi02780	conserved hypothetical
BLi02781	DctR: two-component response regulator involved in C4-dicarboxylate transport
BLi02782	RNA polymerase sigma-K factor
BLi02783	conserved hypothetical
BLi02784	BmrR: transcriptional activator of the bmrUR operon
BLi02789	YdfQ unknown; similar to thioredoxin
BLi02790	YrkJ unknown; similar to unknown proteins
BLi02792	YrkH unknown; similar to unknown proteins
BLi02793	YrkF unknown; similar to unknown proteins from B. subtilis
BLi02794	YrkE unknown; similar to unknown proteins
BLi02795	YrkD unknown; similar to unknown proteins
BLi02796	putative S-adenosylmethionine-dependent methyltransferase
BLi02797	conserved hypothetical

BLi02800	putative ABC transporter permease protein
BLi02801	conserved hypothetical
BLi02804	HxlB: 6-phospho-3-hexuloisomerase
BLi02805	HxIA: 3-hexulose-6-phosphate synthase
BLi02806	HxIR: positive regulator of hxIAB expression
BLi02814	putative dipeptide transport ATP-binding protein
BLi02966	hypothetical
BLi02967	putative phage PBSX protein
BLi02969	putative phage protein
BLi02970	hypothetical
BLi02972	hypothetical
BLi02973	hypothetical
BLi02974	putative modification methylase alpha subunit (EC 2.1.1.73)
BLi02975	hypothetical
BLi02976	conserved hypothetical
BLi02977	putative ABC transporter (binding protein)
BLi02978	putative response regulator aspartate phosphatase
BLi02980	hypothetical
BLi02981	hypothetical
BLi03583	hypothetical
BLi03584	hypothetical
BLi03588	hypothetical
BLi03589	conserved hypothetical
BLi03590	putative transposase
BLi03591	putative transposase
BLi03593	conserved hypothetical
BLi03594	hypothetical
BLi03595	hypothetical
BLi03596	conserved hypothetical
BLi03597	hypothetical
BLi03601	YomQ unknown
BLi03602	putative phage PBSX protein
BLi03603	conserved hypothetical
BLi03604	putative phage autolysin (amidase) homolog
BLi03606	putative phage tail protein
BLi03607	hypothetical
BLi03608	hypothetical
BLi03609	hypothetical
BLi03610	hypothetical
BLi03611	hypothetical
BLi03612	hypothetical
BLi03614	putative phage protein
BLi03615	ClpP family serine protease, possible phage related
BLi03617	putative phage protein
BLi03618	putative terminase large subunit

DI 100040				
BLi03619	putative phage terminase protein, small subunit			
BLi03623	hypothetical			
BLi03624	putative phage protein			
BLi03628	hypothetical			
BLi03630	putative phage protein			
BLi03631	putative phage protein			
BLi03632	putative phage protein			
BLi03635	putative phage protein			
BLi03637	putative phage protein			
BLi03638	hypothetical			
BLi03639	YdcL unknown; similar to integrase			
BLi03658	hypothetical			
BLi03659	YoaZ unknown; similar to unknown proteins from B. subtilis, fragment 2			
BLi03660	YoaZ unknown; similar to unknown proteins from B. subtilis			
BLi04172	putative anaerobic ribonucleoside-triphosphate reductase activating protein (EC 1.97.1.4)			
BLi04173	hypothetical			
BLi04174	hypothetical			
BLi04175	hypothetical			
BLi04176	putative cell-division protein / general stress protein (class III heat-shock)			
BLi04296	hypothetical			
BLi04297	hypothetical			
BLi04299	YvfS unknown; similar to ABC transporter transmembrane subunit, fragment 2			
BLi04300	YvfS unknown; similar to ABC transporter transmembrane subunit			
BLi04305	hypothetical			
BLi04306	hypothetical			
BLi04307	putative ABC transporter ATP-binding protein			
BLi04308	hypothetical			
BLi04310	hypothetical			
BLi04311	conserved hypothetical			
BLi04312	hypothetical			
BLi04313	hypothetical			
BLi04314	putative metal-dependent hydrolase			
BLi04315	putative type I restriction enzyme (restriction chain)			
BLi04316	putative type I restriction enzyme hsdS subunit			
BLi04317	putative DNA-binding protein			
BLi04318	putative type I site-specific deoxyribonuclease			
BLi04319	hypothetical			
BLi04320	hypothetical			
BLi04322	putative site-specific recombinase			
BLi04323	SpoIVCA site-specific DNA recombinase required for creating the sigK gene (excision of the			
	skin element)			
BLi04325	conserved hypothetical			
BLi04328	hypothetical			
BLi04328	hypothetical			

## Danksagung

Mein besonderer Dank gilt Herrn Prof. Gottschalk, der das Genomprojekt *B. licheniformis* ins Leben gerufen hat, vor allem für die Überlassung dieses Themas, die Ermöglichung des selbständigen Arbeitens, sein stetiges Interesse und Engagement am Fortgang dieser Arbeit und seine vielfältig gewährte Unterstützung. Vielen Dank für die Gelegenheit, an vielen Tagungen und innovativen Gesprächen im Rahmen der Kooperation teilnehmen zu können!

Dr. Armin Ehrenreich möchte ich sehr herzlich für die Einführung in die Welt der *Transcriptomics*, seinen Ideenreichtum und seine Diskussionsbereitschaft, sowie für seine umfangreiche Betreuung und Unterstützung der durchgeführten Arbeiten danken.

An dieser Stelle möchte ich mich ebenfalls bei unseren Kooperationspartnern der Firma Henkel, Dr. Maurer, Dr. Feesche und Dr. Evers, sowie der Universität Greifswald, Prof. Hecker, Dr. Schweder und Dr. Voigt, für die Zusammenarbeit bedanken.

Weiterhin gilt mein Dank Prof. Ruth Schmitz-Streit für die Übernahme des Korreferates.

Vielen Dank Prof. Merkel für die Berechnung der Fremd- und hochexprimierten Gene mit dem Programm SIGI.

Allen ehemaligen und jetzigen Mitgliedern des Labors für Genomanalyse bin ich für eine großartige Zusammenarbeit und umfangreiche Unterstützung zu größtem Dank verpflichtet! Vielen Dank an Holger Brüggemann, Rosa Martìnez-Arias und ganz besonders an Anke Henne, die mir das "Gap 4" in die Genomsequenzierung geöffnet haben. Ein riesiges Dankeschön geht natürlich an die Sequenzier-Feen Silke Denker, Silke Steckel, Iwona Decker, Mechthild Bömecke und Monika Grzywa für ihr unermüdliches Gelestückeln. Ohne Euch würde es wohl die ein oder andere der 47.000 Sequenzen nicht geben! Bei Heiko Liesegang bedanke ich mich ganz besonders für das bioinformatische Jonglieren der *B. licheniformis* Datenbank und hitzige Diskussionen, in denen er versucht hat, mir die Bioinformatik wenigstens in Bruchteilen näher zu bringen. Kein leichtes Unterfangen. In diesem Sinne geht ebenfalls ein großes Dankeschön an Jarek Sobkowiak und Arnim Wiezer. Ohne Euch hätten mein Rechner und ich keine so intensive Haß-Liebe aufbauen können...

Weiterhin bedanke ich mich bei Axel Strittmatter für eine produktive Zusammenarbeit.

Das Labor 255: Ein ganz herzliches Dankeschön gilt hier Christina Herzberg für ihren unschlagbaren Enthusiasmus und ihre Fähigkeit einfache Lösungen zu sehen, wenn man selbst um tausend Ecken denkt. Der gute Geist des Labors, ich habe viel von Dir gelernt. Herzlich bedanken möchte ich mich auch bei Silke Steckel für ihre Bereitschaft und Begeisterung sich in kurzer Zeit in völlig neue Methoden einzuarbeiten. Ohne Euch zwei wäre die Hälfte der

Arbeit jetzt noch nicht geschafft. Vielen Dank auch an die Arbeitsgruppe von Prof. Stülke, durch die viele neue Eindrücke und frischer Wind auf die zweite Etage und ins Labor 255 gelangt sind.

Ein ganz großes Dankeschön gilt Armin Ehrenreich, Marc Hoffmeister, Heiko Liesegang, Arnim Wiezer und natürlich Prof. Gottschalk für das Korrekturlesen dieser Arbeit.

Meinen Mitstreitern, Kollegen, Freunden Christel Schmeißer, Florian Fricke, Marc Hoffmeister, Olivia Gohl, Carsten Raasch und Christian Hildmann danke ich sehr für das unentbehrliche sowohl fachliche, als auch private Mentaltraining!

Ein großer Dank geht auch an das Werkstatteam Gerd Birke, Olaf Waase und Patrick Regin für ihre Hilfsbereitschaft bei den unterschiedlichsten Problematiken.

Natascha, Katha, Alexandra, Maike und meiner Lieblings-WG Patricia, Ina und Anna danke ich für ihr Verständnis, ihre Unterstützung und vor allem für ihre Freundschaft. Schön, dass es Euch gibt! An dieser Stelle möchte ich mich auch bei Frau Böddeker bedanken.

Meiner Familie danke ich von Herzen für alles; für ihre Unterstützung, ihre Liebe und ihr Vertrauen.

Lebenslauf	
23.05.1977	geboren in Jena deutsche Staatsbürgerschaft
1983 - 1991	Polytechnische Oberschule (POS) Theodor-Neubauer in Jena
1991 -1995	Carl-Zeiss-Spezialschule in Jena (heute naturwissenschaftlich/mathematischer Spezialschulteil des Carl-Zeiss-Gymnasiums in Jena)
1995	Erlangen der Allgemeinen Hochschulreife
WS1995/96	Immatrikulation an der Georg-August-Universität Göttingen im Studiengang Biologie (Diplom)
20.10.1997	Diplomvorprüfung in den Fächern Mikrobiologie, Zoologie, Anorganische Chemie und Physikalische Chemie
05.05.2000	Diplomprüfung in den Fächern Mikrobiologie, Immunologie und Physikalische Chemie
Juni 2000 - Mai 2001	Anfertigung der experimentellen Diplomarbeit unter Anleitung von Prof. Dr. W. Liebl am Institut für Mikrobiologie und Genetik der Georg-August-Universität Göttingen mit dem Titel: "Die thermostabilen α-Glucosidasen AglB aus <i>Thermotoga maritima</i> MSB8 und MalA aus <i>Thermotoga neapolitana</i> NS-E: Expression, Reinigung und biochemische Charakterisierung der rekombinanten Enzyme"
Juli 2001 - Januar 2002	Beschäftigung als wissenschaftliche Hilfskraft im Labor für Genomanalyse am Institut für Mikrobiologie und Genetik der Georg-August-Universität Göttingen
Februar 2002	Beginn der experimentellen Arbeit zur vorliegenden Dissertation