

**Entschlüsselung der Genome von
Ralstonia eutropha H16 und
Methanosphaera stadtmanae
und vergleichende Untersuchungen
zu Anpassungen der Genomorganisation**

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

vorgelegt von
Wolfgang Florian Fricke
aus Lemgo

Göttingen 2005

D 7

Referent: Prof. Dr. G. Gottschalk

Korreferent: Prof. Dr. B. Bowien

Tag der mündlichen Prüfung: 30.06.2005

Inhaltsverzeichnis

1 Einleitung.....	1
2 Material und Methoden.....	5
2.1 Organismen und Plasmide	5
2.2 Nährmedien und Puffer.....	6
2.2.1 Medien für <i>Escherichia coli</i>	6
2.2.2 Medienzusätze	6
2.2.3 Stamm- und Pufferlösungen	7
2.3 Zellanzucht und Stammhaltung	7
2.3.1 Zellanzucht.....	7
2.3.1.1 Anzucht von <i>Ralstonia eutropha</i> H16	7
2.3.1.2 Anzucht von <i>Methanospaera stadtmanae</i>	8
2.3.1.3 Anzucht von <i>Escherichia coli</i>	8
2.3.2 Stammhaltung	9
2.4 Standardtechniken für die Arbeit mit Nukleinsäuren	9
2.4.1 Geräte und Lösungen.....	9
2.4.2 Dialyse von Nukleinsäuren.....	9
2.4.3 DNA-Konzentrationsbestimmung und Reinheitskontrolle.....	10
2.4.4 Agarose-Gelelektrophorese	10
2.4.5 Saccharose-Dichtegradientenzentrifugation	11
2.4.6 Isolierung chromosomaler DNA.....	12
2.4.7 Isolierung von Plasmid DNA.....	13
2.4.8 Aufreinigung von DNA-Fragmenten.....	13
2.4.9 Isolation von DNA-Fragmenten aus Agarosegelen.....	14
2.4.10 Scheren von DNA.....	14
2.5 Modifikation von Nukleinsäuren	15
2.5.1 Restriktion.....	15
2.5.2 Dephosphorylierung.....	15
2.5.3 Herstellung glatter DNA-Enden (<i>blunt ends</i>)	16
2.5.4 Herstellung von Insert-DNA mit A-Überhang	16
2.5.5 Ligation.....	17

2.6 Transformation und Selektion	17
2.6.1 TOPO TA-Klonierung	17
2.6.1.1 Ligation mit dem TOPO - Vektor	18
2.6.1.2 Transformation von <i>E. coli</i> TOP10 - Zellen	18
2.6.2 Blau-Weiß <i>screening</i> (X-Gal-Test)	18
2.7 Amplifikationen	19
2.7.1 Polymerase-Kettenreaktion (PCR)	19
2.7.2 GenomiPhi-Amplifikation	22
2.8 Genbanken	23
2.8.1 Plasmidbanken	23
2.9 Sequenzierung	24
2.9.1 Sequenzierung mit dem ABI PRISM [®] 377	24
2.9.2 Sequenzierung mit MegaBACE 1000 bzw. MegaBACE 4000	26
2.10 Prozessierung und Assemblierung der DNA-Sequenzen	28
2.11 Lückenschluß	28
2.11.1 Primerwalking	29
2.11.2 Lückenschluß mit PCR	29
2.11.2.1 Ableitung über Cosmidbrücken	29
2.11.2.2 Ableitung über konservierte Genabfolgen (<i>peptide links</i>)	30
2.11.3 Multiplex-PCR	30
2.12 BLAST	31
2.12.1 Bidirektionaler BLAST	32
2.13 Vorhersage und Annotation kodierender Sequenzen	33
2.13.1 Vorhersage kodierender offener Leserahmen (CDSs)	33
2.13.2 Vorhersage von Transfer-RNAs (tRNAs)	34
2.13.3 Annotation in ERGO	34
2.13.4 Kontrolle der CDS-Längen	35
2.14 Kumulativer <i>GC-Skew</i>	35
2.15 Untersuchungen zur <i>Codon Usage</i> (SIGI)	36
3 Ergebnisse	37
3.I Das <i>Ralstonia eutropha</i> H16-Genomprojekt	37
3.I.1 Sequenzierung	37
3.I.1.1 Strategie	37
3.I.1.2 Rohsequenzierungsphase	37

3.I.1.3 Editierung des Rohdatensatzes	39
3.I.1.3.1 Editierung der Consensus-Sequenz	39
3.I.1.3.2 Primerwalking über Sequenzlücken	41
3.I.1.3.3 Verknüpfung von Contig-Enden im Gesamtdatensatz	42
3.I.1.3.4 Auflösung von Fehlassemblierungen	42
3.I.1.4 Strukturierung des Gesamtdatensatzes	44
3.I.1.4.1 Herstellung einer selektiven Genbank für Chromosom 2	45
3.I.1.4.2 Ordnung der Contigs relativ zum Genom	46
3.I.1.4.2.1 Bildung von Supercontigs über Cosmid-Brücken	46
3.I.1.4.2.2 Vergleich konservierter Genabfolgen	46
3.I.1.5 Lückenschluß	47
3.I.1.6 Sekundärstrukturen	47
3.I.1.7 Feinkorrektur des Gesamtdatensatzes	48
3.I.2 Sequenzanalyse	49
3.I.2.1 G+C-Gehalt	52
3.I.2.2 Replikationsursprünge	52
3.I.2.3 RNA-Gene	55
3.I.2.4 Vorhersage kodierender Sequenzen	56
3.I.2.5 Annotation	57
3.I.3 Mosaikstruktur des <i>R. eutropha</i> -Genoms	58
3.I.3.1 Vergleich von <i>R. eutropha</i> und <i>R. solanacearum</i>	58
3.I.3.2 Vorhersage genomischer Inseln	61
3.I.4 Charakterisierung von Chromosom 2	63
3.I.4.1 Fakultativ chemolitho- und organoautotrophes Wachstum	63
3.I.4.2 Erweiterung des Substratspektrums	65
3.I.4.2.1 Alternative Kohlenstoffquellen	65
3.I.4.2.1.1 Abbau von Kohlenhydraten	65
3.I.4.2.1.2 Abbau von aus Zuckern abgeleiteten Säuren	70
3.I.4.2.1.3 Abbau aromatischer Kohlenwasserstoffe	71
3.I.4.2.1.4 Acetoin-Abbau	77
3.I.4.2.2 Erschließung alternativer Stickstoff-Quellen	78
3.I.4.2.2.1 Cyanat-Hydratase	78
3.I.4.2.2.2 Formamidase	78
3.I.4.2.2.3 Ethanolamin-Ammonium-Lyase	79

3.I.4.2.3 Alternative Phosphorquellen	80
3.I.4.3 Synthese von Pyrroloquinolinquinon (PQQ).....	80
3.I.4.4 Kdp-ATPase	81
3.I.4.5 Begeißelung	82
3.I.4.6 Synthese von Sekundärmetaboliten.....	84
3.I.4.6.1 Nicht-ribosomale Peptid-Synthetasen (NRPSs).....	84
3.I.4.6.2 Potentielle Toxine.....	86
3.I.4.6.2.1 RTX-Toxine	86
3.I.4.6.2.2 Tc-Toxine	88
3.II Das <i>Methanosphaera stadtmanae</i> -Genomprojekt.....	90
3.II.1 Sequenzierung und Annotation	90
3.II.1.1 Strategieanpassung	90
3.II.1.2 Rohsequenzierung.....	91
3.II.1.3 Editierung des Rohdatensatzes.....	92
3.II.1.4 Lückenschluß	93
3.II.1.5 Identifikation von zwei Varianten eines ribosomalen Operons	93
3.II.1.6 Vorhersage und Annotation kodierender Sequenzen.....	94
3.II.2 Sequenzanalyse	95
3.II.2.1 Allgemeine Eigenschaften	95
3.II.2.2 Replikationsursprung	98
3.II.2.3 Potentiell hochexprimierte Gene.....	99
3.II.3 Methanogenese in <i>Msp. stadtmanae</i>	101
3.II.4 CO ₂ -Reduktionsweg.....	106
3.II.5 Genomvergleiche	110
3.II.5.1 Nicht identifizierte Proteine	110
3.II.5.2 <i>Msp. stadtmanae</i> -spezifische Gene.....	113
3.II.5.2.1 Enzyme der Zellwandsynthese.....	114
3.II.5.2.2 Überlange offene Leserahmen	115
3.II.5.2.2.1 Zusammensetzung der überlangen ORFs und Proteine	117
3.II.5.2.2.2 Aufbau der überlangen Proteine.....	118
3.II.5.2.3 Restriktions-Modifikations-System	119
4 Diskussion.....	120
4.I Das <i>Ralstonia eutropha</i> H16-Genomprojekt.....	120
4.I.1 Das <i>R. eutropha</i> H16-Genom	120

4.I.2 Charakterisierung von Chromosom 2 aus <i>R. eutropha</i> H16.....	121
4.I.3 Genomorganisation in <i>R. eutropha</i> H16.....	123
4.I.4 Replikonvergleiche innerhalb der <i>Burkholderiales</i>	123
4.I.5 Genomorganisation innerhalb der <i>Burkholderiaceae</i>	126
4.I.6 Verwandtschaftsverhältnisse mutmaßlicher Megaplasmide	127
4.I.7 Gemeinsame Abstammung verschiedener Megaplasmide.....	129
4.I.8 Hypothetische Entwicklung der <i>Burkholderiales</i> -Genome.....	130
4.I.9 Bemerkungen zur taxonom. Einteilung der <i>Burkholderiaceae</i>	132
4.I.10 Ausblick.....	133
4.II Das <i>Methanosphaera stadtmanae</i> -Genomprojekt.....	134
4.II.1 Das <i>Methanosphaera stadtmanae</i> -Genom.....	134
4.II.2 Der methanogene Stoffwechsel.....	134
4.II.3 Kommensalismus	136
4.II.3.1 Zellwandsynthese.....	137
4.II.3.2 Überlange ORFs.....	139
4.II.4 Spezifische Anpassungen des <i>Msp. stadtmanae</i> -Genoms	141
4.II.5 Ausblick	143
5 Zusammenfassung	144
5.I Das <i>Ralstonia eutropha</i> -Genomprojekt.....	144
5.II Das <i>Methanosphaera stadtmanae</i> -Genomprojekt.....	145
6 Literaturverzeichnis	147

Abkürzungsverzeichnis

A	Adenin
Abb.	Abbildung
ABC	<i>ATP-binding cassette</i>
Ala	Alanin
Amp	Ampicillin
APS	Ammoniumpersulfat
Arg	Arginin
Asn	Asparagin
Asp	Aspartat
ATP	Adenosin-5'-triphosphat
AA	<i>amino acid</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
bidest	bidestilliert
<i>Bo.</i>	<i>Bordetella</i>
bp	base pair(s)
<i>Bu.</i>	<i>Burkholderia</i>
bzw.	beziehungsweise
°C	Grad Celsius
C	Cytosin oder Kohlenstoff
ca.	circa
CDS	<i>coding sequence</i>
Cl	Chlor
cm	centimeter
CoA	Coenzym A
CoB	Coenzyme B
CoM	Coenzym M
C-terminal	carboxyterminal
Cys	Cystein
d.h.	das heißt
DNA	<i>deoxyribonucleic acid</i>
dNTP	Desoxyribonukleotid-5'-Triphosphat

DSMZ	Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig
<i>E.</i>	<i>Escherichia</i>
EDTA	<i>ethylene diamine tetraacetic acid</i>
et al.	et alteri
Fa.	Firma
FAD	Flavinadenindinukleotid
g	Gramm
G	Guanin oder Glycin
Gap	<i>genome assembly program</i>
Glimmer	<i>gene locator and interpolated markov modeler</i>
Gln	Glutamin
Glu	Glutamat
Gly	Glycin
GmbH	Gesellschaft mit beschränkter Haftung
h	Stunde
H	Wasserstoff
Hdr	Heterodisulfid-Reduktase
His	Histidin
HPLC	<i>high performance liquid chromatography</i>
H4MPT	Tetrahydromethanopterin
Hrsg.	Herausgeber
i.d.R.	in der Regel
Ile	Isoleucin
inc.	<i>increase</i>
IPTG	Isopropyl- β -Thiogalactopyranosid
K	Kalium oder Lysin
k.A.	keine Angabe
kbp	Kilobasenpaare
kg	Kilogramm
l	Liter
LB	Luria-Bertani
Leu	Leucin
LPS	Lipopolysaccharide

ABKÜRZUNGSVERZEICHNIS

Lys	Lysin
μ	Mikro
M	Molar
Mbp	Megabasenpaare
<i>Mbr.</i>	<i>Methanobrevibacter</i>
<i>Mca.</i>	<i>Methanococcus</i>
MCM	<i>minichromosome maintenance complex</i>
MCR	Methyl-CoM-Reduktase
MCS	<i>multiple cloning site</i>
Met	Methionin
MFR	Methanofuran
min	Minute
<i>Mpy.</i>	<i>Methanopyrus</i>
<i>Msa.</i>	<i>Methanosarcina</i>
<i>Msp.</i>	<i>Methanosphaera</i>
<i>Mth.</i>	<i>Methanothermobacter</i>
N	Asparagin
Na	Natrium
NAD	Nicotinamid-Adenin-Dinukleotid
NADP	Nicotinamid-Adenin-Dinukleotid-Phosphat
NCBI	<i>National Center for Biotechnology Information</i>
Nr.	Nummer
nt.	Nukleotid(e)
N-terminal	aminoterminal
O	Sauerstoff
OD	optische Dichte
ORF	<i>open reading frame</i>
PCR	<i>polymerase chain reaction</i>
Pfam	<i>protein families</i>
pH	negativer dekadischer Logarithmus der Protonenkonzentration
Phe	Phenylalanin
Pro	Prolin
<i>R.</i>	<i>Ralstonia</i>
R	Arginin

R-M-System	Restriktions-Modifikations-System
RNA	<i>ribonucleic acid</i>
rRNA	ribosomale RNA
RT	Raumtemperatur
s	Sekunde
S	Schwefel
s.o.	siehe oben
SDS	<i>sodium dodecyl sulfate</i>
Sec	<i>secretion</i>
Ser	Serin
<i>s-layer</i>	<i>surface layer</i>
sp.	<i>species</i>
T	Thymin
TEMED	N, N, N', N'-Tetramethylethylendiamid
Threonin	Thr
TISS	Typ I Sekretionsystem
Tm	<i>melting temperature</i>
TM	transmembran
Tris	Tris-(hydroxymethyl-)aminomethan
tRNA	Transfer-RNA
TISS	Typ I Sekretionsystem
Trp	Tryptophan
Tyr	Tyrosin
U	<i>unit</i>
u.a.	unter anderem
v/v	<i>volume per volume</i>
V	Volt
V-Typ	vakuolärer Typ
Val	Valin
w/v	<i>weight per volume</i>
X-Gal	5-Brom-4-Chlor-3-Indolyl-D-Galactosid
z.B.	zum Beispiel

1 Einleitung

Die Entschlüsselung eines mikrobiellen Genoms unterteilt sich im wesentlichen in zwei Schritte: Sequenzierung und Sequenzanalyse. Mit zunehmender Größe der Genomprojekte werden beide Schritte soweit wie möglich standardisiert und automatisiert. Bestimmte Aufgaben – wie das Aufreinigen von DNA-Proben – werden von Robotern übernommen; die Sequenzierung selbst findet im Hochdurchsatz an Kapillarsequenzierern statt, die deutlich mehr als 1000 Sequenzreaktionen pro Tag durchführen und dabei Sequenzläufe einer Länge von bis zu 1200 Basen produzieren. Unabhängig von den speziellen physiologischen Eigenschaften eines Organismus verlangt die Technik der Sequenzierung nur geringe spezifische Anpassungen, die sich nach der Größe des Genoms, der Anzahl der Replikons, auf die sich das Genom verteilt, dem G+C-Gehalt und der Anzahl und Länge etwaiger repetitiver Bereiche innerhalb der Genomsequenz richten. Bei der Sequenzanalyse wird versucht, soweit wie möglich auf bioinformatische Werkzeuge zurückzugreifen, um die Bearbeitung der enormen Datenmengen, die bei einem Genomprojekt anfallen, überhaupt gewährleisten zu können. Trotz weitreichender Entwicklungen innerhalb der Bioinformatik besteht in der Genomforschung allerdings ein Ungleichgewicht zwischen der Menge an verfügbaren Sequenzdaten und der angemessenen Auswertung dieser Information. Zu den wichtigsten Schritten der Sequenzanalyse, die mit Hilfe bioinformatischer Algorithmen durchgeführt werden, gehören die Vorhersage kodierender offener Leserahmen und die funktionelle Annotation dieser potentiellen Gene anhand von Sequenzvergleichen mit bekannten Proteinen. Die dynamische Entwicklung der Sequenziertechnik mit dem Ziel einer weiteren Effizienzsteigerung hat bis heute dazu geführt, daß 229 mikrobielle Genome – darunter 208 bakterielle und 21 archaeelle – vollständig entschlüsselt und am National Center for Biotechnology Information veröffentlicht worden sind.

Die weiterführende Analyse vollständig sequenzierter und annotierter Genome kann von verschiedenen Ansprüchen geleitet werden. Zunächst können anhand der Sequenz Erkenntnisse über den Organismus, die auf experimentellem Wege gesammelt worden sind, bestätigt, ergänzt und neu interpretiert werden. Je nach Umfang der zur Verfügung stehenden experimentellen Daten kann dieser Ansatz bei der Analyse des Genoms im Vordergrund stehen. Gleichzeitig liegt es aber im Interesse des Projektes,

Eigenschaften des Genoms zu charakterisieren, die sich erst aus der Untersuchung der Sequenz ergeben und die Hinweise auf unbekannte spezifische Qualitäten des Organismus geben können. Dazu bietet sich die genomweite Analyse mit Hilfe standardisierter bioinformatischer Werkzeuge an. Beispielsweise können aus dem Vergleich der *Codon Usage* potentieller Gene genomische Inseln identifiziert werden (Merkel, 2004), die auf lateralen Gentransfer zurückgeführt und als Grundlage neu erworbener genetischer Fähigkeiten angesehen werden (Doolittle, 1999). Eine weitere und sehr vielversprechende Form der Sequenzanalyse ergibt sich schließlich aus der zunehmenden Verfügbarkeit vollständig entschlüsselter Genome aus phylogenetisch verwandten Bereichen oder von Organismen mit ähnlichen physiologischen Eigenschaften. Durch Bildung von Vergleichsgruppen ähnlicher, bezüglich besonderer Merkmale jedoch unterschiedlicher Organismen kann gezielt nach den genetischen Grundlagen dieser speziellen Eigenschaften gesucht werden. Beispielsweise bieten sich, um genetische Grundlagen dieser Lebensstile zu identifizieren, Vergleiche zwischen auto- und heterotrophen, pathogenen und kommensalen Organismen und zwischen Hochleistungstämmen und Wildtypen an.

In der vorliegenden Arbeit wurden mit Hilfe standardisierter Methoden zwei völlig unterschiedliche mikrobielle Organismen entschlüsselt. Die im Laufe des ersten Teils – der Sequenzierung und Analyse des β -Proteobakteriums *Ralstonia eutropha* H16 – gewonnenen Erfahrungen konnten dabei gewinnbringend für die Entschlüsselung des zweiten Organismus – des methanogenen Archaeons *Methanosphaera stadtmanae* – eingesetzt werden. Über jeden der drei geschilderten Ansätze für die Analyse von Genomsequenzen konnten dabei Ergebnisse erzielt werden, die für beide Organismen im Hinblick auf unterschiedliche Anpassungsstrategien der Genomorganisation interpretiert werden können.

1. Das Gram-negative β -Proteobakterium *Ralstonia eutropha* H16 gehört zur Familie der *Burkholderiaceae*, die sich im wesentlichen aus den Gattungen *Burkholderia* und *Ralstonia* zusammensetzt. Spezies dieser Familie haben sich an äußerst unterschiedliche ökologische Nischen angepasst und damit ein beträchtliches biotechnologisches Interesse geweckt (Coenye & Vandamme, 2003). Zu den Anwendungen, die man sich aus der Untersuchung verschiedener *Burkholderiaceae*-Stämme abzuleiten erhofft, gehören der Abbau von Xenobiotika, die Toleranz gegenüber Schwermetallen, die Wachstumsförderung von Nutzpflanzen und die Bekämpfung pflanzlicher, tierischer und menschlicher

Pathogene. Die industrielle Nutzung vieler Spezies wird allerdings durch ihre schädliche Wirkung auf den Menschen erschwert. Bis heute wurden aus der Familie der *Burkholderiaceae* drei Genome vollständig entschlüsselt: *R. solanacearum* GMI1000 (Salanoubat *et al.*, 2002), *Burkholderia mallei* ATCC 23344 (Nierman *et al.*, 2004) und *Bu. pseudomallei* K96243 (Holden *et al.*, 2004). Eine Reihe weiterer Genome, zu denen aus der Gattung *Ralstonia* die Stämme *R. eutropha* JMP134 und *R. metallidurans* CH34 gehören, befindet sich in einem fortgeschrittenen Stadium der Sequenzierung am DOE Joint Genome Institute in Kalifornien, USA.

Ralstonia eutropha H16, bei dem bisher keine in irgendeiner Form pathogene Wirkung nachgewiesen werden konnte, zeichnet sich durch ein weites Substratspektrum und eine außergewöhnliche metabolische Vielseitigkeit aus, die diesen Stamm zu einem gut untersuchten Modellorganismus gemacht haben. Zu den charakteristischen Eigenschaften von *R. eutropha* H16 gehören das fakultativ chemolithoautotrophe Wachstum, das auf der Oxidation von molekularem Wasserstoff und der Fixierung von Kohlendioxid über den reduktiven Pentose-Phosphat-Cyklus nach Calvin-Benson-Bassham basiert (Hogrefe *et al.*, 1984; Kusian & Bowien, 1997; Bowien & Kusian, 2002) und das fakultativ anaerobe Wachstum durch Denitrifikation (Pfitzner & Schlegel, 1973). Polyhydroxyalkanoate, die von *R. eutropha* H16 als Speicherstoffe synthetisiert werden, finden industriellen Einsatz als Kunststoffersatz (Steinbüchel & Schlegel, 1991; Gorenflo *et al.*, 2001). Aufgrund dieser vielfältigen Eigenschaften wird *R. eutropha* H16 ein beträchtliches Potential für die Konstruktion maßgeschneiderter Produktionsstämme durch gezielte genetische Eingriffe (*metabolic engineering*) zugeschrieben, insbesondere im Zusammenhang mit einer auf Wasserstoff basierenden Technologie.

Diese Arbeit beschäftigt sich mit der Entschlüsselung der Genomsequenz von *R. eutropha* H16, um auf Basis der genetischen Grundlagen ein umfassendes Verständnis für die komplexen Hintergründe dieser Eigenschaften zu gewinnen. Das Genom von *R. eutropha* H16 verteilt sich auf drei Replikons (Schwartz & Friedrich, 2001), von denen das kleinste – das Megaplasmid pHG1 – bereits vollständig sequenziert und annotiert wurde und für Schlüsselfunktionen des fakultativ lithoautotrophen und fakultativ anaeroben Stoffwechsels kodiert (Schwartz *et al.*, 2003). Ziel dieser Arbeit war es, die besondere Funktion von Chr.2 für die Charakterisierung von *R. eutropha* H16 herauszustellen und aus vergleichenden Untersuchungen mit anderen *Burkholderiaceae*-Genomen Hinweise auf allgemeine Prinzipien der Genomorganisation innerhalb dieser Familie abzuleiten.

2. *Methanosphaera stadtmanae* gehört zu den methanogenen Organismen, die ausschließlich in der Domäne der Archaeen gefunden werden und sich auf den letzten Schritt des anaeroben Kohlenstoff-Cyclus spezialisiert haben: die Reduktion von CO₂, von methylierten C₁-Verbindungen oder von Acetat zu Methan (Deppenmeier, 2002b). Im globalen Kohlenstoffhaushalt spielen sie eine wichtige Rolle beim Abbau organischer Verbindungen in anoxischen Umgebungen wie Sedimenten von Flüssen, Seen und Sümpfen (Conrad, 1996). Dem durch die Methanogenese freigesetzten Methan kommt als starkem Treibhausgas eine besondere ökologische Bedeutung zu (Reay, 2003). Gleichzeitig dient Methan als Rohstoffquelle für die Energieerzeugung und lagert in Form von Methanhydraten biologischen Ursprungs wahrscheinlich in beträchtlichen Mengen in Sedimentschichten am Boden der Ozeane (Marchesi *et al.*, 2001; Wood *et al.*, 2002).

Ein weiteres anoxisches Habitat, in dem sich *Msp. stadtmanae* auf eine spezielle Form der Methanogenese – die Reduktion von Methanol mit H₂, spezialisiert hat – stellt der menschliche Enddarm dar (Miller & Wolin, 1983; Miller & Wolin, 1985). Der überwiegende Anteil der endogenen Methanproduktion geht allerdings auf das häufigste methanogene Archaeon im menschlichen Darm, *Methanobrevibacter smithii*, zurück (Miller & Wolin, 1982). Ein Teil der im Darm produzierten Gase wird absorbiert und über den Atem abgegeben, in dem sich bei ca. 30-50 % der westlichen Bevölkerung signifikante Mengen Methan nachweisen lassen (Levitt & Ingelfinger, 1968; Pitt *et al.*, 1980). Unter den methanproduzierenden Menschen liegt der Anteil der Frauen (63 %) höher als der der Männer (37 %). Kinder unter drei Jahren produzieren kein Methan (Peled *et al.*, 1985; Florin *et al.*, 2000). Genetische Veranlagungen scheinen für die Methanproduktion keine Rolle zu spielen.

Mit Ausnahme der *Methanosarcina*-Spezies zeichnen sich methanogene Archaeen durch sehr kleine Genome von weniger als 2 Mbp Länge aus. Auf Basis dieser eingeschränkten genetischen Ressourcen ist es *Msp. stadtmanae* gelungen, sich an das Wachstum im Gastrointestinaltrakt des Menschen anzupassen und dort besondere Formen des Stoffwechsels zu entwickeln. In der vorliegenden Arbeit wurden die genetischen Grundlagen dieser Anpassungen untersucht und in Bezug auf die Genomorganisation dargestellt, die die Physiologie von *Msp. stadtmanae* hervorgebracht hat.

2 Material und Methoden

2.1 Organismen und Plasmide

Die im Rahmen dieser Arbeit verwendeten Bakterien-Stämme sind in Tab. 2.1 dargestellt. Der sequenzierte Bakterien-Stamm *Ralstonia eutropha* H16 wurde ursprünglich als *Hydrogenomonas eutropha* isoliert (Wilde, 1962) und später umbenannt zu *Alcaligenes eutrophus* (Davis *et al.*, 1969; Davis *et al.*, 1970), *Ralstonia eutropha* (Yabuuchi *et al.*, 1995), *Wautersia eutropha* (Vaneechoutte *et al.*, 2004) und *Cupriavidus necator* (Vandamme & Coenye, 2004).

Tab. 2.1: Verwendete Bakterien-Stämme.

Stamm	Genotyp	Referenz
<i>Ralstonia eutropha</i> H16	Stammsammlung: DSM428, ATCC 17699	als <i>Hydrogenomonas eutropha</i> : Wilde (1962)
<i>Methanosphaera stadtmanae</i>	Stammsammlung: JCM11832, ATCC 43021	Miller & Wolin (1985)
<i>Escherichia coli</i> TOP10	F ⁻ , <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)$, $\phi 80lacZ\Delta M15$, $\Delta lacX74$, <i>recA1</i> , <i>araD139</i> , $\Delta(ara-leu)7697$, <i>galU</i> , <i>galK</i> , <i>rpsL</i> , (Str ^R), <i>endA1</i> , <i>nupG</i>	Fa. Invitrogen GmbH, Karlsruhe

Tab. 2.2 gibt eine Übersicht über verwendete Plasmid- und Cosmid-Vektoren.

Tab. 2.2: Verwendete Plasmid- und Cosmid-Vektoren

Plasmid/ Cosmid	Größe [kbp]	Merkmal	Verwendung	Referenz
pGEM-3Z	2.7	<i>lacZ</i> , MCS, Amp ^r	<i>R. eutropha</i> H16: Plasmid-Genbank	Fa. Promega, Mannheim
Lorist6	5.3	kan ^r	<i>R. eutropha</i> H16: Cosmid-Genbank	Gibson <i>et al.</i> (1987)
pCR2.1- TOPO	3.9	kan ^r , amp ^r	<i>R. eutropha</i> H16, Chromosom 2: GenomiPhi-Genbank	Fa. Invitrogen
pCR4-TOPO	4.0	kan ^r , amp ^r	<i>Methanosphaera stadtmanae</i> : Plasmid-Genbank	Fa. Invitrogen

2.2 Nährmedien und Puffer

Alle aufgeführten Medien wurden durch Autoklavieren für 30 min bei 121 °C sterilisiert. Glucose wurden in separater Lösung autoklaviert oder sterilfiltriert und dem Medium unmittelbar vor dem Animpfen zugegeben. Anaerobe Medien wurden nach der von Costilow (1981) beschriebenen Methode unter Begasung mit sauerstofffreiem Stickstoff hergestellt. Zur Herstellung von Agarplatten wurde das Medium vor dem Autoklavieren mit 1.5 % (w/v) Agar versetzt.

2.2.1 Medien für *Escherichia coli*

Die Zusammensetzungen von LB-Medium und TB-Medium wurden aus Sambrook *et al.* (1989) entnommen.

LB-Medium		TB-Medium	
Trypton	10 g	Trypton	12 g
Hefeextrakt	5 g	Hefeextrakt	24 g
NaCl	10 ml	Glycerin (87 %, v/v)	4 ml
H ₂ O _{bidest}	ad 1000 ml	H ₂ O _{bidest}	ad 1000 ml

Für die Anzucht in *flat bottom blocks* (Fa. Qiagen, Hilden) wurden zu 400 ml TB-Medium je 8 ml 5x Phosphatpuffer und Glucose-Stammlösung (10 %, v/v) gegeben.

2.2.2 Medienzusätze

Antibiotika-Stammlösungen sowie IPTG und X-Gal wurden dem autoklavierten und auf mindestens 50 °C abgekühlten Medium zur Selektion entsprechender rekombinanter *E. coli*-Stämme zugesetzt. Verwendete Konzentrationen der verschiedenen Medienzusätze sind in Tab. 2.3 zusammengefaßt.

Tab. 2.3: Medienzusätze

Zusatz	Stammkonzentration	Endkonzentration
Ampicillin	100 mg/ml in H ₂ O _{bidest} ^a	100 µg/ml
Kanamycin	100 mg/ml in H ₂ O _{bidest} ^a	25-50 µg/ml
IPTG	100 mg/ml in H ₂ O _{bidest} ^a	40 µg/ml
X-Gal	40 mg/ml in N,N-Dimethylformamid ^b	40 µg/ml

^a bei 4 °C gelagert^b vor Gebrauch frisch angesetzt

2.2.3 Stamm- und Pufferlösungen

10x TE-Puffer		5x Phosphat-Puffer	
1 M Tris-HCl (pH 8.0)	100 ml	KH ₂ PO ₄	57.75 g
0.5 M EDTA (pH 8.0)	10 ml	K ₂ HPO ₄	313.50 g
H ₂ O _{bidest}	ad 1000 ml	H ₂ O _{bidest}	ad 500 ml

2.3 Zellanzucht und Stammhaltung

2.3.1 Zellanzucht

2.3.1.1 Anzucht von *Ralstonia eutropha* H16

Zur Anzucht von *R. eutropha* H16 wurde ein Mineralmedium nach Schlegel *et al.* (1961), modifiziert nach Friedrich *et al.* (1982), verwendet. Als organische Kohlenstoffquelle diente Fruktose in einer Endkonzentration von 0.4 % [w/v]. Als Stickstoffquelle wurde Ammoniumchlorid (0.2 % [w/v]) zugesetzt (FN-Medium). Für die DNA-Isolierung wurden die Zellen bis zur frühen stationären Phase angezogen und dann geerntet. Die erhaltenen Zellen wurden bei -80 °C aufbewahrt.

2.3.1.2 Anzucht von *Methanospaera stadtmannae*

Methanospaera stadtmannae wurde unter anaeroben Bedingungen bei 37 °C nach der Methode von Sparling *et al.* (1993) angezogen. Methanol und H₂ dienten als Energiequelle, H₂ und CO₂ als Elektronendonator und Kohlenstoffquelle. Das Medium enthielt zusätzlich 10 % filtriertes, zentrifugiertes und autoklaviertes Pansenfluid. Die Kultivierung erfolgte in mit Butyl-Gummistopfen verschlossenen 2 l-Schottflaschen. Das Kulturvolumen betrug 1 l. Zum Anaerobisieren des Mediums wurde dieses in einem Anaerobenzelt unter N₂/H₂ (95 % : 5 %) gerührt, bis der dem Medium zugesetzte Redoxindikator Resazurin farblos wurde. Anschließend wurde das Medium in die Kulturflaschen abgefüllt und im verschlossenen Zustand autoklaviert. Zuvor wurde noch Na₂S als Schwefelquelle und zum Anaerobisieren zugeführt. Die Gasphase wurde anschließend gegen H₂/CO₂ (80 % : 20 %) (0.7 bar) ausgetauscht und das Medium 10 %ig mit einer frischen Vorkultur angeimpft. Die Kulturflaschen wurden auf einem Inkubationsschüttler (RFI-125, Fa. Infors, Bottmingen, Schweiz) bei 150 U/min und 37 °C inkubiert. Die Gasphase wurde zwei mal täglich gegen H₂/CO₂ (80 % : 20 %) ausgetauscht. Beim Erreichen einer optischen Dichte von 2.5 (gemessen bei 578 nm) wurde die Kultur mit einer Durchflußzentrifuge (Heraeus Contifuge 17RS, Fa. Heraeus Sepatech GmbH, Osterode) bei 4 °C abzentrifugiert. Die erhaltenen Zellen wurden bei -80 °C aufbewahrt.

2.3.1.3 Anzucht von *Escherichia coli*

Die Anzucht von rekombinanten *Escherichia coli* DH5 α -Zellen erfolgte in 96er *flat bottom blocks* (Fa. Qiagen), die nach dem Animpfen mit luftdurchlässigen Folien (*aerosheets*) abgeklebt wurden. Die einzelnen 96 Wells der Blöcke wurden jeweils mit 1.3 ml TB-Medium, Ampicillin, Glucose und Phosphatpuffer befüllt. Die Inkubation erfolgte unter Schütteln (360 rpm) für 20-26 Stunden bei 37 °C im Brutschrank.

2.3.2 Stammhaltung

Zur kurzfristigen Stammhaltung wurden die rekombinanten *E. coli*-Stämme auf LB-Agar-Platten ausgestrichen, über Nacht bei 37 °C inkubiert und dann bei 4 °C gelagert. Für die Stammhaltung der rekombinanten *E. coli*-Klone der Sequenzierungs-Genbanken im 96er Format wurden 1.3 ml-Kulturen wie oben beschrieben inkubiert und in Mikrotiterplatten mit 230 µl TB-Medium pro Ansatz überstempelt. Diese Platten wurden über Nacht bei 37 °C inkubiert, anschließend mit 23 µl DMSO pro Ansatz versetzt und bei -80 °C gelagert.

2.4 Standardtechniken für die Arbeit mit Nukleinsäuren

2.4.1 Geräte und Lösungen

Zur Inaktivierung von Nukleasen wurden hitzestabile Geräte und Lösungen autoklaviert (20 min bei 121 °C). Nicht autoklavierbare Materialien wurden mit 70 % Ethanol (v/v) gespült oder abgeflammt. Hitzelabile Biochemikalien wurden nach Möglichkeit in Lösung sterilfiltriert.

2.4.2 Dialyse von Nukleinsäuren

Um DNA-Lösungen von Salzen und anderen niedermolekularen Substanzen zu befreien, wurden sie einer Mikrodialyse unterzogen. Dieses Verfahren erhöhte die Ligations-effizienz (siehe 2.6.1.1) und ermöglichte die Transformation durch Elektroporation, die durch Salze gestört wird. Die Dialyse wurde mit auf bidestilliertem Wasser schwimmenden sterilen Membranfiltern durchgeführt (VS-Filter, 0.025 µm Porengröße, Fa. Millipore, Eschborn). Darauf wurden 20-50 ml für ca. 60 min dialysiert und danach vorsichtig abgenommen.

2.4.3 DNA-Konzentrationsbestimmung und Reinheitskontrolle

Die Konzentration wässriger DNA-Lösungen wurde photometrisch über die optische Dichte (OD) bestimmt. Dazu wurde im Photometer (Uvikon 810 Photometer, Kontron Instruments, Eching) die Absorption bei 260 nm in einer Quarzküvette der Schichtdicke 1 cm gemessen. Eine Absorption bei 260 nm von 1 gegen Wasser (OD_{260}) entspricht einer Konzentration von 50 $\mu\text{g} / \text{ml}$ doppelsträngiger DNA bzw. 40 $\mu\text{g} / \text{ml}$ einzelsträngiger DNA/RNA bzw. 31 $\mu\text{g} / \text{ml}$ Oligonukleotiden (Sambrook *et al.*, 1989). Um den Reinheitsgrad der DNA-Lösung abzuschätzen, wurden zusätzlich die optischen Dichten bei 230 nm und 280 nm bestimmt. Verunreinigung mit Protein führen zu einer verstärkten Absorption bei 280 nm, Verunreinigung mit Polysacchariden zur verstärkten Absorption bei 230 nm. Für reine DNA gilt:

$$\begin{aligned} OD_{260} : OD_{280} &= 1.8 \\ OD_{230} : OD_{260} : OD_{280} &= 0.45 : 1 : 0.515 \end{aligned}$$

DNA-Konzentrationen von Plasmiden aus der automatisierten Präparation mit dem *BioRobot* 8000 (Fa. Qiagen; siehe 2.4.7) wurden in 96er Mikrotiterplatten eluiert und im Plattenphotometer gemessen (PowerWave X, Fa. Bio-Tek Instruments, Winooski, Vt., USA). Zur groben Abschätzung von DNA-Konzentrationen wurden Aliquots der DNA-Lösungen elektrophoretisch im Agarosegel aufgetrennt und die Stärke der Banden mit einem Marker bekannter Konzentration verglichen.

2.4.4 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese wurde zur analytischen und präparativen Auftrennung von DNA genutzt. Die Auftrennung erfolgte in horizontalen Kammern der Größe 11 cm x 6.5 cm. In TAE-Puffer wurden 0.8 %ige Agarosegele (Seakem LE Agarose, Fa. Cambrex, Me., USA) angesetzt. Derselbe Puffer diente auch als Laufpuffer während der Elektrophorese. Vor der Elektrophorese wurden die Proben mit 0,2 Vol. *loading buffer* (Fa. Fermentas, St. Leon-Rot) versetzt, um die DNA zu beschweren und die Lauffront während der Auftrennung sichtbar zu machen. Als Größenreferenz wurden 3 μl eines

Größenstandards aufgetragen (*GeneRuler DNA Ladder Mix*, Fa. Fermentas). Nach der Elektrophorese wurden die Gele in einem Ethidiumbromid-Bad (1 µg/ml) für ca. 10 min gefärbt und dann gewässert, um überschüssiges Ethidiumbromid zu entfernen. Die Visualisierung der DNA-Banden erfolgte durch UV-Licht (254 nm) mit einem Gel-Dokumentationssystem (Multimage LightCabinet, Fa. Alpha Innotech, Ca., USA).

20x TAE-Puffer

Tris	96.9 g
Eisessig	22.8 ml
EDTA	7.5 g
H ₂ O _{bidest}	ad 1000 ml

2.4.5 Saccharose-Dichtegradientenzentrifugation

Eine Auftrennung der DNA durch Dichtegradientenzentrifugation (Sambrook *et al.*, 1989) wurde bei der Herstellung von Genbanken vorgenommen (siehe 2.8). Diese Methode hatte den Vorteil, daß die DNA-Fragmente nahezu unbeschädigt aufgereinigt wurden. Die Zentrifugation wurde mit einem Saccharose-Gradienten von 10 % bis 40 % (w/v) in 10 ml Zentrifugenröhrchen (Sorvall Polyallomer 3699, Fa. DuPont, Dreieich) durchgeführt. Dazu wurden unterschiedliche Saccharose-Lösungen (10 %, 17.5 %, 25 %, 32.5 %; w/v) in der Stammlösung für den Dichtegradienten angesetzt. Die Lösungen wurden beginnend mit der höchsten Konzentration in die Zentrifugenröhrchen geschichtet und eingefroren. Auf den Gradienten wurden zwischen 100 und 200 g DNA aufgetragen. Die anschließende Zentrifugation erfolgte mit 27 000 rpm für 24 h bei 20 °C in einem *swing out*-Rotor (Sorvall TH641) in einer Ultrazentrifuge (Sorvall OTD 55 B, beide Fa. Kendro Laboratory Products GmbH, Langenselbold). Nach der Zentrifugation wurde der Gradient vorsichtig in Fraktionen von 300 µl von oben nach unten entnommen. Zur anschließenden Fällung der DNA wurden die Fraktionen mit 0.5 Vol. 20 %iger PEG-Lösung (v/v) in 2.5 M NaCl versetzt und über Nacht bei RT inkubiert. Am nächsten Tag wurde die Lösung 30 min bei 22 650 rpm zentrifugiert. Nachdem der Überstand abgenommen und das Pellet in 500 µl Ethanol (70 %; v/v) aufgenommen worden war, erfolgte eine weitere Zentrifugation unter gleichen Bedingungen. Der Ü-

Überstand wurde abgenommen, das Pellet getrocknet und in 30 µl autoklaviertem H₂O_{bidest} gelöst. Zur Überprüfung der Fraktionen wurden 10 µl von jeder dritten Fraktion auf einem Agarosegel überprüft (siehe 2.4.4).

2x Dichtegradienten-Stammlösung

NaCl	35.06 g
Tris	1.45 g
EDTA	1.12 g
H ₂ O _{bidest}	ad 300 ml
	pH 8

2.4.6 Isolierung chromosomaler DNA

Für die Isolierung chromosomaler DNA wurden zwei verschiedene Kits verwendet: *AquaPure Genomic DNA Kit* (Bio-Rad Laboratories, Ca., USA) und *E.Z.N.A. Bacterial DNA Kit* (Fa. PEQLAB Biotechnologie GmbH, Erlangen). Der Vorteil des Bio-Rad-Kits lag in der guten Qualität der präparierten DNA. Demgegenüber lieferte das PEQLAB-Kit eine höhere Ausbeute an DNA, die insgesamt jedoch stärker geschert vorlag. Für PCR-Amplifikationen wurde i.d.R. genomische DNA aus Bio-Rad-Präparationen verwendet.

Zellen aus ca. 500 µl Übernacht-Kultur bzw. aus einer entsprechenden Menge eingefrorener Zellen wurden für die Isolierung verwendet und daraus die genomische DNA nach Protokoll präpariert. Bei der Isolierung mit dem Bio-Rad-Kit wurde ein zusätzlicher lytischer Schritt eingeführt. Dazu wurden die Zellen nach Inkubation mit *Genomic DNA Lysis Solution* mit 1.5 µl Proteinase K (20 mg/ml) versetzt und für 1 h bei 55 °C lysiert. Die isolierte chromosomale DNA wurde in 75-100 µl *Hydration Solution* (Bio-Rad), bzw. in zwei Schritten in 100 µl und 50 µl Elutions-Lösung (PEQLAB) aufgenommen. Kurz- und mittelfristig wurde chromosomale DNA bei 4 °C gelagert, für längere Aufbewahrung aliquotiert und bei -20 °C gelagert.

2.4.7 Isolierung von Plasmid DNA

Für die Isolierung von Plasmid-DNA wurden je nach Anzahl der Proben verschiedene Methoden verwendet. Für die Präparation weniger Proben, beispielsweise für Restriktionsverdauere rekombinanter Plasmide (siehe 2.5.1), wurde das *QIAprep Spin Miniprep Kit* (Fa. Qiagen) genutzt und nach Protokoll vorgegangen. Die Präparation rekombinanter Plasmide ganzer Genbänke wurde dagegen automatisiert an Robotern des Typs *BioRobot 8000* (Fa. Qiagen) mit Hilfe des *QIAprep 96 Turbo BioRobot Kit* durchgeführt. Die Plasmid-Isolation selbst basiert auf einer alkalischen Lyse mit anschließender Aufreinigung der DNA durch Bindung an eine Silikat-Gel-Membran. Hierzu wurden die Klone wie unter 2.3.1.3 beschrieben im 96er Maßstab in *flat bottom blocks* angezogen und von Robotern weiterverarbeitet. Nach der Anzucht wurden die Blöcke für 20 min bei 3300 rpm abzentrifugiert (Megafuge 2.0, Fa. Heraeus, Hanau) und der Überstand verworfen. Die Zellpellets wurden im Block in den *BioRobot 8000* überführt und automatisch weiterbearbeitet. Die gebundene DNA wurde in mehreren Schritten gewaschen und mit 120 µl Wasser (LiChrosolv, Fa. Merck KGaA, Darmstadt) gelöst und in 96er Mikrotiterplatten mit Flachboden (UV-geeignet; Fa. Greiner Bio-one, Frickenhausen) gesammelt.

2.4.8 Aufreinigung von DNA-Fragmenten

Um überschüssige Nukleotide, Primer, Puffer und Enzymreste aus dem PCR-Ansatz zu entfernen, wurden PCR-Produkte generell aufgereinigt. Ähnlich wie bei der Präparation von Plasmid-DNA (siehe 2.4.7) wurde die DNA je nach Anzahl der Proben manuell oder automatisiert präpariert. Für diese manuelle Aufreinigung wurde das *QIAquick PCR Purification Kit* (Fa. Qiagen) verwendet. Ergänzend zu den Vorschriften des Protokolls wurde nach der Fällung der DNA ein zusätzlicher Zentrifugationsschritt durchgeführt, um Reste des PE-Puffer vollständig zu entfernen.

2.4.9 Isolation von DNA-Fragmenten aus Agarosegelen

DNA-Fragmente wurden z.B. aus Agarosegelen isoliert, wenn eine PCR-Reaktion zur unspezifischen Amplifikation mehrerer Produkte geführt hatte. Um mehrere Produkte einzeln sequenzieren zu können, wurde nach der Auftrennung im Agarosegel jede Bande mit dem Skalpell ausgeschnitten. Die Entfernung von Gelrückständen und die Aufreinigung der DNA wurde mit dem *QIAquick Gel Extraction Kit* (Fa. Qiagen) durchgeführt und die isolierte DNA direkt für die Sequenzierung verwendet.

2.4.10 Scheren von DNA

Für die Herstellung von Genbanken war eine gleichmäßige Fragmentierung der chromosomalen DNA nötig, um eine proportionale Verteilung des gesamten Genoms auf die Genbank zu gewährleisten. Das mechanische Scheren der DNA bot den Vorteil, daß es gegenüber der Restriktion unabhängig von Schnittstellen war.

Die chromosomale DNA von *Methanosphaera stadtmanae* wurde mit Hilfe eines *nebulizer* (Fa. Invitrogen) geschert. Bei diesem Gerät wird die DNA-Lösung in einen nach unten hin spitz zulaufenden Zylinder gegeben, in dem die DNA durch Druckluft verwirbelt und dabei geschert wird. Für eine Scherung wurden etwa 15 µg DNA mit 750 µl *shearing buffer* in den *nebulizer* gegeben, der *nebulizer* fest verschraubt und über eine Stickstoffleitung ein Druck von 0.62 bar angelegt. Während des gesamten Schervorgangs wurde der *nebulizer* auf Eis gelagert. Die Dauer des Scherens wurde in Abhängigkeit von der gewünschten Fragmentgröße empirisch ermittelt. Zur Weiterverarbeitung der gescherten DNA wurde die DNA-Lösung aus dem *nebulizer* in einer *SpeedVac* auf ca. 200 µl eingedampft.

Shearing Buffer

Tris	121.1 g
EDTA	37.23 g
Glycerin	10 ml
	pH 8

2.5 Modifikation von Nukleinsäuren

2.5.1 Restriktion

Die enzymatische Restriktion mit Enzymen bekannter Spezifität wurde dazu genutzt, DNA an vorhersagbaren Stellen zu schneiden. Chromosomale DNA wurde zur Herstellung von Genbanken geschnitten. Die Restriktion von Plasmiden wurde für die Erstellung von Genkarten genutzt, mit deren Hilfe die korrekte Assemblierung der Plasmid-Sequenzläufe im Datensatz überprüft werden konnte.

Für die Reaktionsansätze wurden für jedes Enzym die jeweiligen Puffersysteme unter den angegebenen Reaktionsbedingungen der Hersteller verwendet. Pro Mikrogramm DNA wurden 3-5 U Enzym eingesetzt. Zur vollständigen Restriktion von Plasmiden oder Cosmiden wurden die Ansätze für ca. 2 Stunden inkubiert. Bei der partiellen Restriktion wurde die Inkubationsdauer variiert, um eine möglichst statistische Verteilung der Fragmente zu erreichen. Zur Inaktivierung der Restriktionsenzyme wurden die Ansätze nach der Restriktion für ca. 20 Minuten bei 65 °C inkubiert oder mit 50 mM EDTA versetzt. Durch die Komplexbildung freier Mg^{2+} -Ionen entzieht EDTA dem Enzym die für die Restriktion benötigten Cofaktoren.

2.5.2 Dephosphorylierung

Bei 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 mehrere Inserts aneinander ligieren.

Die Insert-DNA für die Ligation wurde im Restriktionsansatz mit 1 U alkalischer Phosphatase pro 1 µg DNA versetzt und 1 h bei 37 °C inkubiert. Danach wurde die dephosphorylierte DNA mit Hilfe des *QIAquick PCR Purification Kit* aufgereinigt (siehe 2.4.8).

2.5.3 Herstellung glatter DNA-Enden (*blunt ends*)

Beim Scheren der chromosomalen DNA entstanden undefinierte Enden, die eine Ligati-on mit dem Vektor erschwerten. Aus diesem Grund wurden vor der Klonierung die En-den der gescherten DNA-Fragmente mit Hilfe der T4-DNA-Polymerase zu *blunt ends* aufgefüllt. Der folgende Ansatz wurde für 1.5 h bei RT inkubiert und anschließend mit dem *QIAquick PCR Purification Kit* (siehe 2.4.8) aufgereinigt.

50 µl Ansatz

Insert-DNA	ca. 38 µl
5x Puffer für T4-DNA-Polymerase	10 µl
dNTP-Mix (je 10 mM)	1.5 µl
T4-DNA-Polymerase	1 µl

2.5.4 Herstellung von Insert-DNA mit A-Überhang

Bei der Klonierung mit dem *TOPO TA Cloning[®] Kit* (Fa. Invitrogen) (siehe 2.6.1), das für die Herstellung von Genbanken verwendet wurde, war es erforderlich, daß die In-sert-DNA am 5'-Ende eine Adenin-Base als Überhang aufwies, da der verwendete TO-PO-Vektor einen entsprechenden Thymidin-Überhang enthielt. Glatte Enden wurden von der Taq-Polymerase bei Zugabe von dATP um einen Adenin-Überhang verlängert. Zunächst wurden DNA-Fragmente mit *blunt ends* hergestellt (siehe 2.5.3).

70 µl Ansatz

DNA	50 µl
dATP (0.5 mM)	6 µl
MgCl ₂	6 µl
10x PCR-Puffer ohne MgCl ₂	7 µl
Taq DNA-Polymerase (5 U/µl)	0.5 µl

Die Herstellung von A-Überhängen wurde in dem oben angegebenen Ansatz für 25 min bei 72 °C durchgeführt. Anschließend wurde die DNA mit dem *QIAquick PCR Purification Kit* aufgereinigt (siehe 2.4.8).

2.5.5 Ligation

In der Ligation wurden Phosphodiesterbindungen zwischen doppelsträngigen DNA-Fragmenten gebildet, die freie 3'-Hydroxyl-Enden und 5'-Phosphat-Enden aufwiesen. Um die Ligationseffizienz zu erhöhen, wurde ein Überschuß an Insert-DNA gegenüber der Vektor-DNA eingesetzt (ca. 1 : 3). Für die Ligation wurde T4-DNA-Ligase mit entsprechendem Puffersystem eingesetzt (Fa. Promega). Der folgende Reaktionsansatz wurde über Nacht bei 16 °C bzw. 2 h bei 37 °C inkubiert.

20 µl Ansatz

Vektor-DNA	0.2-2 µl
Insert-DNA	0.8-8 µl
10x T4-Ligase-Puffer	2 µl
T4-DNA-Ligase (1 U/µl)	1-2 µl
H ₂ O _{bidest}	ad 20 µl

2.6 Transformation und Selektion

2.6.1 TOPO TA-Klonierung

Die Klonierung mit dem *TOPO TA Cloning[®] Kit* (Fa. Invitrogen) lieferte eine Methode zur direkten Klonierung von DNA-Fragmenten mit A-Überhang und wurde für die Herstellung von Genbanken aus gescherter, chromosomaler DNA und aus PCR-Fragmenten verwendet.

Das Kit macht sich dabei die Eigenschaft der Taq-Polymerase zunutze, an das 3'-Ende amplifizierter DNA eine einzelne Adenin-Base zu hängen. Der in dem Kit verwendete

TOPO-Vektor besitzt in entsprechender Weise einen Thymidin-Überhang, der durch eine spezielle Topoisomerase erzeugt wird, die bis zur Ligation mit dem Insert kovalent gebunden bleibt. Auf diese Weise wird einerseits die Ligation mehrerer Inserts miteinander verhindert als auch der enzymatische Verdau des Vektors erschwert (Shuman, 1991).

2.6.1.1 Ligation mit dem TOPO - Vektor

Der folgende Ansatz wurde 2 h bei RT inkubiert und danach zur Transformation chemisch kompetenter *E. coli* TOP10 -Zellen verwendet (2.6.1.2):

4 µl Ansatz

Insert DNA mit A-Überhang	3 µl
Salt Solution (TOPO-Kit, Fa. Invitrogen)	0.5 µl
PCR TOPO-Vektor (Fa. Invitrogen)	0.5 µl

2.6.1.2 Transformation von *E. coli* TOP10 - Zellen

Der gesamt Reaktionsansatz der TOPO-Ligation (siehe 2.6.1.1) wurde für die Transformation eingesetzt. Dazu wurden die 4 µl Reaktionsansatz zu 70 µl *E. coli* TOP10-Zellen gegeben und der Ansatz zuerst 10 min auf Eis und anschließend 45 s bei 42 °C inkubiert. Danach wurden die Zellen sofort auf Eis gestellt und mit 400 µl LB-Medium vermischt. Die transformierten Zellen wurden anschließend für 1.5 h bei 37° C und 180 rpm geschüttelt. Die Selektion erfolgte auf LB-Platten mit Ampicillin, die über Nacht bei 37 °C inkubiert wurden.

2.6.2 Blau-Weiß *screening* (X-Gal-Test)

Das Blau-Weiß *screening* diente neben der Antibiotikaresistenz der schnellen Selektion erfolgreich transformierter Klone. Anhand der Färbung konnten solche Klone, die ein

rekombinantes Plasmid mit Insert tragen, von denjenigen unterschieden werden, die nur den leeren Vektor enthielten.

Das Enzym β -Galactosidase spaltet neben dem eigentlichen Substrat Lactose auch 5-Brom-4-Chlor-Indoyl- μ -D-Galactosid (X-Gal). Dabei entsteht der Farbstoff 5-Brom-4-Chlor-Indigo, der sich in Anwesenheit von Sauerstoff spontan blau verfärbt (Horwitz *et al.*, 1964). Im X-Gal-Test werden *E. coli*-Stämme eingesetzt, die durch eine Deletion im Chromosom eine inaktive β -Galactosidase ohne N-Terminus bilden. Geeignete Vektoren sind in der Lage, diesen Effekt zu kompensieren. Sie tragen neben dem Promotor- und Operatorbereich des *lac*-Operons auch das 5'- Ende des *lacZ*-Gens, welches für das sogenannte α -Peptid kodiert. Dies kann *in vivo* mit der verkürzten β -Galactosidase zum aktiven Enzym assoziieren (α -Komplementation). Da die multiple Klonierungsstelle des Vektors innerhalb des *lacZ*-Gens liegt, wird die Bildung des α -Peptids verhindert, wenn der Vektor ein Insert trägt. Auf Indikatorplatten, die IPTG (zur Induktion der β -Galactosidase) und X-Gal enthalten, können blaue Kolonien, die nur den Plasmid-Vektor enthalten von weiße Kolonien, die mit hoher Wahrscheinlichkeit ein Plasmid mit Insert tragen, unterschieden werden.

2.7 Amplifikationen

2.7.1 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion (PCR) ermöglicht die hocheffiziente *in vitro*-Amplifikation spezifischer DNA-Bereiche. Sie wurde verwendet, um gezielt einzelne Bereiche des Genoms zu amplifizieren, um Assemblierungen im Datensatz zu überprüfen und um Sequenzlücken, die von der Genbank nicht erfaßt wurden, zu überbrücken.

Die Amplifikationsreaktion wurde mit zwei unterschiedlichen Ansätzen durchgeführt. Für die Amplifikation kurzer DNA-Bereiche (0.5-5.0 kbp), wie sie zur Erhöhung der Sequenzabdeckung durchgeführt wurde, kam eine Taq-Polymerase (Fa. Eppendorf, Hamburg) zum Einsatz (siehe Standard-Ansatz). Längere Amplifikationen (4-8 kbp), z.B. über größere Sequenzlücken hinweg, wurden mit dem *TripleMaster long range* - Ansatz des *TripleMasterTM PCR-Systems* durchgeführt. Als Template diente genomische DNA bzw. Plasmid- und Cosmid-DNA. Die verwendeten Primer wurden in Länge

(17-25 nt) und Schmelztemperatur (*R. eutropha* H16: 56-62 °C, *Msp. stadtmanae*: 44-47 °C) je nach Beschaffenheit des zu amplifizierenden Bereiches variiert. Die Berechnung des Schmelzpunktes erfolgte nach Chester & Marshak (1993):

$$T_m = 69.3 + 0.41 * x - 650/L$$

x = G+C-Gehalt [%]
L = Länge des Primers [bp]

Die Spezifität der Amplifikation konnte durch Zugabe von DMSO bzw. durch Verringerung der MgCl₂-Konzentration im Standard-Ansatz erhöht werden.

Standard-Ansatz	<i>Msp. stadtmanae</i>	<i>R. eutropha</i> H16
Template-DNA	1 µl	1 µl
Primer (5 pmol)	je 4 µl	je 4 µl
dNTP-Gemisch (10 mM)	1 µl	1 µl
Taq DNA-Polymerase (5 U/µl)	0.3 µl	0.3 µl
10x PCR-Puffer ohne MgCl ₂	5 µl	5 µl
MgCl ₂	3 µl	0.5 µl
DMSO	-	2.5 µl
H ₂ O _{bidest}	ad 50 µl	ad 50 µl

Der Standard-Ansatz für *R. eutropha* H16 und *Msp. stadtmanae* und der *TripleMaster*-Ansatz für *Msp. stadtmanae* wurden in einem Master-Mix zusammenpipettiert. Der *TripleMaster*-Ansatz für *R. eutropha* wurde dagegen in zwei Ansätzen (Master I und Master II) zusammenpipettiert, die erst direkt vor der Reaktion vereinigt wurden.

<i>TripleMaster</i> -Ansatz (long range)	<i>Msp. stadtmanae</i>	<i>R. eutropha</i> H16
Template-DNA	2 µl	Master I: 1 µl
Primer (5 pmol)	je 3 µl	je 4 µl
H ₂ O _{bidest}	17.84 µl	11 µl
10x <i>Tuning Buffer</i>	0.16 µl	Master II: 0.3 µl
dNTP-Gemisch (10 mM)	5 µl	2.5 µl
<i>TripleMaster</i> -Polymerase	1 µl	0.4 µl

Tab. 2.4: PCR-Programme, Standard-Ansatz.

<i>R. eutropha</i> H16			<i>Msp. stadtmanae</i>		
LID	110 °C	ON	LID	110 °C	ON
PAUSE	98 °C	FOREVER	PAUSE	98 °C	FOREVER
TEMP	96 °C	2 min	TEMP	96 °C	
LOOP [30X		LOOP [30X	
TEMP	96 °C	20 s	TEMP	96 °C	20 s
TEMP	55 °C	30 s	TEMP	42 °C	30 s
TEMP	72 °C	3 min	TEMP	72 °C	3 min
LOOP]			LOOP]		
TEMP	72 °C	10 min	TEMP	72 °C	10 min
LID		OFF	LID		OFF
TEMP	4 °C	FOREVER	TEMP	4 °C	FOREVER
END			END		

Tab. 2.5: PCR-Programme, TripleMaster-Ansatz (long range)

<i>R. eutropha</i>			<i>M. stadtmanae</i>		
LID	110 °C	ON	LID	110 °C	ON
PAUSE	93 °C	FOREVER	PAUSE	96 °C	FOREVER
TEMP	93 °C	3 min	TEMP	96 °C	2 min
LOOP [10X		LOOP [10X	
TEMP	93 °C	15 s	TEMP	96 °C	15 s
TEMP	55 °C	30 s	TEMP	42 °C	30 s
TEMP	68 °C	8 min	TEMP	68 °C	5 min
LOOP]			LOOP]		
LOOP [10X		LOOP [10X	
TEMP	93 °C	15 s	TEMP	96 °C	15 s
TEMP	55 °C	30 s	TEMP	42 °C	30 s
TEMP	68 °C	8 min	TEMP	68 °C	5 min
LOOP]		(inc. 20 s) ^a	LOOP]		(inc. 20 s) ^a
TEMP	72 °C	10 min	TEMP	72 °C	
LID		OFF	LID		10 min
TEMP	4 °C	FOREVER	TEMP	4 °C	OFF
END			END		FOREVER

^a inc. = *increase*: Elongationsschritt wird in jedem Durchlauf um 20 s verlängert.

Die in den Tab. 2.4 und Tab. 2.5 dargestellten Programme für die PCR-Amplifikation mit Standard- und *TripleMaster*-Ansatz wurden gegebenenfalls angepaßt, um ungewöhnliche DNA-Bereiche zu amplifizieren. Durch Erhöhung der *annealing*-Temperatur und Verlängerung der Denaturierungs-Schritte konnte die Spezifität der Amplifikationen bzw. die Erfolgsquote für Bereiche hoher G+C-Konzentrationen verbessert werden.

2.7.2 GenomiPhi-Amplifikation

Das *GenomiPhiTM DNA Amplification Kit* (Amersham Biosciences, Freiburg) wurde für die hocheffiziente, unspezifische Amplifikation gesamtgenomischer DNA genutzt. Es beruht auf der Verwendung einer hochprozessiven modifizierten Phi29-Polymerase mit Strang-Verdrängungs-Eigenschaften in Verbindung mit einem Gemisch zufälliger Hexamer-Primer. Innerhalb des *R. eutropha* H16-Genomprojektes wurden geringe Mengen des aufgereinigten Chromosoms 2 nach Protokoll mit dem GenomiPhi-Kit amplifiziert und für die Herstellung einer selektiven Genbank genutzt.

2.8 Genbanken

Eine gute Genbank ist die Grundlage eines erfolgreichen Genomprojektes. Bei der Sequenzierung mit dem *whole shotgun approach* ist es entscheidend, eine gleichmäßige Verteilung des gesamten Genoms auf die Genbank zu erreichen. Bei der Sequenzierung einer Genbank, in der das Genom statistisch verteilt vorliegt, kann die Abnahme der verbliebenen Sequenz-Lücken nach der Assemblierung in Abhängigkeit von der Anzahl der Sequenzläufe nach der Formel von Lander & Waterman (1988) berechnet werden. Für die Sequenzierung größerer Genomprojekte bietet sich die Herstellung mehrerer Genbanken unterschiedlicher Insertgröße an. Dabei kann eine gleichmäßige Abdeckung des gesamten Genoms durch Assemblierung einer hohen Anzahl von Sequenzläufen einer Plasmid-Genbank mit geringer Insertgröße erreicht werden. Für die genomweite Anordnung der Contigs aus den assemblierten Sequenzläufen der Plasmid-Genbank kann dagegen eine geringe Anzahl von Sequenzläufen einer Cosmid-Genbank mit langer Insertgröße verwendet werden.

Das verhältnismäßig kleine Genom von *Msp. stadtmanae* (1.8 Mbp) wurde mit Hilfe einer einzelnen Plasmid-Genbank sequenziert. Das große Genom von *R. eutropha* H16 (7.4 Mbp) dagegen wurde mit Hilfe einer Plasmid-Genbank und einer zusätzlichen Cosmidbank sequenziert, die jedoch von der Fa. Integrated Genomics auf Basis gesamtochromosomaler DNA hergestellt wurden.

2.8.1 Plasmidbanken

Für das *R. eutropha*-Genomprojekt wurde neben der von der Fa. Integrated Genomics bezogenen Plasmid- und Cosmid-Genbanken eine weitere selektive Plasmid-Genbank von Chr.2 hergestellt. Dazu wurde Chr.2 einer GenomiPhi-Amplifikation (siehe 2.7.2) unterzogen und partiell mit dem Restriktionsenzym *Bsp143I* geschnitten. Für die Herstellung einer Plasmid-Genbank für das *Msp. stadtmanae*-Genomprojekt wurde gesamtochromosomale DNA mechanisch geschert (siehe 2.4.10). Die resultierenden DNA-Fragmente wurden über einen Saccharose-Dichtegradienten aufgetrennt (siehe 2.4.5), gefällt und dephosphoryliert (siehe 2.5.2). Die geschnittene bzw. gescherte DNA beider Projekte wurde mit Hilfe der T4-DNA-Polymerase zu *blunt ends* aufgefüllt (siehe 2.5.3), die Enden mit Hilfe der Taq-Polymerase mit A-Überhängen versehen (siehe 2.5.4), für die Ligation mit dem TOPO-Vektor eingesetzt (siehe 2.6.1.1) und schließlich in *E. coli* TOP10-Zellen transformiert (siehe 2.6.1.2).

2.9 Sequenzierung

In den Sequenzierungs-Projekten von *R. eutropha* H16 und *Msp. stadtmanae* kamen drei verschiedene Sequenzierungsgeräte zum Einsatz: MegaBACE 1000, MegaBACE 4000 (Fa. Amersham Bioscience) und ABI PRISM 377 (Fa. Applied Biosystems). Diese Geräte arbeiten nach dem Verfahren der Kettenabbruchmethode (Sanger *et al.*, 1977). Bei diesem Verfahren werden der Sequenzreaktion neben den Standard-Desoxynukleotiden (dNTPs) Fluoreszenz-markierte Didesoxynukleotide (ddNTPs, *dye terminators*) zugesetzt. Die markierten ddNTPs werden statistisch mit der gleichen Wahrscheinlichkeit in die DNA-Synthese einbezogen wie dNTPs, führen jedoch zum Abbruch der Polymerase-Reaktion. Das entstandene Fragment trägt eine basenspezifische

sche 5'-Markierung, deren Fluoreszenz vom Sequenzierautomaten detektiert wird. Aus der Integrierung der gemessenen Fluoreszenz der unterschiedlichen Farbstoffe über die Zeit leitet das *base calling* die DNA-Sequenz ab.

2.9.1 Sequenzierung mit dem ABI PRISM[®] 377

Bei den ABI 377-Geräten handelt es sich um Plattensequenzierer. Die Sequenzen werden elektrophoretisch in einem Polyacrylamid-Gel zwischen zwei Glasplatten aufgetrennt. Dazu werden 96 Proben manuell mit Hilfe einer Mehrkanal-Hamilton-Pipette auf das Gel aufgetragen. Der Vorteil dieser Geräte liegt in der hohen Toleranz gegenüber unterschiedlichen eingesetzten DNA-Mengen. Nachteilig wirkt sich der hohe Arbeitsaufwand bei der Präparation der Gele und der manuellen Auftragung der Proben aus. Zudem kann es durch Ausfall einzelner Spuren bzw. Überstrahlung von schwachen Signalen durch starke, benachbarte Signale zu Verwechslungen einzelner Sequenzen kommen.

Für die Sequenzierung am ABI 377 wurde ein „1/2 LGA“-Ansatz zusammengestellt und nach dem Programm in Tab. 2.6 amplifiziert. Um überschüssige Primer und *dye-terminators* zu entfernen, wurde die DNA nach der Amplifikation mit 99 % EtOH (v/v) und 3 M NaAcetat (pH 4.6) gefällt. Sequenziergele wurden nach Tab. 2.7 angesetzt. Vor dem Auftragen auf das Sequenziergel wurden die Proben zunächst mit 1 µl Stop-Mix versetzt. Für die elektrophoretische Auftrennung der DNA wurden die Parameter aus Tab. 2.7 verwendet.

„1/2-LGA“-Sequenzieransatz

Premix	4 µl
(DMSO*)	(1 µl*)
Primer (5 pmol/µl)	1 µl
DNA	400 ng
H ₂ O _{bidest}	ad 20 µl

* nur *R. eutropha* H16

Tab. 2.6: ABI 377-Sequenzierprogramm.

LID	110 °C	ON
PAUSE	98 °C	FOREVER
TEMP	98 °C	2 min
LOOP [25X	
TEMP	96 °C	20 s
TEMP	55 °C ¹ (42 °C ²)	15 s
TEMP	60 °C	4 min
LOOP]		
LID		OFF
TEMP	4 °C	FOREVER
END		

¹ *R. eutropha*, ² *Msp. stadmanae*

Tab. 2.7: ABI 377-Sequenziergel

Zusammensetzung:	
7 M Harnstoff	18 g
40 % PAGE-Plus Polyacrylamid	5.3 ml
10X TTE-Puffer	6 ml
H ₂ O _{bidest}	21.5 ml
10 % APS	250 µl
TEMED	25 µl
Laufparameter:	
Geldicke	0,2 mm
Gellänge	48 cm
Spannung	3 500 V
Stromstärke	60 mA
Leistung	40 W
Temperatur	45 °C
CCD <i>offset</i>	0
CCD <i>gain</i>	4
Vorlaufzeit	90 sek
Laufzeit	1.5 h + 14.5 h
Laufpuffer	2x TTE oben, 1x TTE unten
Minimale Anzahl von <i>scans</i>	22 000

loading buffer

EDTA	25 mM
Blue Dextran	0.5 μ l
H ₂ O _{bidest}	ad 500 ml

Stop-Mix

Formamid : loading buffer	=	5 : 1
---------------------------	---	-------

2.9.2 Sequenzierung mit MegaBACE 1000 bzw. MegaBACE 4000

Bei den Geräten der MegaBACE-Reihe handelt es sich um Kapillarsequenzierungseinheiten mit 96 (MegaBACE 1000) bzw. 384 (MegaBACE 4000) Kapillaren. Bei diesen Sequenzierautomaten wird markierte DNA in dünnen Glaskapillaren von ca. 100 μ m Durchmesser, die eine Polyacrylamid-Matrix enthalten, elektrophoretisch aufgetrennt. Da die Proben mittels angelegter Spannung in kleinsten Mengen in die Kapillaren injiziert werden, können mit jedem Reaktionsansatz fast unbegrenzt viele Sequenzierungsreaktionen durchgeführt werden. Zudem ist keine Möglichkeit zur Verwechslung der Ansätze gegeben, die im Gegensatz zur Sequenzierung im Plattensequenzierer einzeln injiziert werden. Allerdings kann es aufgrund der hohen Empfindlichkeit der Auftragungsmethode leicht zu Verstopfungen der Kapillaren kommen, wenn die Injektionszeit zu lang oder die Injektionsspannung oder die Konzentration der Template-DNA im Reaktionsansatz zu hoch gewählt werden.

10 μ l Sequenzieransatz wurden nach Programm in Tab. 2.8 amplifiziert und danach mit 24 μ l 80 % Isopropanol (v/v) und 40 μ l 80 % EtOH gefällt.

MegaBACE-Sequenzieransatz

DNA	400 ng
Primer (5 pmol)	1 μ l
<i>DYEnamic ET Dye Terminator Kit</i>	4 μ l
H ₂ O _{HPLC}	ad 10 μ l

Tab. 2.8: MegaBACE-Sequenzierprogramm

LID	110 °C	ON
PAUSE	95 °C	FOREVER
TEMP	95 °C	5 min
LOOP [30X	
TEMP	95 °C	20 s
TEMP	55 °C ¹ (42 °C ²)	15 s
TEMP	60 °C	1 min
LOOP]		
LID		OFF
TEMP	4 °C	FOREVER
END		

¹ *R. eutropha*, ² *Msp. stadtmanae*

2.10 Prozessierung und Assemblierung der DNA-Sequenzen

Die Prozessierung der Daten aus den Sequenzierungseinheiten erfolgt zu großen Teilen vollautomatisch und mit folgenden Programmen:

- Für das *base calling*, die Zuweisung der DNA-Sequenz zu den chromatographischen Daten aus der Sequenziereinheit, wurde das Programm Phred verwendet (Ewing & Green, 1998). Dabei wird der Sequenz ein Zuverlässigkeitswert (*base confidence*) zugeordnet, der die Fehlerwahrscheinlichkeit für jede einzelne Base angibt.
- In einem *screening*-Schritt wurde die Sequenz auf Übereinstimmungen mit der Sequenz des Plasmid-Vektors und der *E. coli*-Wirtszellen überprüft. Dazu wurde das Programm Pregap aus dem Staden Package verwendet, das unerwünschte Sequenzen für die folgenden Schritte maskiert (Staden, 1996).
- Die Assemblierung (*assembly*) der resultierenden Sequenzläufe zu Contigs wurde mit dem Programm Gcphrap aus dem Staden Package durchgeführt.
- Die Herstellung eines Datensatzes aus den prozessierten Sequenzläufen sowie die weiterführende Bearbeitung dieses Datensatzes erfolgte mit dem Programm Gap4 aus dem Staden Package.

Nach dem Ende der Rohsequenzierungsphase des Genomprojekts wurden alle Sequenzläufe (*readings*), die sich aus dem Ansequenzieren von Plasmid- und Cosmid-Inserts (*templates*) ergaben, computergestützt assembliert (**Gcphrap**). Readings wurden zu *Contigs* assembliert, kontinuierlichen Sequenzabschnitten, die aus vielen miteinander überlappenden Sequenzbereichen bestehen. Contigs wurden zu *Supercontigs* geordnet, die weitreichende Genombereiche überspannen, aber noch Sequenzlücken zwischen den Contigs enthalten (**Gap4**).

2.11 Lückenschluß

Um Überlappungen zwischen zwei Contigs zu identifizieren, diese Contigs miteinander zu verbinden und damit eine Sequenzlücke zu schließen, wurde entweder die Funktion *Find Internal Joins* aus dem Programm **Gap4** verwendet oder ein Vergleich mittels BLAST-Algorithmus (Altschul *et al.*, 1990) durchgeführt, der sich in der Praxis als empfindlicher und zuverlässiger erwies (siehe 2.12).

2.11.1 Primerwalking

Beim Primerwalking handelt es sich um die einfachste und schnellste Methode zum Lückenschluß direkt nach der Rohsequenzierungsphase. Bei einer durchschnittlichen Insertgröße von 2-3 kbp in den Plasmid-Genbanken der Genomprojekte (siehe 2.8.1) kann mit der Ansequenzierung der Inserts von beiden Seiten nicht die komplette Sequenz des Templates erfaßt werden. Beim sogenannten Primerwalking wird diese Lücke durch Sequenzierung mit geeigneten Primern auf den Templates geschlossen. In der Contig-Darstellung (*Template Display*) von **Gap4** werden zusammengehörende Sequenzläufe eines Templates markiert und als Plasmidbrücken dargestellt, wenn sie mit verschiedenen Contigs assembliert sind (Abb. 3.4). Sequenzlücken, die von sogenannten Plasmidbrücken überspannt werden, können über die Sequenz des Templates erschlossen werden.

2.11.2 Lückenschluß mit PCR

Mit Hilfe von PCR-Amplifikationen wurden die verbliebenen Sequenzlücken geschlossen, nachdem die Möglichkeiten des Lückenschlusses über Primerwalking ausgeschöpft worden waren. Die dabei auftretenden Schwierigkeiten bestanden zum einen darin, die Anordnung der Contigs im Genom herauszufinden, zum anderen darin, die unterschiedlich großen Lücken und teilweise sehr schwer sequenzierbaren Bereiche zu überbrücken.

2.11.2.1 Ableitung über Cosmidbrücken

Ähnlich wie Plasmid-Inserts können Cosmid-Inserts, wenn sie Sequenzlücken überbrücken, Informationen über die Anordnung der Contigs im Genom liefern. Aufgrund der Größe der Cosmid-Templates von 20-50 kbp überspannen sie nicht nur größere Lücken zwischen den Contigs sondern auch Sequenzwiederholungen (*repeats*) innerhalb der Contigs, wie zum Beispiel ribosomale RNA-Operons, die auf Basis der Plasmid-Sequenzläufe schwer aufzulösen sind.

Nachdem die Contigs nach Cosmidbrücken geordnet worden waren, wurden die verbliebenen Lücken je nach erwarteter Größe mit Standard- bzw. *long range* PCR amplifiziert und sequenziert. Teilweise wurden die überbrückenden Cosmide direkt als Template für die PCR genutzt, weil so die Spezifität der Reaktion gegenüber der Amplifikation auf chromosomaler DNA erhöht werden konnte.

2.11.2.2 Ableitung über konservierte Genabfolgen (*peptide links*)

Bei den sogenannten *peptide links* handelt es sich um eine weitere Methode, Informationen über die Anordnung der Contigs im Genom zu bekommen, um Supercontigs bilden zu können. Dabei macht man sich die konservierte Anordnung der Gene zwischen verwandten Organismen zunutze (*Synthenie*). Voraussetzung für die Nutzung von *peptide links* ist damit das Vorliegen größerer annotierter DNA-Bereiche möglichst nah verwandter Organismen. Für den Vergleich der Annotation verschiedener Genome

wurde das ERGO Programmpaket verwendet (Overbeek *et al.*, 2003). Unter der Funktion *contig regions* können mit ERGO die Umgebungen ausgewählter, homologer Gene in verschiedenen Organismen verglichen werden, um konservierte Gen-Cluster zu identifizieren (Abb. 3.11). Zwischen nah verwandten Organismen besteht häufig eine Konservierung der Genabfolge, die sich nicht nur über funktionelle Operons sondern über große Teile des gesamten Genoms erstreckt. Aus der Zuordnung der Gene eines Contig-Endes zu einem konservierten Cluster in einem verwandten Organismus kann dementsprechend auf Gene desselben Clusters geschlossen werden, die am Ende eines benachbarten Contigs zu erwarten sind.

2.11.3 Multiplex-PCR

Nach der Bearbeitung aller Informationen aus Plasmid- und Cosmidbrücken und konservierten Genabfolgen bietet sich für die Verknüpfung der restlichen Contig-Enden die Multiplex-PCR an (Tettelin *et al.*, 1999). Deren Prinzip beruht auf der Bildung von Mischungen unterschiedlicher Primer (*pools*), die in verschiedenen PCR-Reaktionen miteinander kombiniert werden. Dazu wurden an allen Contig-Enden längere Primer abgeleitet (ca. 22 nt) und zu *pools* aus jeweils fünf Primern vereinigt. In mehreren PCR-Runden wurden zunächst Kombinationen verschiedener *pools* identifiziert, mit deren Hilfe erfolgreiche Amplifikationen durchgeführt werden konnten und anschließend einzelne Primer dieser *pools* über Standard-PCRs miteinander kombiniert. Multiplex-PCRs wurden im folgendem Ansatz nach dem PCR-Programm für *TripleMaster*-PCRs (*long range*) durchgeführt:

Multiplex-PCR-Ansatz

Primer- <i>pool</i> 1 (je 1 : 10)	2 μ l
Primer- <i>pool</i> 2 (je 1 : 10)	2 μ l
<i>Tuning Buffer</i>	2 μ l
dNTPs	1 μ l
<i>TripleMaster</i> -Polymerase	0.16 μ l
DNA	120 ng
H ₂ O _{bidest}	ad 20 μ l

2.12 BLAST

Für die meisten Sequenzvergleiche auf DNA- und Proteinebene bietet sich wegen seiner Geschwindigkeit und Empfindlichkeit der BLAST-Algorithmus an (Altschul *et al.*, 1990). In dieser Arbeit wurde der BLAST-Algorithmus eingesetzt, um Sequenzen nach Vektorkontaminationen zu durchsuchen (siehe 2.10), um Überlappungen zwischen Contigs zu finden (siehe 2.11) und um ganze Genome miteinander zu vergleichen (siehe 2.12.1). Für die vergleichende Analyse kompletter Genome wurden bidirektionale BLAST-Vergleiche durchgeführt. Vergleiche auf DNA-Ebene wurden mit dem Programm `blastn` vorgenommen, Vergleiche auf Proteinebene mit `blastp`.

2.12.1 Bidirektionaler BLAST

Für den Vergleich ganzer Genome ist es sinnvoll, das Programm `blastp` zu verwenden, um nur die übersetzten Proteinsequenzen miteinander zu vergleichen und nicht die DNA-Sequenzen der zugrundeliegenden kodierenden Bereiche. Ein Vergleich auf DNA-Ebene würde, bedingt durch das BLAST-Prinzip des *local alignment*, viele kurze Treffer liefern und nicht die in diesem Zusammenhang wesentliche Einheit der kodierenden Sequenzen berücksichtigen. Das besondere Prinzip des bidirektionalen BLASTs zwischen zwei Sequenzen beruht darauf, daß von jeder Seite aus ein BLAST-Vergleich durchgeführt wird und nur diejenigen Proteine als Trefferpaare gewertet werden, die von beiden Seiten aus die größte Ähnlichkeit zueinander zeigen.

Das Vorgehen beim Vergleich ganzer Genome (A und B) unter Unix gliederte sich in folgende Schritte:

- Zusammenfassung aller Proteinsequenzen innerhalb eines *multiple fasta files* (a.fas und b.fas)
- Erstellen zweier BLAST-Datenbanken:

```
formatdb -i a.fas -p T -o T
```

```
formatdb -i b.fas -p T -o T
```

- BLAST-Abgleich der Genome gegeneinander:

```
blastall -p blastp -i a.fas -d b.fas -e 1.0e-15
```

```
blastall -p blastp -i b.fas -d a.fas -e 1.0e-15
```

- skriptgesteuerte Auswertung der Ergebnisse

Mit Hilfe eines Skriptes wurden die Ausgaben der beiden *blastall*-Aufrufe eingelesen, miteinander verglichen und diejenigen Proteine als Treffer ausgegeben, die wechselseitig aufeinanderwiesen, d.h. Proteine 1 und 2 wurden als Trefferpaar ausgegeben, wenn Protein 1 aus Organismus A größte Ähnlichkeit zu Protein 2 aus Organismus B zeigt UND Protein 2 aus Organismus B größte Ähnlichkeit zu Protein 1 aus Organismus A zeigte. Mit diesem Verfahren können falsch positive Treffer zu paralogen Proteinen ausgeschlossen werden (Proteine, die in mehreren Kopien im selben Genom vorliegen). Der Schwellenwert für signifikante BLAST-Treffer wurde als *e-value* ($1.0e-15$, d.h. e^{-15}) im *blastall*-Aufruf angegeben.

2.13 Vorhersage und Annotation kodierender Sequenzen

Um die kodierenden Eigenschaften eines Genoms bestimmen zu können, bedarf es der Annotation aller potentiellen Gene. Dazu müssen zunächst alle kodierenden Sequenzen (*coding sequences*, CDSs) bestimmt werden und diesen in einem zweiten Schritt eine funktionelle Annotation zugewiesen werden. Funktionszuweisungen werden anhand von Sequenzähnlichkeiten zu bekannten Proteinen vorgenommen, beruhend auf der Beobachtung, daß Proteine ab einer Ähnlichkeit von 30 % mit hoher Wahrscheinlichkeit übereinstimmende räumliche Strukturen zeigen (Sander & Schneider, 1991).

2.13.1 Vorhersage kodierender offener Leserahmen (CDSs)

Als zuverlässigste Methode zur CDS-Vorhersage erwies sich eine Kombination verschiedener automatischen Vorhersageprogramme in Verbindung mit einer manuellen Kontrolle der CDS-Längen. Dazu wurde das von Tech & Merkl (2003) entwickelte Programm YACOP verwendet, das auf einer Kombination der Programme Critica

(Badger & Olsen, 1999), Glimmer (Delcher *et al.*, 1999) und ZCurve (Guo *et al.*, 2003) beruht und ab einer Contig-Länge von 50 kbp zuverlässige Ergebnisse liefert. Die Vorhergehensweise gliederte sich in folgende Schritte:

1. Contigs größer 50 kbp wurden einer YACOP-Vorhersage unterzogen.
2. Kleinere Contigs wurden mit Glimmer bearbeitet. Als Trainingsmenge für die Glimmer-Vorhersage wurde der größte verfügbare Contig verwendet.
3. Verbliebene Lücken ab einer Länge von 400 bp wurden mit Einträgen aus der aktuellen Proteindatenbank von NCBI aufgefüllt. Dazu wurde ein BLAST-Vergleich mit dem Programm *tblastx* unter Verwendung eines strikten Grenzwertes von e^{-50} vorgenommen.

Für die Aktualisierung der Vorhersagen nach Veränderung der zugrundeliegenden DNA-Sequenz im Datensatz wurden alle vorhandenen CDSs über BLAST mit den veränderten Sequenzdaten abgeglichen. Ehemalige Sequenzlücken wurden mit CDSs aus einer aktualisierten YACOP-Vorhersage aufgefüllt und nur die Bereiche mit veränderten bzw. hinzugekommenen CDSs einer erneuten Überprüfung unterzogen. In *R. eutropha* H16 wurde neben der YACOP-Vorhersage eine weitere mit dem Programm FrameD durchgeführt (Schiex *et al.*, 2003).

2.13.2 Vorhersage von Transfer-RNAs (tRNAs)

Für die Identifikation der Transfer-RNAs (tRNAs) wurde das Programm *tRNAscanSE* verwendet, das neben den Koordinaten der tRNA das von dieser tRNA abgelesene Codon und die entsprechende Aminosäure ausgibt (Lowe & Eddy, 1997).

2.13.3 Annotation in ERGO

Die automatische Annotation wurde mit dem Programm ERGO durchgeführt (Fa. Integrated Genomics; (Overbeek *et al.*, 2003)). Dazu wurden die von YACOP in Form von START- und STOP-Positionen vorhergesagten Gene in Proteinsequenzen übersetzt und über FASTA (Pearson & Lipman, 1988) mit einer nicht-redundanten, internen Proteindatenbank von ERGO abgeglichen. Diese interne Datenbank setzt sich aus den lokal

annotierten Genomen und den regelmäßig aktualisierten externen Datenbanken *Swiss-Prot* und *TrEMBL* zusammen (<ftp.expasy.org/databases/uniprot/knowledgebase/>). Die automatische Annotation jeder CDS wurde anschließend manuell überprüft und die Funktionszuweisung je nach Übereinstimmung mit Dateneinträgen, Zugehörigkeit zu konservierten Gen-Clustern, etc. angepaßt. Beispielsweise wurde bevorzugt nach Ähnlichkeiten zu Proteinen gesucht, deren Annotation bereits experimentell bestätigt worden war. Übereinstimmungen mit externen Datenbanken wurden stärker gewertet als ERGO-interne Ähnlichkeiten, Treffer in der stark überarbeiteten *Swiss-Prot*-Datenbank stärker gewichtet als in *TrEMBL*. Generell wurde auf eine zurückhaltende Annotation Wert gelegt. Zu jeder CDS wurden über die in die ERGO-Oberfläche eingebundenen Hilfsprogramme verschiedene externe Zusatzinformationen eingeholt:

1. Mit Hilfe von *TMpred*-Vorhersagen (Hofmann & Stoffel, 1993) wurden mögliche Transmembran-Bereiche identifiziert. Danach wurden Proteine als „*putative membrane spanning protein*“ (mehrere Transmembran-Helices) bzw. „*putative membrane associated protein*“ (eine Transmembran-Helix) annotiert.
2. Über *NCBI CD-Search* (Marchler-Bauer *et al.*, 2005) bzw. *Pfam* (Bateman *et al.*, 2004) wurde jede CDS nach bekannten konservierten Domänen durchsucht. Funktionszuweisungen wurden nach Möglichkeit mit Hinweisen auf Domänen belegt.
3. Die Zuordnung eines Gens unbekannter Funktion zu einem konservierten Cluster wurde nach Vergleich der Umgebungen dieses Gens in verschiedenen Organismen mit der ERGO-Funktion *contig regions* (siehe Abb. 3.11) vorgenommen und als Hinweis auf einen möglichen Funktionszusammenhang gewertet.

2.13.4 Kontrolle der CDS-Längen

Die Koordinaten der CDS-Vorhersagen für die einzelnen Contigs wurden für die skriptgesteuerte Berechnung von *Genbank-Files* verwendet. Diese Dateien wurden mit dem Programm *Artemis* (Rutherford *et al.*, 2000) bearbeitet und für die manuelle Korrektur der START- und STOP-Positionen genutzt. Mit der Funktion des *GC Frame Plots* (Ishikawa & Hotta, 1999) liefert *Artemis* ein nützliches Hilfsmittel, um die Grenzen einer CDS zu bestimmen. Dabei werden die G+C-Gehalte für jedes der drei Leseraster innerhalb einer Sequenz unabhängig voneinander berechnet. Da innerhalb kodierender Se-

quenzen starke Unterschiede zwischen den G+C-Gehalten der drei Codon-Positionen bestehen, kann aus dem *GC Frame Plot* die START-Position einer CDS abgeschätzt werden.

2.14 Kumulativer *GC-Skew*

Für die Lokalisation des Replikationsursprunges hat sich die Bestimmung des kumulativen *GC-Skews* nach Grigoriev (1998) als nützlich erwiesen. Dieses Verfahren nutzt eine Ungleichverteilung von G- und C-Basen im *leading* bzw. *lagging strand* aus. Die Ursachen für diese Verteilung liegen vermutlich in unterschiedlichen Selektionsdrücken begründet, die auf den kodierenden und den nicht-kodierenden Strang ausgeübt werden (Lobry, 1996). Der kumulative *GC-Skew* wird in einer Abwandlung des einfachen *GC-Skews* von Lobry (1996) nach folgender Formel berechnet:

$$GC - Skew_{kumulativ} = \frac{G_{i,i+n} - C_{i,i+n}}{G_{i,i+n} + C_{i,i+n}}$$

Dabei gibt (n) die Fenstergröße an, die für die Berechnung der G- und C-Gehalte herangezogen werden ($G_{i,i+n}$ und $C_{i,i+n}$). In einer graphischen Darstellung des kumulativen *GC-Skews* können der Replikationsursprung als globales Minimum und der Terminationspunkt als globales Maximum identifiziert werden.

2.15 Untersuchungen zur *Codon Usage* (SIGI)

Das Programm SIGI liefert basierend auf dem Vergleich von *Codon Usages* und Cluster-Analysen eine Vorhersage für genomische Inseln (Merkl, 2004). Dazu wird aus der Gesamtheit der kodierenden Sequenzen zunächst ein für jeden Organismus spezifischer Durchschnittswert der *Codon Usage* berechnet. Sogenannte Fremdgene (*alien genes*) weichen stark von diesem Durchschnittswert ab und werden zu genomischen Inseln zusammengefaßt, die möglicherweise auf lateralen Gentransfer zurückgeführt werden können. Diejenigen Gene, die in Bezug auf die Translationseffizienz eine perfekte An-

passung der *Codon Usage* an den tRNA-Gehalt des jeweiligen Organismus darstellen, werden von dem Programm SIGI als sogenannte hochexprimierte Gene (*high expression genes*) charakterisiert.

3 Ergebnisse

3.1 Das *Ralstonia eutropha* H16-Genomprojekt

3.1.1 Sequenzierung

3.1.1.1 Strategie

Vor Beginn der Sequenzierung war die Genomorganisation von *R. eutropha* H16 bereits untersucht (Schwartz & Friedrich, 2001) und das kleinste der drei Replikons, das Megaplasmid pHG1, vollständig sequenziert und annotiert worden (Schwartz *et al.*, 2003). Dementsprechend sind alle Angaben, die im folgenden zu pHG1 gemacht werden – sofern nicht anders angegeben –, der Veröffentlichung von Schwartz *et al.* (2003) entnommen. Die Angabe *R. eutropha* bezieht sich, soweit nicht anders vermerkt, auf den Stamm *Ralstonia eutropha* H16.

Eine spezifische Auftrennung der verbliebenen Replikons in ausreichenden Mengen, wie sie für die Herstellung selektiver Genbanken nötig gewesen wäre, war praktisch nicht durchführbar. Deshalb wurde die gesamte chromosomale DNA für die Genomsequenzierung eingesetzt. Eine Sortierung der Replikons erfolgte nach der Assemblierung aufgrund von Sequenzinformationen im Datensatz (siehe 3.1.1.4). Die Sequenzierung wurde mittels zweier Genbanken durchgeführt, welche von der Firma Integrated Genomics (Chicago, IL, USA) hergestellt wurden, einer Plasmid-Genbank mit geringer Insertgröße (2-5 kbp) und hoher Sequenzdichte sowie einer Cosmid-Genbank mit langer Insertgröße (30-50 kbp) und geringer Sequenzdichte.

3.1.1.2 Rohsequenzierungsphase

Während der Rohsequenzierung wurden die Inserts der rekombinanten Genbanken von beiden Enden aus mit Hilfe von aus der Vektorsequenz abgeleiteten Standardprimern ansequenziert. Die resultierenden Sequenzläufe wurden mit dem Programm Gcphrap

zu sogenannten Contigs assembliert, kontinuierlichen Sequenzabschnitten, die aus vielen miteinander überlappenden Sequenzbereichen bestehen (siehe 2.10). Die Anzahl der Sequenzläufe (N), die für eine bestimmte Abdeckung des gesamten Genoms benötigt werden, kann nach folgender Formel abgeschätzt werden:

$$N = \frac{A * L}{w}$$

Dabei entspricht (N) dem Quotienten aus dem Produkt der gewünschten Abdeckung (A) mit der Größe des Genoms (L) und der durchschnittlichen Leseweite einer einzelnen Sequenzierung (w). Beispielsweise wären bei einer durchschnittlichen Leseweite von 600 bp ca. 99 000 Sequenzläufe nötig, um eine 8fache Abdeckung des gesamten Genoms von *R. eutropha* H16 (7.4 Mbp) zu erzielen. In dieser Rechnung wird von einer perfekten statistischen Verteilung der erstellten Genbank über das Genom ausgegangen. Da diese in der Realität nicht zu erreichen ist, wurde die Anzahl der zu sequenzierenden rekombinanten Klone entsprechend höher gewählt.

Für die von der Fa. Integrated Genomics durchgeführte Rohassemblierung wurden die Inserts von 67 000 Klonen einer Plasmidbank und von 2 688 Klonen einer Cosmidbank ansequenziert. Aus den resultierenden Sequenzen wurden 112 012 zu einem Rohdatensatz assembliert und zusammen mit den isolierten Plasmiden und Cosmiden ins Göttingen Genomics Laboratory überstellt.

Von den zunächst assemblierten Sequenzen dieses Rohdatensatzes enthielt ein großer Teil Abschnitte sehr schlechter Qualität, die zu massiven Fehlassemblierungen und Ungenauigkeiten in der Consensus-Sequenz führten. Aus diesem Grund wurden alle Sequenzläufe einer weiteren, strikten Qualitätskontrolle mit dem Programm Phred unterzogen (siehe 2.10). Daraus gingen 101 060 Sequenzen hervor, die zu 5 867 Contigs reassembliert wurden, davon 564 Contigs größer 3 kbp. Obwohl die durchschnittliche Länge der Sequenzen und damit die allgemeine Abdeckung des Datensatzes verringert worden war, konnte die Konsistenz des Datensatzes durch die zweite Assemblierung erheblich verbessert werden.

3.1.1.3 Editierung des Rohdatensatzes

Der aus der Rohassemblierung hervorgegangene Datensatz wurde mit dem Programm Gap4 in der Version 4.6 aus dem Staden Package (Staden, 1996) bearbeitet. Alle weiteren Arbeiten zur Kontrolle und Editierung der Consensus-Sequenz sowie zur Sequenzierung der verbliebenen Lücken wurden mit Hilfe der darin enthaltenen Funktionen *Contig Editor* (Abb. 3.1) und *Template Display* (Abb. 3.3) durchgeführt. Um eine effiziente Bearbeitung des Genomprojektes zu gewährleisten, wurde der Datensatz auf mehrere Unterprojekte aufgeteilt. Verschiedene Arbeitsschritte, die eine manuelle Editierung der gesamten Genomsequenz erforderten, wurden in diesen Unterprojekten von den beteiligten Arbeitsgruppen in Göttingen, Berlin und Münster parallel bearbeitet. Über einen SecureBLAST-Server (Wiezer & Merkl, 2003), der im Göttingen Genomics Laboratory eingerichtet und aktualisiert wurde, erhielten die beteiligten Arbeitsgruppen ständigen Zugang zu den aktuellen Sequenzdaten.

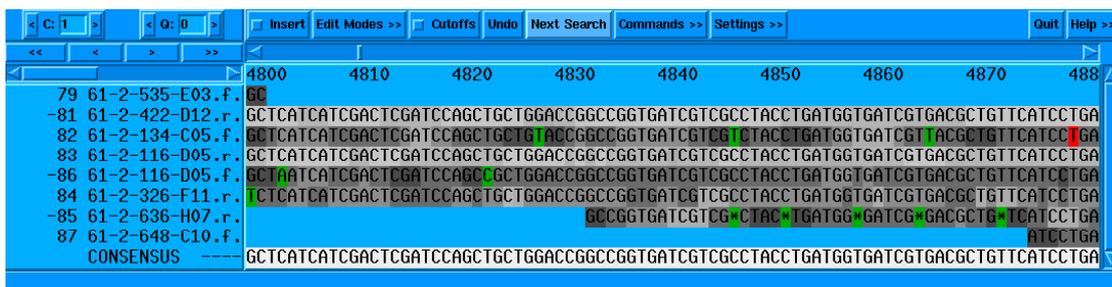


Abb. 3.1: Contig Editor, zufälliger Sequenzausschnitt. Die Fehlerwahrscheinlichkeit der einzelnen Basen (*base confidence*) eines Sequenzlaufes sowie die daraus abgeleitete Zuverlässigkeit der Consensus-Sequenz werden durch verschiedene Grautöne angezeigt. Abweichungen vom Consensus werden in grün hervorgehoben, rot markierte Basen wurden in der Consensus-Sequenz editiert.

3.1.1.3.1 Editierung der Consensus-Sequenz

In einer ersten Editierungsrunde nach der Rohassemblierung wurde die gesamte Consensus-Sequenz aller Contigs größer 3 kbp manuell auf ihre Zuverlässigkeit überprüft. Aufgrund der Größe des Datensatzes wurde diese Editierung auf verschiedene Arbeitsgruppen verteilt und in den Teildatensätzen der Unterprojekte durchgeführt. Für diesen Arbeitsschritt wurde der *Contig Editor* aus Gap4 verwendet (Abb. 3.1), der jeder ein-

zelen Base der Consensus-Sequenz einen Zuverlässigkeitswert (*base confidence*) zuordnet. Dieser Wert errechnet sich aus der Anzahl und Qualität der assemblierten Sequenzläufe, aus denen die Consensus-Sequenz zusammengesetzt wird. Um die einer Sequenz zugrundeliegenden chromatographischen Daten bewerten zu können, wurden zu jeder Sequenz die entsprechenden *Trace Files* aufgerufen (Abb. 3.2).

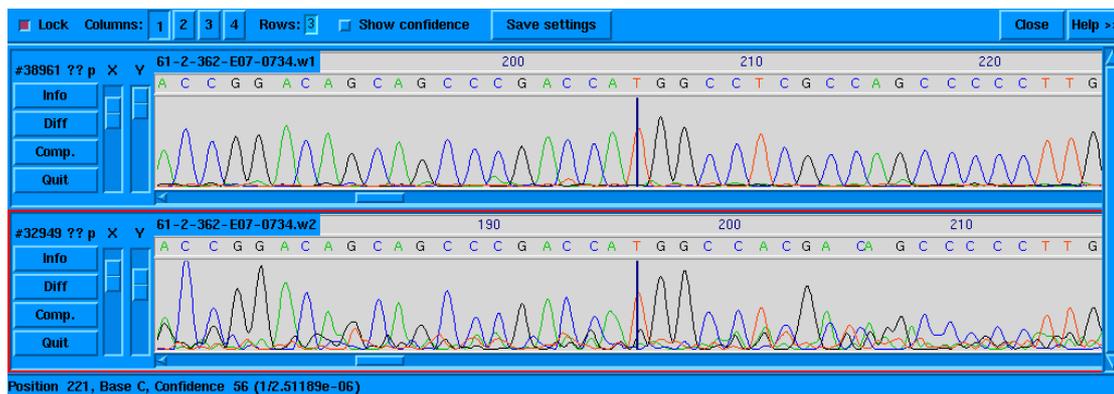


Abb. 3.2: *Trace Files* der Sequenzen von zwei Primerwalkings (61-2-362-E07-0734.w1 und 61-2-362-E07-0734.w2). Chromatographische Daten und Basenzuweisung (nach Phred; siehe 2.10). Einander überlagernde Signale verschiedener Sequenzen erschweren die Basenzuweisung für den zweiten Sequenzlauf.

Als Richtwerte für eine vertrauenswürdige Consensus-Sequenz galt ein Wert von 40 für die *base confidence* und eine Abdeckung von mindestens zwei Sequenzläufen. Im Zweifelsfall wurde anhand der *Trace Files* entschieden. Während der ersten Editierrunde wurden zweifelhafte Bereiche i.d.R. ausgehend von der Plasmid-Genbank sequenziert. Dazu wurden gezielt einzelne Sequenzläufe wiederholt oder durch Primerwalkings mit spezifischen Primern neue Sequenzläufe erstellt (siehe 3.I.1.3.2). Bei Zweifeln an der korrekten Assemblierung der Sequenzläufe eines Contigs oder bei Mangel an überspannenden Plasmid-Inserts wurden PCR-Amplifikationen durchgeführt und PCR-Produkte als Templates für weitere Sequenzierungen genutzt. Der zunehmende Verfall der isolierten Plasmid-DNA resultierte im Verlauf der Editierungsphase zunächst in verkürzten Sequenzläufen und schließlich im Totalausfall der Sequenzierungen. Für die Feinkorrektur der Consensus-Sequenz wurde deshalb vollständig auf PCR-Amplifikationen ausgewichen (siehe 3.I.1.7).

Die Umgebung einer Sequenz, d.h. die Darstellung aller zu einem Contig assemblierten Sequenzläufe, wurde über die Funktion *Template Display* abgerufen (Abb. 3.3). In Bereichen geringer Abdeckung oder schlechter Consensus-Qualität ließ sich auf diese Weise schnell ein Überblick über die Konsistenz des Contigs gewinnen.

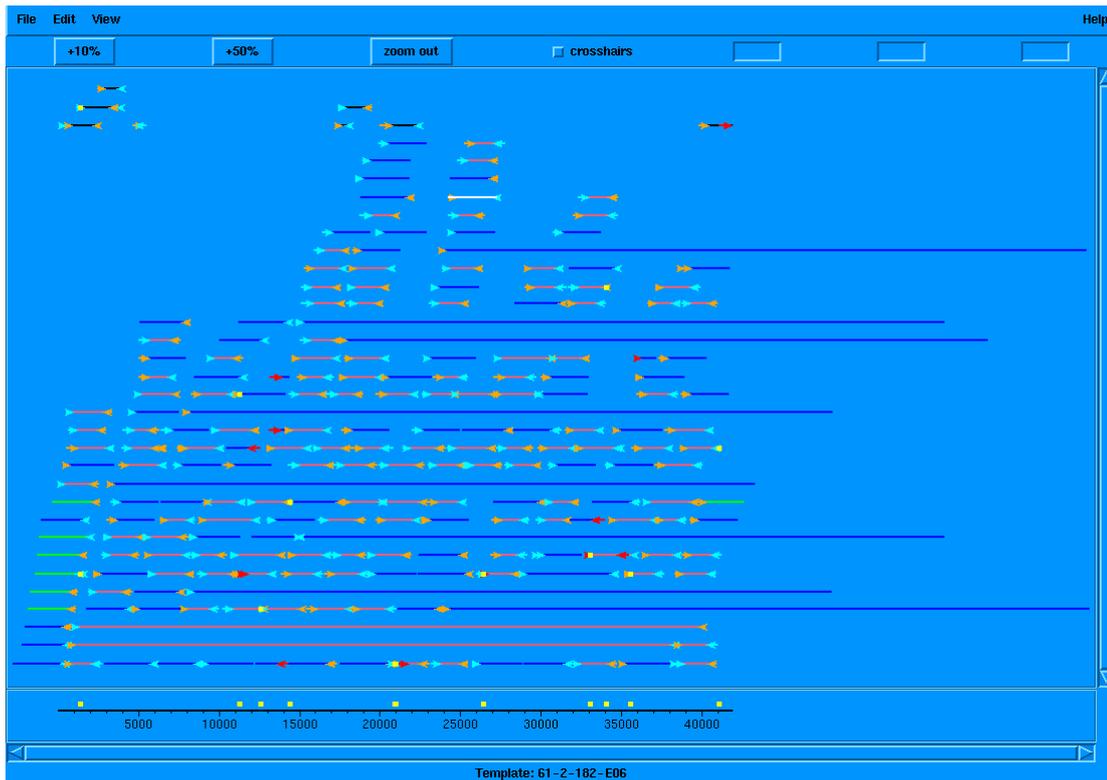


Abb. 3.3: Template Display des Contig 61-2-182-E06. Sequenzläufe werden durch Pfeile dargestellt und einem Template (Plasmid- oder Cosmid-Insert) zugeordnet (blaue, rote, grüne oder schwarze Striche). Hellblaue und orange Pfeile, Sequenzläufe mit Standard-for/rev-Primern; blaue Striche, erwartete Länge des Templates, wenn kein weiterer Sequenzlauf im Datensatz vorhanden ist; rote Striche, Template-Länge zwischen zusammengehörenden Sequenzläufen, die der Erwartung entspricht; schwarze Striche, Template-Länge, die unter der Erwartung liegt; grüne Striche, zusammengehörende Sequenzläufe liegen in unterschiedlichen Contigs (Plasmidbrücke); gelbe Punkte, Primerpositionen.

3.1.1.3.2 Primerwalking über Sequenzlücken

Wie beschrieben gibt das *Template Display* zu jedem Sequenzlauf Informationen über Existenz und Assemblierung des Gegenlaufes vom gleichen Plasmid-Insert an (Abb. 3.3; Abb. 3.4). Sequenzläufe an den Rändern eines Contigs, deren Gegenläufe an den Rändern eines anderen Contigs assemblieren, bilden sogenannte Plasmidbrücken und können beim Primerwalking als Template für die Sequenzierung von Sequenzlücken genutzt werden (siehe 2.11.1). Deshalb wurden parallel zur Editierung der Consensus-Sequenz an den Rändern von jedem Contig Primer abgeleitet und bei Existenz von Plasmidbrücken für das Primerwalking verwendet.

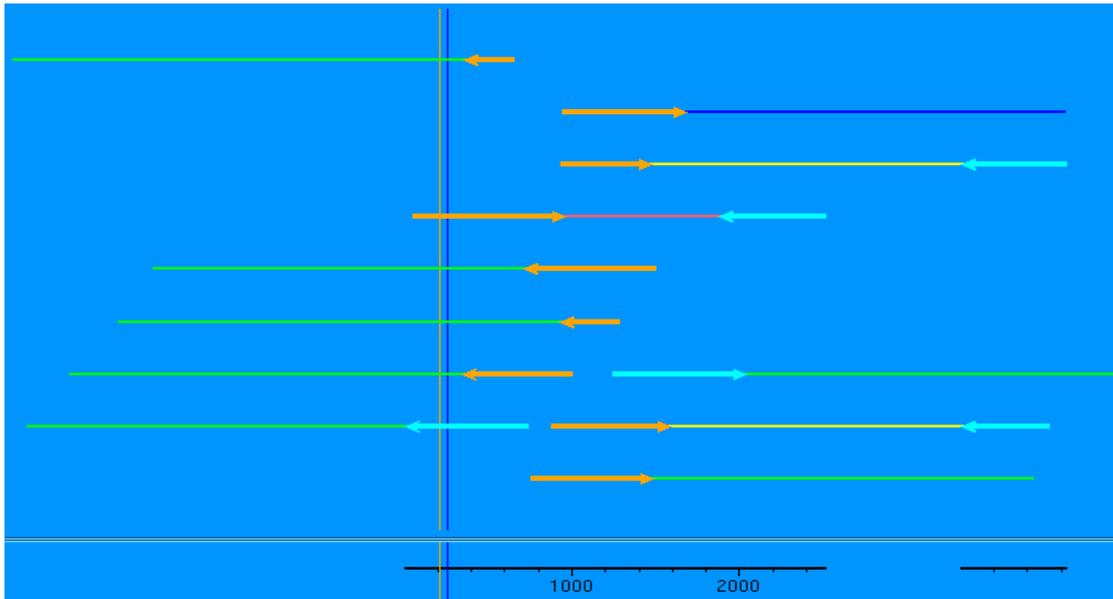


Abb. 3.4: *Template Display* von zwei Contigs, die über Plasmidbrücken (gelbe Templates) miteinander verbunden sind. Farbkodierungen wie in Abb. 3.3.

3.1.1.3.3 Verknüpfung von Contig-Enden im Gesamtdatensatz

Contigs, die auf Fehlassemblierungen überprüft, deren Consensus-Sequenz editiert und an deren Rändern Primer abgeleitet worden waren, wurden in einem neuen Gesamtdatensatz vereinigt. Nach jeder Ergänzung des Gesamtdatensatzes um einen neuen Contig wurden die Enden aller Contigs abgeglichen und gegebenenfalls miteinander verknüpft (siehe 2.11).

3.1.1.3.4 Auflösung von Fehlassemblierungen

Der Rohdatensatz des *R. eutropha*-Genoms enthielt überdurchschnittlich viele Fehlassemblierungen, die entstehen, wenn Sequenzläufe aus unterschiedlichen Bereichen des Genoms aufgrund von Sequenzähnlichkeiten zu sogenannten *Pile-ups* assemblieren (Abb. 3.5). Typische repetitive Regionen, die in einigen wenigen Contigs zu *Pile-ups* assemblieren, sind die ribosomalen RNA-Operons.

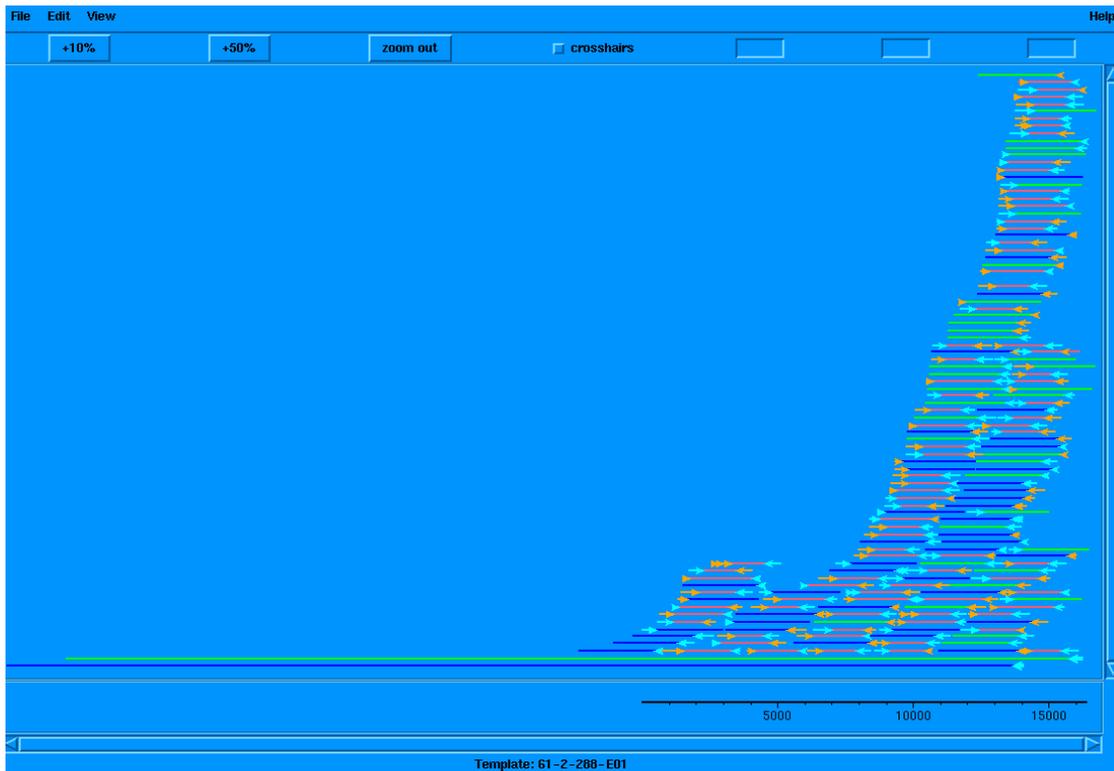


Abb. 3.5: Ribosomaler Contig mit Fehlassemblierungen (Pile up). Über dem gezeigten Abschnitt assemblieren unnatürlich viele Sequenzläufe. Eine hohe Anzahl grün markierter Plasmid- und Cosmid-Templates weist auf Gegenläufe in anderen Contigs hin. Diese Plasmid- und Cosmidbrücken geben Hinweise auf die tatsächliche Zugehörigkeit der fehlassemblierten Sequenzläufe.

Fehlassemblierungen wurden über das *Template Display* als *Pile-ups* oder aufgrund von inkonsistenten Plasmid- und Cosmidbrücken identifiziert und nach Möglichkeit aufgrund von Sequenzunterschieden im *Contig Editor* (Abb. 3.6) sortiert und reassembliert.

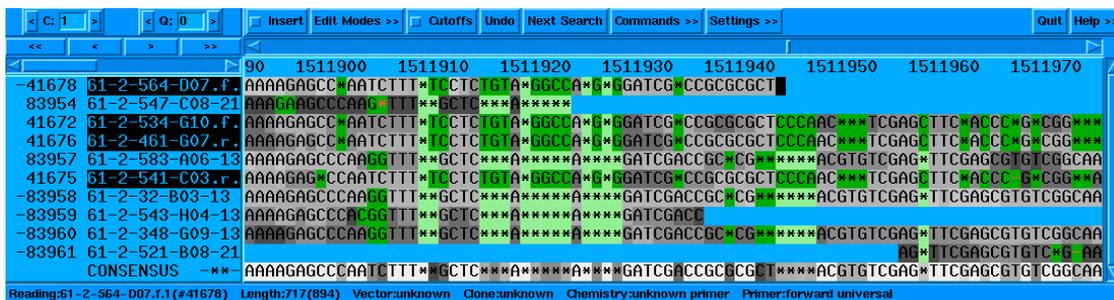


Abb. 3.6: Contig Editor mit Fehlassemblierung. Die assemblierten Sequenzläufe können zwei unterschiedlichen Gruppen aus vier bzw. sechs Sequenzläufen zugeordnet werden.

Die Grenzen der repetitiven Regionen wurden über BLAST-Vergleiche mit dem Gesamtdatensatz oder mit bereits veröffentlichten Genomsequenzen identifiziert (Altschul *et al.*, 1990). Außerhalb der repetitiven Region wurden spezifische Primer abgeleitet,

mit denen der gesamte Sequenzabschnitt über PCR amplifiziert werden konnte. PCR-Produkte dienten als Template für Primerwalking. In einigen Fällen enthielt die amplifizierte DNA interne Wiederholungen, so daß innerhalb des PCR-Produktes keine spezifischen Primer für das Primerwalking abgeleitet werden konnten. In diesen Fällen wurden ausreichende Mengen des repetitiven Bereiches amplifiziert, gesichert und für die Herstellung einer Genbank verwendet. Die Fragmente der Genbank wurden mit den Standardprimern für Plasmide ansequenziert und aus den Sequenzläufen die repetitiven Bereiche zusammengesetzt.

R. eutropha enthält fünf ribosomale RNA-Operons mit einer Größe von ca. 5.1 kbp (siehe 3.1.2.3). Zwischen den RNA-Operons bestehen nur geringe Sequenzunterschiede, so daß in der Rohassemblierung fast alle Sequenzläufe dieser Regionen zu einigen wenigen *Pile-ups* assembliert wurden (Abb. 3.5). Aus diesem Grund war die genaue Anzahl der RNA-Operons zunächst nicht bekannt. Da die Größe der RNA-Operons die durchschnittliche Länge der Inserts der Plasmidbank übersteigt, konnten diese Fehlassemblierungen allein über die Plasmid-Sequenzläufe nicht aufgelöst werden. Um zunächst die Anzahl der rRNA-Operons im Genom zu bestimmen, wurden unspezifische Primer innerhalb der 23S rRNA abgeleitet und rekombinative PCRs mit allen verbliebenen offenen Contig-Enden durchgeführt. In einer zweiten Runde rekombinativer PCRs wurden danach die Enden aller identifizierten rRNA-Cluster miteinander kombiniert. Diese Amplifikationen über die gesamten RNA-Operons hinweg wurden mit Hilfe von *TripleMaster*-PCRs durchgeführt, die speziell für die Amplifikation langer DNA-Fragmente entwickelt wurden (siehe 2.7.1). PCR-Produkte wurden als Templates für schrittweise Primerwalkings durch die gesamten rRNA-Operons genutzt.

3.1.1.4 Strukturierung des Gesamtdatensatzes

Nach der Vereinigung aller editierten Contigs in einem Gesamtdatensatz wurde nach Möglichkeiten gesucht, die Übersichtlichkeit des Datensatzes durch weitergehende Strukturierung zu erhöhen.

1. Zunächst sollten diejenigen Contigs aussortiert werden, deren Sequenz das Megaplasmid pHG1 abbildete. Dazu wurden aus Textdateien künstliche Sequenzläufe hergestellt und daraus im Gesamtdatensatz ein Contig mit der bekannten Sequenz des Megaplasmidas assembliert. Über Find Internal Joins wurden alle Contigs des Datensatzes

mit diesem pHG1-Konstrukt abgeglichen und bei übereinstimmender Sequenz miteinander assembliert.

2. Um die restlichen Contigs des Datensatzes soweit wie möglich den beiden anderen Replikons – Chromosom 1 (Chr.1) und Chromosom 2 (Chr.2) – zuzuordnen, wurde auf Basis der speziellen GenomiPhi-Amplifikation der chromosomalen DNA eine selektive Genbank von Chr.2 hergestellt (siehe 3.I.1.4.1).

3. Schließlich wurde nach Informationen über die Reihenfolge der Contigs im Vergleich zum Genom gesucht, um sogenannte Supercontigs zu bilden, die dann über PCR verknüpft werden sollten (siehe 3.I.1.4.2).

3.I.1.4.1 Herstellung einer selektiven Genbank für Chromosom 2

Über Pulsfeld-Gelelektrophorese (*pulsed-field gel electrophoresis*) wurde die chromosomale Gesamt-DNA von *R. eutropha* aufgetrennt, die drei Replikons wurden aus dem Gel präpariert (A. Pohlmann, Humboldt-Universität, Berlin). Wie bereits erwähnt (siehe 3.I.1.1) konnten auf diesem Wege jedoch keine ausreichenden Mengen an chromosomaler DNA für die Herstellung spezifischer Genbanken isoliert werden. Deshalb wurde zunächst eine unspezifische GenomiPhi-Amplifikation der DNA von Chr.2 durchgeführt (siehe 2.7.2). Die amplifizierte DNA wurde anschließend einer partiellen *Bsp143I*-Restriktion unterzogen (siehe 2.5.1) und für die Herstellung einer Plasmid-Genbank verwendet (siehe 2.8.1). Von dieser Genbank wurden insgesamt 757 Sequenzläufe mit dem Gesamtdatensatz assembliert. Während der GenomiPhi-Amplifikation muß es jedoch zur Bildung artifizieller Konstrukte gekommen sein, die aus benachbarten, scheinbar willkürlich verknüpften DNA-Fragmenten bestanden. Zusätzlich wurden offenbar neben Chr.2 auch geringe Anteile von Chr.1 amplifiziert. Trotzdem konnte mit Hilfe der GenomiPhi-Sequenzläufe etwa die Hälfte der großen Contigs des Gesamtdatensatzes dem kleineren Replikon Chr.2 zugeordnet werden. Darauf aufbauend konnten ca. 95 % der Sequenz von Chr.2 zu Supercontigs zusammengesetzt werden (siehe 3.I.1.4.2.1).

3.1.1.4.2 Ordnung der Contigs relativ zum Genom

3.1.1.4.2.1 Bildung von Supercontigs über Cosmid-Brücken

Analog zur Darstellung der Plasmidbrücken eines Contigs gibt das *Template Display* von Gap4 Auskunft über Cosmidbrücken (Abb. 3.5). Über die Anordnung der Contigs entsprechend ihrer Cosmid- und Plasmidbrücken konnte etwa die Hälfte beider Replikons in sogenannten Supercontigs dargestellt werden. Cosmide eigneten sich nicht für Primerwalkings, konnten allerdings als Template für PCR-Amplifikationen genutzt werden. Über die Größe der Cosmid-Inserts wurden die ungefähren Größen der Sequenzlücken bestimmt. Für genauere Vorhersagen mußten i.d.R. zunächst kleinere Contigs identifiziert werden, die innerhalb der überspannten Bereiche lagen und direkt über PCR-Amplifikationen miteinander verknüpft werden konnten.

3.1.1.4.2.2 Vergleich konservierter Genabfolgen

Parallel zur Editierung der Consensus-Sequenz des Gesamtdatensatzes in Gap4 (siehe 3.1.1.3.1) wurden in regelmäßigen Abständen automatische Vorhersagen der kodierenden Sequenzen (CDSs) mit YACOP (Tech & Merkl, 2003) sowie automatische Annotationen der vorhergesagten CDSs in ERGO (siehe 2.13) durchgeführt. Für den Vergleich der Umgebungen eines Gens in verschiedenen Organismen (*peptide links*; siehe 2.11.2.2) bietet ERGO die Funktion *Contig Regions* an (Abb. 3.11). Aus der Betrachtung konservierter Bereiche eines Contigs in anderen Organismen konnten Hinweise auf die Umgebungen der Gene an den unmittelbaren Contig-Enden in *R. eutropha* abgeleitet werden. Wenn daraus auf angrenzende Contig-Enden geschlossen werden konnte, wurden beide Enden über PCR-Amplifikationen miteinander verknüpft. Besonders nützlich für den Vergleich konservierter Genabfolgen in ERGO erwiesen sich das vollständig entschlüsselte Genom von *R. solanacearum* GMI1000 und die laufenden Genomprojekte von *R. eutropha* JMP134 und *R. metallidurans* CH34, deren Sequenzdaten am DOE Joint Genome Institute, California, USA öffentlich zugänglich sind.

Mit Hilfe der Informationen aus Plasmid- und Cosmidbrücken in Verbindung mit dem Vergleich konservierter Genabfolgen konnte der gesamte Datensatz des

R. eutropha-Genomprojektes auf zwei Supercontigs für Chr.1 und Chr.2 verteilt werden.

3.1.1.5 Lückenschluß

Für die Sequenzierung der verbliebenen Sequenzlücken innerhalb der Supercontigs wurde so weit wie möglich auf Primerwalkings auf Plasmid-Templates zurückgegriffen. Alle weiteren Lücken, für die keine Plasmid-Templates vorlagen, wurden mit Hilfe kombinatorischer PCRs amplifiziert und PCR-Produkte als Templates für Primerwalkings genutzt. Teilweise wurden für die PCR-Amplifikationen Cosmide als Template-DNA genutzt, wenn unspezifische Bindungen der PCR-Primer die Amplifikation auf chromosomaler DNA erschwerten.

3.1.1.6 Sekundärstrukturen

Aufgrund ihres hohen G+C-Gehalts (66.3 %) können einzelsträngige Sequenzabschnitte des *R. eutropha*-Genoms verhältnismäßig stabile Sekundärstrukturen ausbilden. Durch diese Strukturen kann die DNA-Polymerase blockiert werden, wodurch es zu Abbrüchen der Amplifikationsreaktionen von PCR und Sequenzierung kommt (Abb. 3.7). Deshalb wurden für die Amplifikation problematischer Bereiche stringenter Denaturierungsbedingungen gewählt, mit deren Hilfe auch die letzten Lücken sequenziert werden konnten. Durch die Zugabe von DMSO zum Reaktionsansatz sowie durch Verlängerung des Denaturierungsschrittes und Erhöhung der Denaturierungstemperatur konnten potentielle Sekundärstrukturen soweit geschwächt werden, daß eine Sequenzierung möglich war. In einigen Fällen erwies sich die Verwendung von Primern von Nutzen, die möglichst nah am Beginn der Sekundärstruktur abgeleitet worden waren.

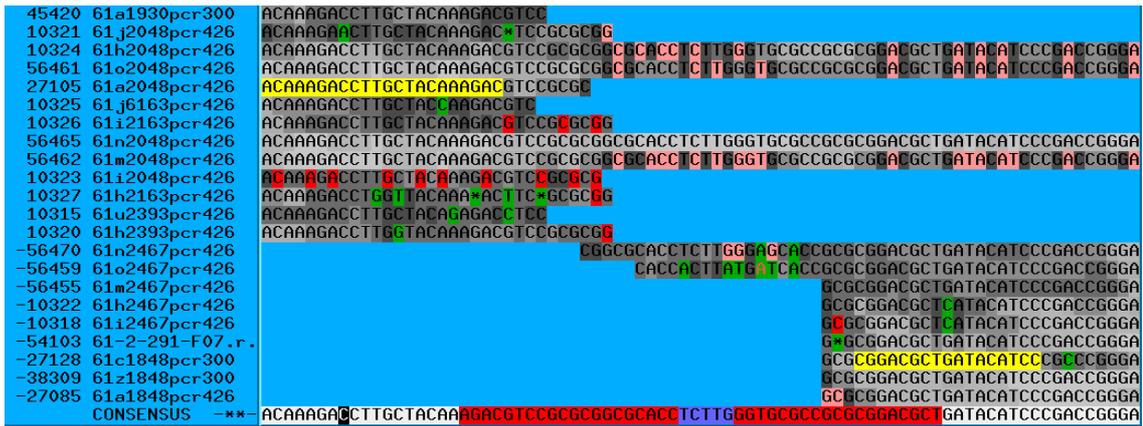


Abb. 3.7: Abgebrochene Sequenzläufe aufgrund der Bildung einer Sekundärstruktur. In einer Darstellung der Sekundärstruktur als *hairpin* würde die palindromische Sequenz (rot) den Stamm der *hairpin*-Struktur, die blauen Basen deren *loop* darstellen. Weitere Markierungen: Gelb, Primersequenzen; grün, Basenabweichungen innerhalb der Sequenzläufe; rot, korrigierte Basen innerhalb der Sequenzläufe; rosa, manuell zugewiesen Basen innerhalb der Sequenzläufe.

3.I.1.7 Feinkorrektur des Gesamtdatensatzes

Vor dem endgültigen Abschluß der Sequenzierung von *R. eutropha* H16 werden noch weitere, sehr zeitaufwendige Feinkorrekturen an der Consensus-Sequenz von Chr.1 benötigt. Deshalb wurde vereinbart, die Sequenzanalyse von Chr.2 sowie allgemeine vergleichende Analysen der beiden Replikons untereinander und im Vergleich zu anderen Spezies der Familie der *Burkholderiaceae* in den Mittelpunkt dieser Arbeit zu stellen.

3.1.2 Sequenzanalyse

Das Genom von *Ralstonia eutropha* H16 besitzt nach jetzigem Stand eine Gesamtgröße von 7 417 135 bp, die sich auf Chr.1 mit 4 052 527 bp (Abb. 3.8), Chr.2 mit 2 912 452 bp (Abb. 3.9) und das Megaplasmid pHG1 mit 452 156 bp verteilt. In Tab. 3.1 sind einige allgemeine Eigenschaften des Genoms von *R. eutropha* H16 im Vergleich zu *R. solanacearum* GMI1000, dem nächsten sequenzierten Verwandten (Salanoubat *et al.*, 2002), dargestellt. Alle Angaben, die im folgenden über das Genom von *R. solanacearum* gemacht werden, beziehen sich auf den Stamm *R. solanacearum* GMI1000 und sind der Veröffentlichung von Salanoubat *et al.* (2002) entnommen.

Das Genom von *R. solanacearum* setzt sich aus einem Chromosom (3.716 Mbp) und einem zweiten, als Megaplasmid bezeichneten Replikon (2.095 Mbp), zusammen. Ein pHG1 entsprechendes Megaplasmid fehlt in *R. solanacearum*.

Tab. 3.1: Vergleich ausgewählter Eigenschaften der Genome von *R. eutropha* H16 und *R. solanacearum* GMI1000 (Salanoubat *et al.*, 2002).

Eigenschaft	<i>R. eutropha</i> H16				<i>R. solanacearum</i> GMI1000		
	gesamt	Chr.1	Chr.2	pHG1	gesamt	Chr.	Mpl.
Länge [Mbp]	7,417	4,053	2,912	0,452	5,811	3,716	2,095
G+C-Gehalt [%]	66,3	66,5	66,3	62,3	67,0	67,0	66,9
kodierende Sequenz	82,9	87,8	88,8	80,5	87,3	87,8	86,5
tRNAs	59	51	7	1	58	55	3
rRNA-Operons	5	3	2	-	4	3	1
CDSs	6956	3882	2654	420	5129	3448	1681

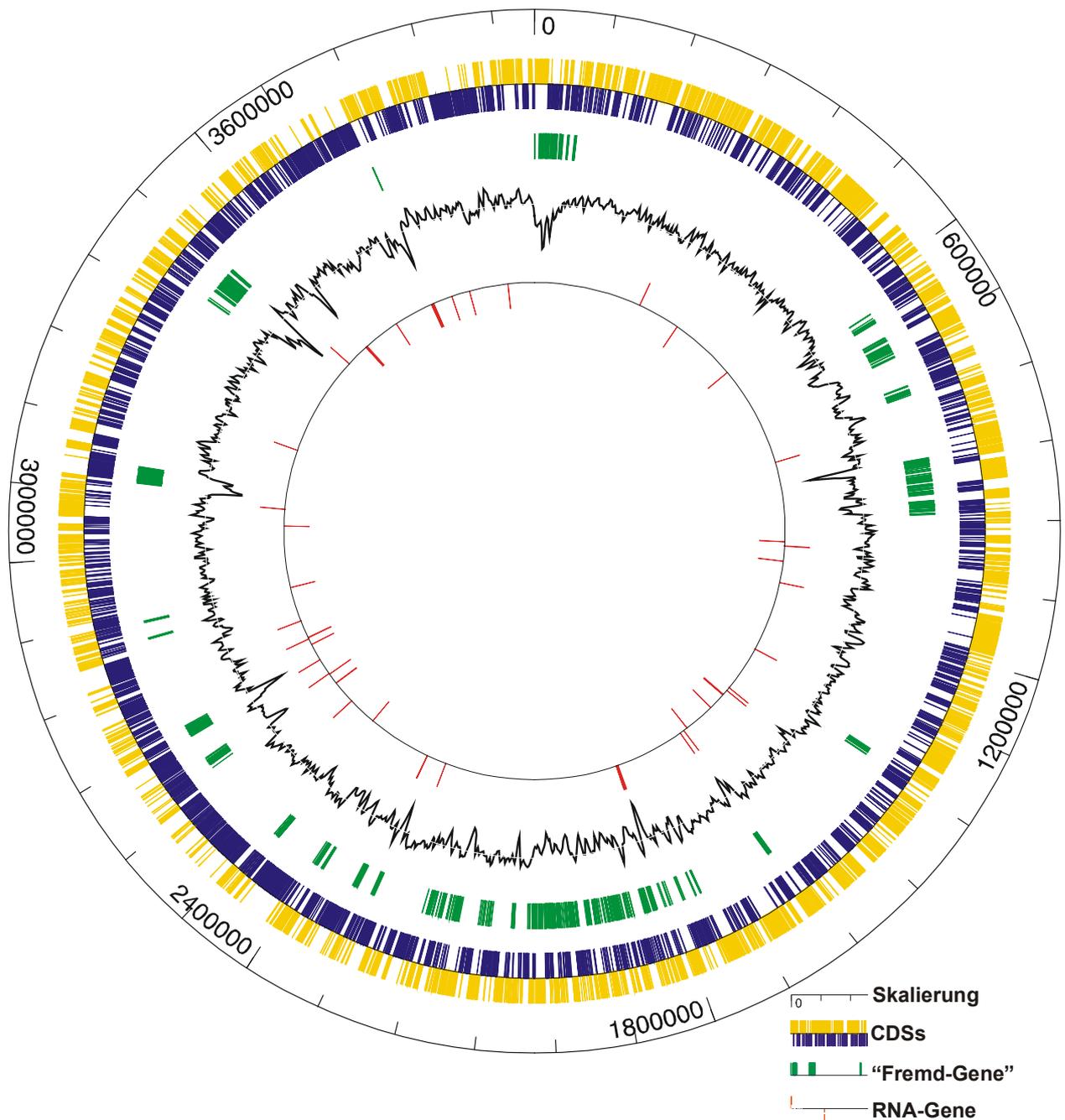


Abb. 3.8: Schematische Darstellung von Chr.1. Von innen nach außen werden folgende Eigenschaften des Replikons dargestellt: 1. Rote Balken, RNA-Gene, d.h. rRNA-Operons und tRNA-Gene. Von der Linie aus nach außen weisende Balken stehen für Gene, die in Richtung des Uhrzeigersinns kodiert sind. 2. Schwarze Linie, G+C-Gehalt. Bereiche niedrigen G+C-Gehaltes weichen von der Durchschnittslinie nach Innen ab. 3. Grüne Balken, potentielle Fremdgene nach SIGI-Vorhersage in strangunabhängiger Darstellung. 4. Blaue Balken, CDSs in Richtung gegen den Uhrzeigersinn; gelbe Balken, CDSs in Richtung des Uhrzeigersinns. 5. Skalierung ausgehend vom Replikationsursprung in Basenpaaren.

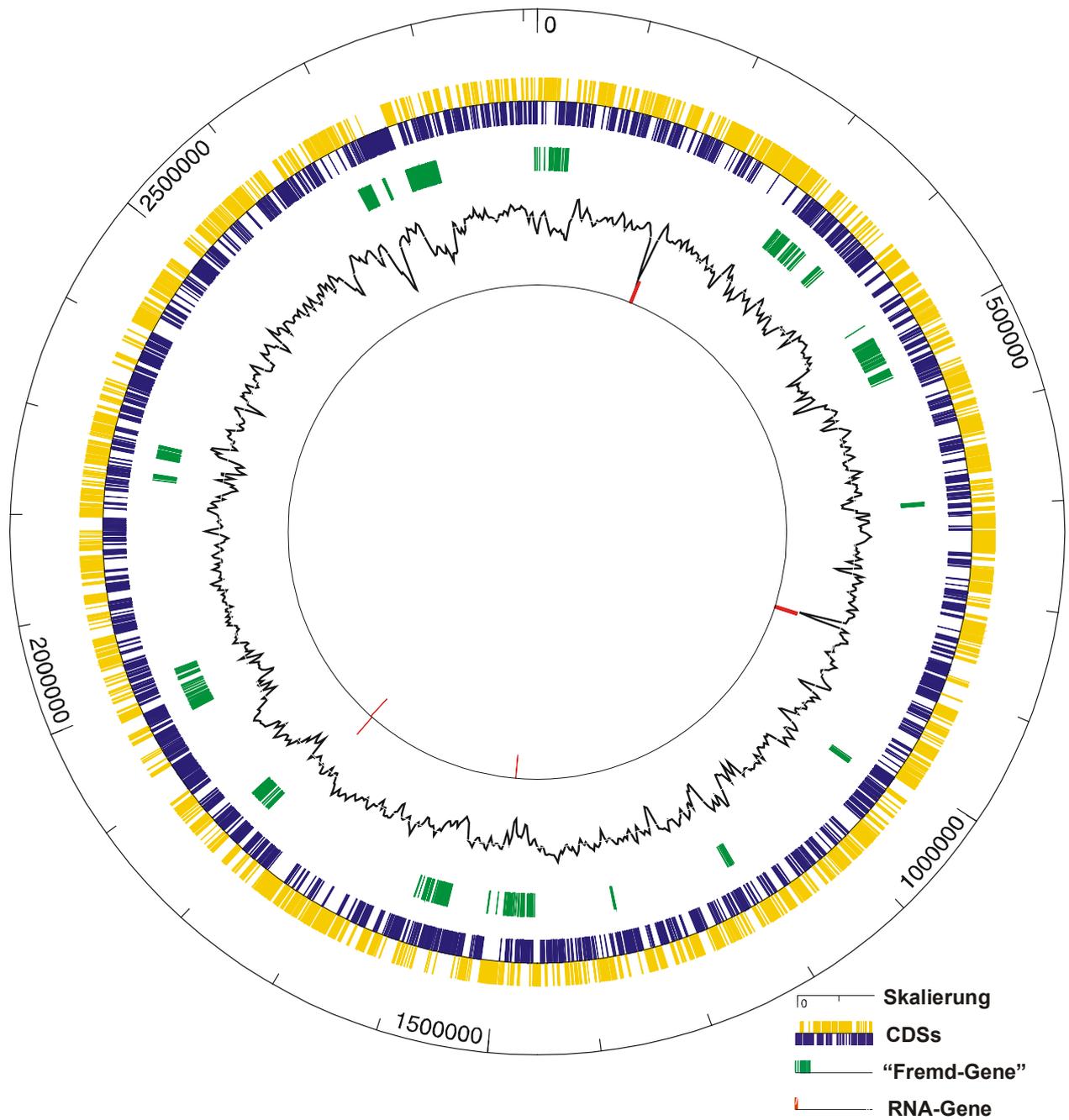


Abb. 3.9: Schematische Darstellung von Chr.2. Reihenfolge der dargestellten Eigenschaften entsprechend der Darstellung von Chr.1 (Abb. 3.8). Von innen nach außen: 1. Rote Balken, rRNA-Operons und tRNAs, strangspezifisch. 2. G+C-Gehalt. 3. Fremdgene, strangunspezifisch. 4. CDSs, strangspezifisch. 6. Skalierung in Basenpaaren.

3.1.2.1 G+C-Gehalt

Mit 66.3 % besitzt *R. eutropha* einen verhältnismäßig hohen G+C-Gehalt, der sich zwischen Chr.1 (66.5 %) und Chr.2 (66.3 %) kaum unterscheidet. Das Megaplasmid pHG1 weist demgegenüber einen niedrigeren G+C-Gehalt auf (62.3 %). In *R. solanacearum* (67.0 %) unterscheiden sich die G+C-Gehalte von Chromosom (67.0 %) und Megaplasmid (66.9 %) ebenfalls nicht wesentlich. In beiden Genomen läßt sich eine leichte Abnahme des G+C-Gehaltes vom größeren zum kleineren Replikon feststellen, in *R. eutropha* zusätzlich eine starke Abnahme von den beiden großen Replikons zum Megaplasmid pHG1. Innerhalb der Replikons von *R. eutropha* kommt es zu starken Schwankungen des G+C-Gehaltes (Abb. 3.8; Abb. 3.9), die mit zwei besonderen Eigenschaften des Genoms korrelieren:

1. Die ribosomalen RNA-Operons fallen im G+C-Gehalt gegenüber dem Rest des Genoms deutlich ab (ca. 53 %).
2. Genomische Inseln, die auf Basis von Unterschieden in der *Codon Usage* bestimmt wurden (siehe 3.1.3.2) und möglicherweise durch lateralen Gentransfer übertragen wurden, weisen ohne Ausnahme einen niedrigeren G+C-Gehalt auf als das restliche Genom.

3.1.2.2 Replikationsursprünge

Die Replikationsursprünge der beiden großen Replikons wurden in *R. eutropha* durch eine Kombination genomweiter Analysen der Sequenz, der Annotationen und des kumulativen *GC-Skews* bestimmt. *GC-Skew*-Analysen machen sich die ausgeprägte Asymmetrie mancher Organismen in Bezug auf die Nukleotid-Zusammensetzung der beiden DNA-Stränge zunutze (Grigoriev, 1998). In der Darstellung des kumulativen *GC-Skews* über das gesamte Replikon stellt der Replikationsursprung das globale Minimum dar (siehe 2.14). Die einzelnen Replikons von *R. eutropha* eignen sich sehr gut für die Bestimmung des Replikationsursprunges mit Hilfe kumulativer *GC-Skew*-Analysen (Beispiel: Abb. 3.10).

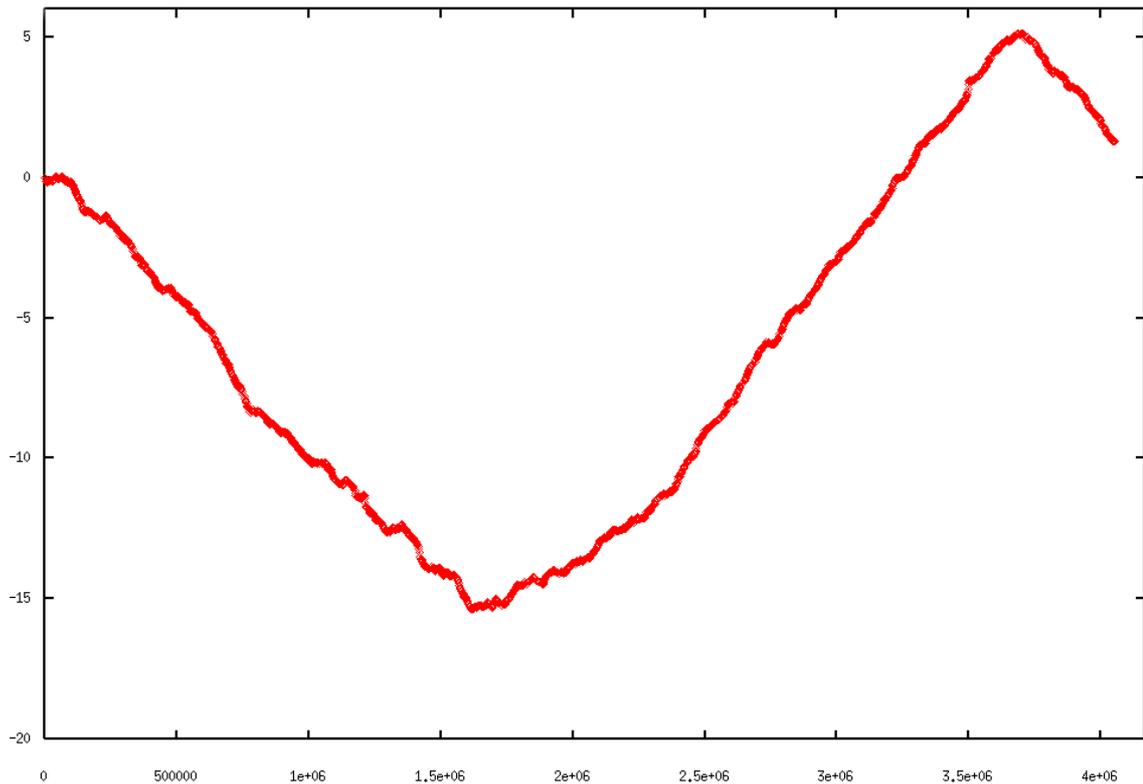


Abb. 3.10: Kumulative GC-Skew-Analyse von Chr.1 aus *R. eutropha* nach Grigoriev (1998).

Ähnlich wie *R. solanacearum* besitzt *R. eutropha* zwei unterschiedliche Typen der *Origins of replication*, einen chromosomalen *oriC* auf Chr.1 sowie zwei plasmidäre *oriV* auf Chr.2 und pHG1:

Der chromosomale Replikationsursprung fällt in Bakterien i.d.R. in eine intergenische Region in direkter Nachbarschaft einer CDS für das DnaA-Protein, das mit der Initiation der Replikation in Verbindung gebracht wird (Messer, 2002). Die Replikation selbst wird durch Bindung von DnaA an repetitive, nicht palindromische, nonamere Sequenzen in der *oriC*-Region initialisiert. Auf Chr.1 ist am globalen Minimum des kumulativen *GC-Skews* ein Gen (H16_A0001) mit 88 % Identität zu *dnaA* aus *R. solanacearum* lokalisiert. In unmittelbarer Nachbarschaft stromaufwärts von *dnaA* befinden sich außerdem zwei potentielle DnaA-Bindestellen (TTATCCACA), die Ähnlichkeiten zu der Modell-Bindestelle aus *E. coli* aufweisen (Fuller *et al.*, 1984). In der Umgebung von *dnaA* sind Proteine der genetischen Informationsprozessierung kodiert: Stromaufwärts werden die DnaA-Bindestellen von einem Protein der kleinen ribosomalen Untereinheit (L34P, H16_A3747) und einer Untereinheit der tRNA-prozessierenden Ribonuklease P (H16_A3746) flankiert. Stromabwärts von *dnaA* sind eine β -Untereinheit der DNA-Polymerase III (EC 2.7.7.7, H16_A0002), eine β -Untereinheit

der DNA-Gyrase B (EC 5.99.1.3; H16_A0003) und drei Untereinheiten eines Typ I Restriktions-Modifikationssystems (H16_A0004-0006) kodiert. Während die Sequenz des *oriC* zwischen unterschiedlichen Spezies variiert, sind Zusammensetzung und Reihenfolge der Gene in der Umgebung des chromosomalen Replikationsursprunges auch zwischen entfernteren Organismen weitgehend konserviert (Abb. 3.11).

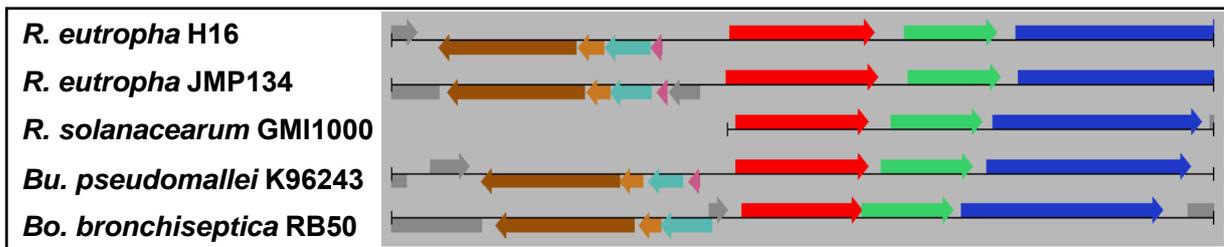


Abb. 3.11: Vergleich der Umgebungen ausgewählter *dnaA*-Homologe (rot) in verschiedenen Organismen (Contig Regions). Gene mit ähnlicher Sequenz werden in gleicher Farbe dargestellt. Rot, *dnaA*; grün, β -Untereinheit der DNA-Polymerase III; blau, β -Untereinheit der DNA-Gyrase B; rosa, L34P; türkis, Ribonuklease P. Abkürzungen: *R.*, *Ralstonia*; *Bu.*, *Burkholderia*; *Bo.*, *Bordetella*.

Im Gegensatz zur chromosomalen Replikation wird mit der Initiation der plasmidären Replikation die Anzahl der Kopien des Plasmids kontrolliert (del Solar & Espinosa, 2000). An diesem Mechanismus ist das plasmidäre Replikationsprotein *repA* beteiligt (Burian *et al.*, 2003). Ebenso wie pHG1 enthält Chr.2 ein potentielles *repA*-Gen (H16_B0001), das mit einem Minimum des kumulativen *GC-Skews* zusammenfällt (Daten nicht gezeigt). Die *RepA*-Proteine von Chr.2 aus *R. eutropha* und dem Megaplasmid aus *R. solanacearum* weisen große Ähnlichkeit zueinander auf (92 % Identität auf Proteinebene). Dagegen unterscheiden sich beide Varianten stark vom *repA*-Gen auf pHG1, das mehr dem *repA* vom Megaplasmid pMOL28 aus *R. metallidurans* CH34 (Taghavi *et al.*, 1996) ähnelt (Abb. 4.2). In der intergenischen Region stromaufwärts von *repA* befinden sich – in unterschiedlicher Orientierung – acht potentielle *RepA*-Bindestellen der Consensus-Sequenz CGTACCCGTTTCTGCC. In gleicher Reihenfolge wie auf dem Megaplasmid von *R. solanacearum* folgen auf *repA* stromabwärts ein potentielles *par*-Operon (H16_B0003-0004) und ein Gen der Transposase/Integrase-Familie (H16_B0005). Ein *par*-Operon ist auch in unmittelbarer Nachbarschaft von *repA* auf pHG1 lokalisiert. *parA* und *parB* kodieren für ein plasmidäres Partitionierungssystem, das während der Zellteilung für eine korrekte Verteilung der Plasmide auf die Tochterzellen sorgt (Gerdes *et al.*, 2000).

3.1.2.3 RNA-Gene

Mit dem Programm tRNAscanSE wurden im Genom von *R. eutropha* insgesamt 59 tRNAs identifiziert (siehe 2.13.2), 51 tRNAs auf Chr.1, sieben tRNAs auf Chr.2 und eine tRNA auf pHG1. *R. solanacearum* dagegen enthält 58 tRNAs, davon 55 tRNAs auf dem Chromosom und drei tRNAs auf dem Megaplasmid (Tab. 3.1). *R. eutropha* kodiert auf Chr.1 für einen vollständigen Satz tRNAs für alle möglichen Codons. Mit Ausnahme eines ribosomalen Operons auf Chr.1 sind in den intergenischen Bereichen aller rRNA-Operons zwischen der 16S rRNA und der 23S rRNA zwei tRNAs kodiert (Abb. 3.12). Neben diesen vier tRNAs aus zwei rRNA-Operons enthält Chr.2 drei weitere einzelne tRNAs (Leu-tRNA, Arg-tRNA und Met-tRNA). Die einzige auf pHG1 kodierte tRNA (Met-tRNA) weist auf DNA-Sequenzebene keine Ähnlichkeit zu einer der anderen tRNAs auf. Überraschenderweise kodiert *R. eutropha* auf Chr.2 für mehrere potentielle Proteine aus dem Selenocystein-Metabolismus: ein Selenium-Bindeprotein (H16_B1000), eine Selenophosphat-Synthase (EC 2.7.9.3, H16_B0295) und einen Selenocystein-spezifischen Elongationsfaktor (SelB, H16_B0947). Allerdings kann im gesamten Genom weder eine Selenocysteinyl-tRNA noch eine Selenocysteinyl-tRNA-Synthetase identifiziert werden.

Durch BLAST-Vergleich von *R. eutropha* mit *R. solanacearum* wurden die stark konservierten Bereiche der ribosomalen RNA-Gene (rRNAs) identifiziert. Das Genom von *R. eutropha* enthält fünf rRNA-Operons, zwei davon auf Chr.2; *R. solanacearum* dagegen enthält insgesamt vier rRNA-Operons, davon eines auf dem Megaplasmid. Innerhalb der rRNA-Operons im Genom von *R. eutropha* lassen sich leichte Unterschiede feststellen, die allerdings nicht an der Lokalisation auf Chr.1 oder Chr.2 festgemacht werden können: Drei der fünf Operons – zwei auf Chr.1 und eines auf Chr.2 – setzen sich aus identischen rRNA-Genen zusammen. Die beiden restlichen Operons weisen veränderte 16S bzw. 23S rRNA-Gene auf (Abb. 3.12). Operon 5 auf Chr.2 unterscheidet sich in zwei Basenaustauschen innerhalb der 16S rRNA von allen anderen 16S rRNA-Genen. Operon 1 auf Chr.1 enthält nur den ersten dieser beiden Basenaustausche und unterscheidet sich außerdem durch das Fehlen der beiden tRNAs in den intergenischen Bereichen sowie durch zwei Insertionen innerhalb der 23S rRNA deutlich von allen anderen rRNA-Operons. Das Vorkommen solcher Insertionen, sogenannter *intervening sequences* (IVs), von unterschiedlicher Länge und Sequenz innerhalb der 23S rRNA

wurde bereits für eine Reihe eukaryotischer und prokaryotischer Organismen beschrieben (Ralph & McClelland, 1993; Miller *et al.*, 2000), unter anderem auch für das α -Proteobakterium *Rhizobium leguminosarum* (Selenska-Pobell *et al.*, 1996). *R. etropha* ist der erste Organismus aus der Gruppe der β -Proteobakterien, für den IVs beschrieben werden können. Interessanterweise entspricht die Lokalisation der IVs innerhalb des 23S rRNA-Gens dem für verschiedene *Salmonella*- (Burgin *et al.*, 1990) und *Haemophilus*-Spezies (Song *et al.*, 1999) beschriebenen Muster, nicht jedoch der in *Rhizobium* charakterisierten Verteilung (Evguenieva-Hackenberg & Selenska-Pobell, 1995).

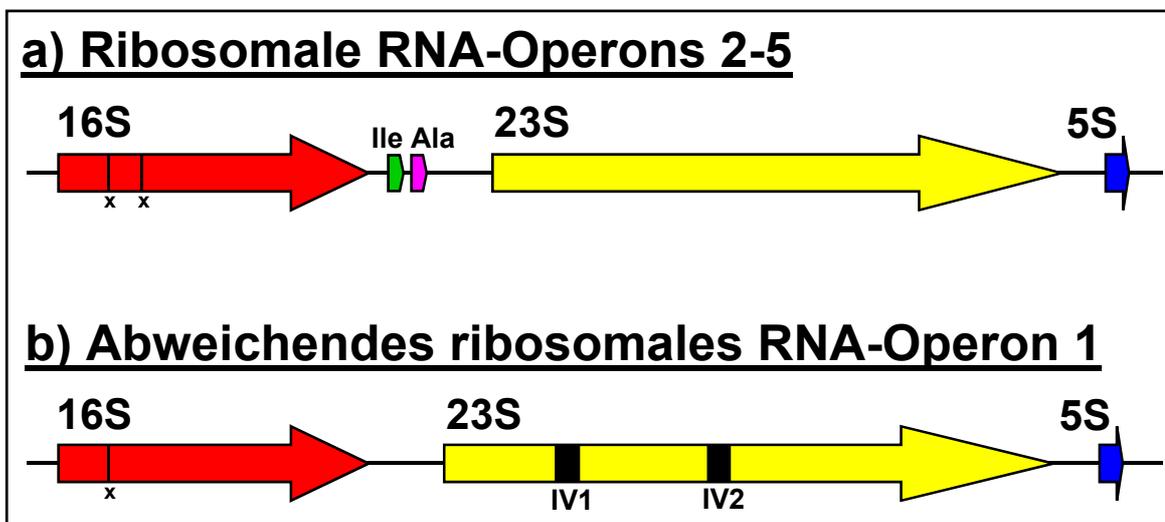


Abb. 3.12: Schematische Darstellung der ribosomalen RNA-Operons in *R. etropha*. a) Die rRNA-Operons 2 und 3 auf Chr.1 sowie 4 und 5 auf Chr.2 unterscheiden sich innerhalb der rRNA-Gene nur durch zwei Basenaustausche (x) innerhalb der 16S rRNA. b) Das rRNA-Operon 1 auf Chr.1 enthält zwei *intervening sequences* (IV1 und IV2) innerhalb der 23S rRNA. In diesem Operon fehlen die zwei tRNA-Gene (Ile, Ala) im intergenischen Bereich zwischen 16S und 23S rRNA.

3.1.2.4 Vorhersage kodierender Sequenzen

Die Vorhersage kodierender offener Leserahmen (siehe 2.13.1) wurde in *R. etropha* durch zwei wesentliche Eigenschaften des Genoms behindert:

1. Der hohe G+C-Gehalt senkt die statistische Wahrscheinlichkeit für das Auftreten der A/T-reichen STOP-Codons TAA, TGA und TAG. Dadurch steigt nicht nur die Gesamtzahl aller möglichen offenen Leserahmen (ORFs), sondern

gleichzeitig wird die Vorhersage der START-Codons und damit die Längenbestimmung aller kodierenden Sequenzen erschwert.

2. Als Folge regelmäßiger intrachromosomaler Rekombinationsereignisse (siehe 3.I.3.1) und der Integration mobiler genetischer Elemente durch lateralen Gentransfer (siehe 3.I.3.2) weist das Genom von *R. eutropha* eine Mosaikstruktur auf. Diese Plastizität der *R. eutropha*-Genoms kann für Schwankungen des G+C-Gehaltes (siehe 3.I.2.1), der *Codon Usage* und anderer Eigenschaften der Sequenz verantwortlich gemacht werden. Dadurch wird die Definierung statistischer Parameter erschwert, mit deren Hilfe automatische CDS-Vorhersagen erstellt werden.

Nach Abschluß der ersten Editierungrunde wurde eine CDS-Vorhersage durchgeführt (siehe 2.13.1) und parallel zur fortlaufenden Bearbeitung der Gesamtdatensatzes manuell korrigiert (siehe 2.13.4) und annotiert (siehe 2.13.3). In regelmäßigen Abständen wurde diese CDS-Vorhersage aktualisiert und an Veränderungen der zugrundeliegenden Sequenzdaten angepaßt. Die standardmäßige CDS-Vorhersage mit YACOP wurde für *R. eutropha* um eine weitere Vorhersage mit Framed ergänzt (Schiex *et al.*, 2003). Dieses Programm wurde im Zusammenhang mit dem *R. solanacearum*-Genomprojekt entwickelt und lieferte dementsprechend gute Ergebnisse für das *R. eutropha*-Genom. Die manuelle Kontrolle der Vorhersagen wurde anhand von *Genbank-Files* mit dem Programm Artemis (Rutherford *et al.*, 2000) durchgeführt. Während der manuellen Kontrolle der CDS-Vorhersagen mit Artemis (siehe 2.13.4) wurden die Ergebnisse von YACOP und Framed verglichen und gegeneinander abgewogen. Im Zuge der letzten Vorhersage, die auf den aktuellen Sequenzdaten beruht (März 2005) und die Grundlage der bestehenden Annotation bildet, wurden insgesamt 6956 CDSs identifiziert, von denen 2654 (ca. 38 %) auf Chr.2 liegen.

3.I.2.5 Annotation

Parallel zur letzten Editierungrunde von Chr.1, nach weitgehend abgeschlossener Feinkorrektur der Sequenz von Chr.2, wurde von den am Genomprojekt beteiligten Arbeitsgruppen eine erste sorgfältige Annotation des gesamten Genoms in ERGO durchge-

führt (siehe 2.13.3). Diese Annotation bildet die Grundlage der in dieser Arbeit durchgeführten Sequenzanalysen.

3.1.3 Mosaikstruktur des *R. eutropha*-Genoms

Die beobachteten Schwankungen im G+C-Gehalt (siehe 3.1.2.1) sowie Probleme bei der CDS-Vorhersage (siehe 3.1.2.4) belegen eine Mosaikstruktur des *R. eutropha*-Genoms, das darin Übereinstimmungen zu *R. solanacearum* und anderen Spezies der Ordnung *Burkholderiales* aufweist (Parkhill *et al.*, 2003; Nierman *et al.*, 2004). Im folgenden sollen zwei weitere Untersuchungen beschrieben werden, die Hinweise auf eine starke strukturelle Plastizität des Genoms von *R. eutropha* geben.

3.1.3.1 Vergleich von *R. eutropha* und *R. solanacearum*

Um zunächst einen Überblick über Ähnlichkeiten zwischen den Genomen von *R. eutropha* und *R. solanacearum* zu gewinnen, wurden direkte BLAST-Vergleiche der Replikons auf DNA-Ebene durchgeführt (siehe 2.12).

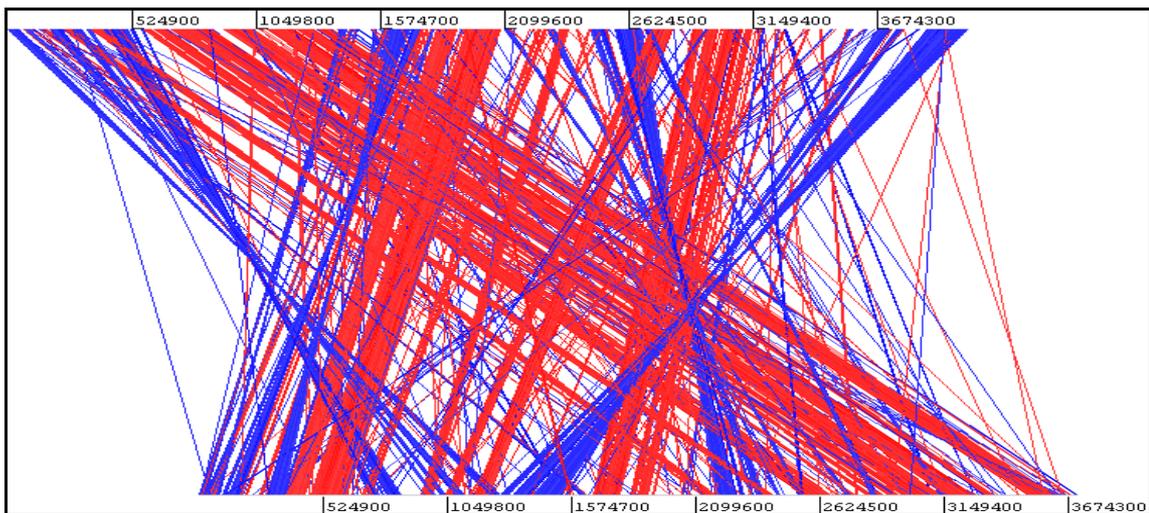


Abb. 3.13: BLAST-Vergleich auf DNA-Ebene zwischen Chr.1 aus *R. eutropha* und dem Chromosom aus *R. solanacearum*. Ausschnitt einer Darstellung in ACT (Rutherford *et al.*, 2000). Treffer ab einem *e-Value* von e^{-15} von *R. eutropha*, Chr.1 (oben) gegen *R. solanacearum*, Chr. (unten) werden als rote und blaue Balken angezeigt. Rote Balken, Treffer in gleicher Orientierung; blaue Balken, Treffer in entgegengesetzter Orientierung.

Bei diesen Vergleichen zeichneten sich starke Übereinstimmungen von Chr.1 aus *R. eutropha* und dem Chromosom aus *R. solanacearum* ab (Abb. 3.13). Chr.2 aus *R. eutropha* weist dagegen nur geringe Ähnlichkeiten zu beiden Replikons aus *R. solanacearum* auf (Daten nicht gezeigt).

Da sich übereinstimmende Bereiche fast ausschließlich auf kodierende Sequenzen beschränkten, wurden weitere Vergleiche auf Proteinebene durchgeführt und dabei das Verfahren des bidirektionalen BLASTs angewandt, um mehrfache Treffer auszuschließen (siehe 2.12.1). Alle Vergleiche wurden mit einem relativ strikten Grenzwert von e^{-15} vorgenommen.

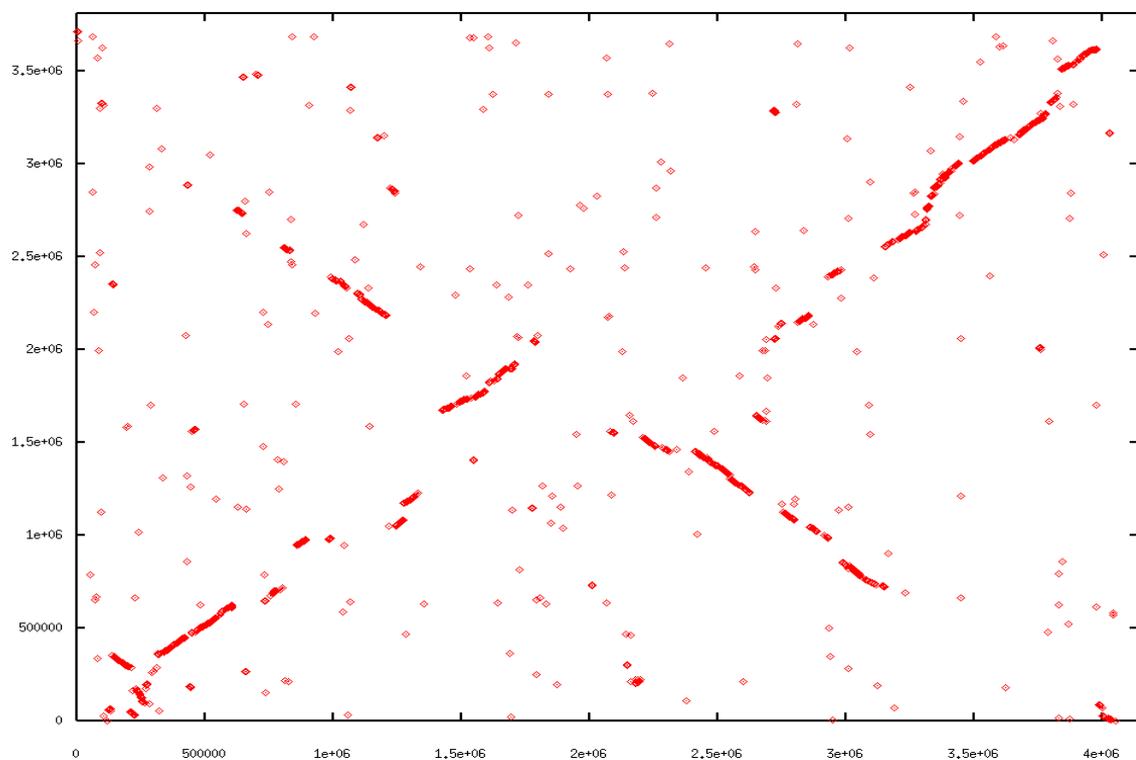


Abb. 3.14: Bidirektionaler BLAST-Vergleich auf Proteinebene zwischen Chr.1 aus *R. eutropha* und dem Chromosom aus *R. solanacearum*. Jeder Punkt stellt einen bidirektionalen Treffer zwischen *R. eutropha*, Chr.1 (x-Achse) und *R. solanacearum*, Chr. (y-Achse) mit einem e-Value ab e^{-15} dar. Nullpunkte beider Achsen stellen die Replikationsursprünge der Replikons dar.

Mehr als 50 % der vorläufig annotierten Proteine von Chr.1 aus *R. eutropha* (2008 von 3882 CDSs) zeigten signifikante Übereinstimmungen mit Proteinen vom Chromosom aus *R. solanacearum* (entsprechend 58 % der Proteine vom Chromosom aus *R. solanacearum*) (Abb. 3.14). Die überwiegende Mehrheit dieser Treffer besitzt zudem in beiden Genomen eine konservierte Umgebung: Ca. 87 % der bidirektionalen Trefferpaare haben direkte Nachbarn, die ebenfalls bidirektionale Trefferpaare darstellen. Die

konservierte Anordnung der Gene im Genom erstreckt sich über die gesamten Chromosomen und wird nur an Replikationsursprung und -terminus unterbrochen. Die Abbildung der Chromosomen aufeinander zeigt eine ungewöhnliche X-Form, die nahezu symmetrisch in Bezug auf den Replikationsursprung ist. In regelmäßigen Abständen findet also eine Umkehrung der konservierten Reihenfolge der Gene statt. Solche Richtungswechsel konservierter genomischer Bereiche wurden bereits für andere Vergleiche zwischen verwandten Organismen beobachtet (Eisen *et al.*, 2000). Als Ursache wurden chromosomale Inversionen angenommen, die aus bisher ungeklärten Gründen in gleichen Abständen vom Replikationsursprung bzw. -terminus stattfinden müssten. CDSs aus *R. eutropha*, die innerhalb potentieller genomischer Inseln liegen (siehe 3.1.3.2), weisen keine bidirektionalen BLAST-Treffer im Genom von *R. solanacearum* auf.

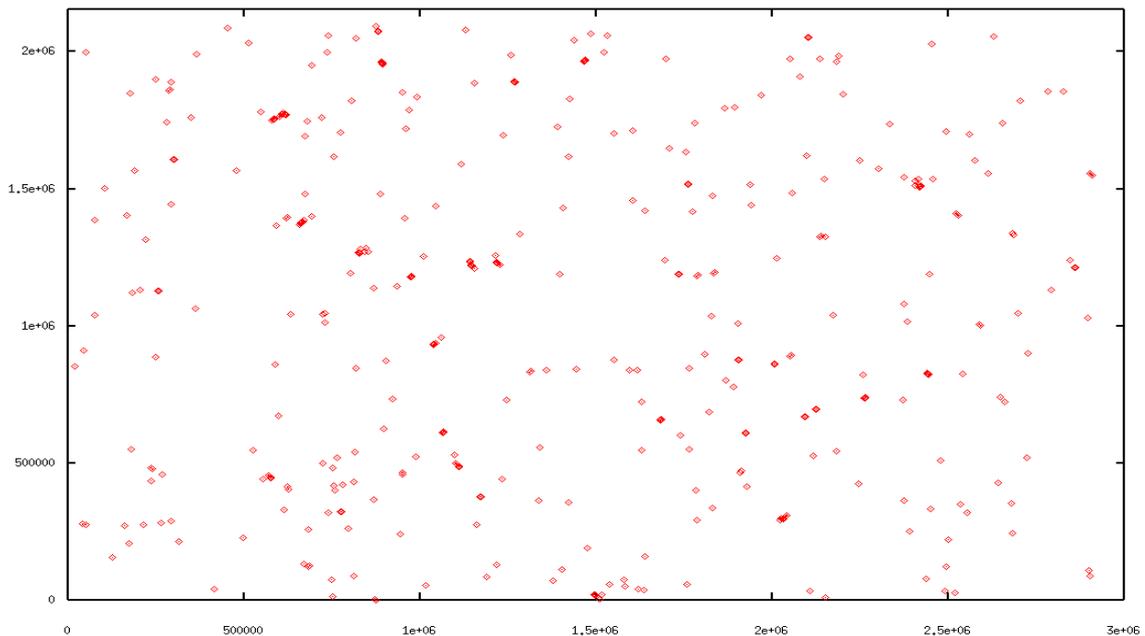


Abb. 3.15: Bidirektionaler BLAST-Vergleich auf Proteinebene zwischen Chr.2 aus *R. eutropha* und dem Megaplasmid aus *R. solanacearum*, Megaplasmid.

Die hohen Übereinstimmungen zwischen Proteinsequenzen aus *R. eutropha* und *R. solanacearum* erstrecken sich nicht auf die kleineren Replikons beider Genome (Abb. 3.15): Nur 18 % der Proteine von Chr.2 aus *R. eutropha* (465 von 2654 CDSs) zeigen signifikante Ähnlichkeit zu Proteinen vom Megaplasmid aus *R. solanacearum* (entsprechend 28 % der Proteine vom Megaplasmid aus *R. solanacearum*). Eine übergeordnete Konservierung der Genabfolgen ist zwischen diesen Replikons nicht erkennbar.

3.1.3.2 Vorhersage genomischer Inseln

Als genomische Inseln werden DNA-Bereiche bezeichnet, die aufgrund ihrer Zusammensetzung auffallend von ihrer Umgebung abweichen. Von genomischen Inseln wird angenommen, daß sie aus der Integration fremder DNA durch horizontalen Gentransfer resultieren und einen evolutiven Mechanismus des Genoms zur Anpassung an veränderte Umweltbedingungen darstellen (Doolittle, 1999). Die Integration mobiler genetischer Elemente wird durch Enzyme der Transposase/Integrase-Familie katalysiert (Haren *et al.*, 1999). In *E. coli* K12 wird von etwa 18 % aller Gene angenommen, daß sie durch horizontalen Gentransfer aufgenommen wurden (Lawrence & Ochman, 1998).

R. eutropha enthält insgesamt 92 CDSs, die für Enzyme der Transposase/Integrase-Familie kodieren und über sämtliche Replikons verteilt sind. Von insgesamt 92 CDSs, entfallen 22 auf Chr.1, 11 auf Chr.2 und 59 auf pHG1. Die meisten dieser potentiellen Gene liegen stark verkürzt vor und weisen auf Proteinebene nur fragmentarische Ähnlichkeit zu bekannten Proteinen auf. Um einen Überblick über den Anteil sogenannter Fremdgene im Genom von *R. eutropha* zu gewinnen, wurden alle CDSs einer SIGI-Analyse unterzogen (siehe 2.15). Das Programm SIGI liefert basierend auf dem Vergleich von *Codon Usages* und Cluster-Analysen eine Vorhersage für genomische Inseln (Merkl, 2004). Da die *Codon Usage* eines Proteins in besonderem Maße von der Genauigkeit der zugrundeliegenden Sequenz abhängt, können bis zum Abschluß der Feinkorrekturen keine endgültigen Aussagen über die Dichte genomischer Inseln getroffen werden. Insbesondere für Chr.1 dürfte die tatsächliche Anzahl von Fremdgenen niedriger liegen als bisher angenommen. Auf Basis der aktuellen Sequenzdaten wurden 15 % der Gene von Chr.1 (571 von 3822 CDSs), 13 % der Gene von Chr.2 (334 von 2654 CDSs) und 34 % der Gene von pHG1 (141 von 420 CDSs) als Fremdgene charakterisiert (Abb. 3.8, Abb. 3.9). Auf Chr.1 ist eine starke Konzentrierung der potentiellen Fremdgene am Replikations-Terminus zu beobachten, wohingegen sich potentielle Fremdgene auf Chr.2 über das gesamte Replikon verteilen. Die meisten identifizierten Gen-Cluster zeichnen sich durch kurze offene Leserahmen aus, die geringe Ähnlichkeit zu bekannten Proteinen aufweisen. Dementsprechend wurden die meisten potentiellen Fremdgene als „hypothetical proteins“ annotiert (22-25 %) bzw. mit sehr allgemeinen Funktionszuweisungen belegt. Die Annotation der kodierenden

Sequenzen innerhalb potentieller Genominseln wird zudem durch zwei weitere Probleme erschwert:

1. Die von ihrer Umgebung abweichenden Eigenschaften genomischer Inseln erschweren die CDS-Vorhersage (siehe 3.I.2.4). Daher enthalten genomische Inseln möglicherweise artifizielle CDSs.
2. Lateral transferierte Gene, die keinen direkten Nutzen für den Organismus darstellen, unterliegen nicht dem Evolutionsdruck und können deshalb verstärkt Mutationen ansammeln. Aus diesem Grund beinhalten genomische Inseln viele degenerierte CDSs mit geringer Ähnlichkeit zu bekannten Proteinen.

Bei ca. 7 % der von SIGI identifizierten Fremdgene handelt es sich um kodierende Sequenzen für regulatorische Proteine (Chr.1: 35 CDSs; Chr.2: 25 CDSs). 2 % wurden als potentielle Transporter annotiert (Chr.1 und Chr.2.: jeweils 10 CDSs). Von 33 CDSs der Transposase/Integrase-Familie auf den beiden großen Replikons wurden 19 CDSs als Fremdgene charakterisiert. Darüberhinaus wurden drei potentielle Phagenproteine auf Chr.1 als Fremdgene identifiziert (H16_A1800, H16_A0010, H16_A3206), nicht jedoch eine potentielle Muramidase (EC 3.2.1.17, H16_B1313), die auf Chr.2 inmitten eines Clusters potentieller Fremdgene liegt. Interessanterweise wird auf Chr.1 ein ca. 28 kbp großer Cluster aus 24 CDSs (H16_A2888-2910) als genomische Insel identifiziert, der wahrscheinlich für Enzyme der O-Antigen-Biosynthese kodiert. Innerhalb dieses Clusters lassen sich kleinere Operons unterscheiden, die jeweils große Ähnlichkeit in Sequenz und Organisation zu Operons anderer Organismen aufweisen. Das O-Antigen ist ein wichtiger Bestandteil der äußeren Membran Gram-negativer Bakterien. Strukturelle Variationen innerhalb der O-Antigene werden auf unterschiedliche Organisationsformen der kodierenden Cluster in verschiedenen Organismen zurückgeführt (Samuel & Reeves, 2003).

3.1.4 Charakterisierung von Chromosom 2

Der hohe Grad an Übereinstimmung zwischen den beiden größten Replikons aus *R. eutropha* und *R. solanacearum* (siehe 3.1.3.1) deutet darauf hin, daß sich phänotypische Unterschiede zwischen diesen verwandten Organismen hauptsächlich auf kodierende Eigenschaften der kleineren Replikons zurückführen lassen. Da die phänotypische Relevanz von pHG1 bereits gut untersucht wurde (Schwartz *et al.*, 2003), soll der folgende Teil dieser Arbeit die Bedeutung von Chr.2 für die Charakterisierung von *R. eutropha* herausstellen. Dazu wurden exemplarisch einige Eigenschaften von *R. eutropha* anhand dieses Replikons untersucht und im Verhältnis zu Chr.1 dargestellt.

3.1.4.1 Fakultativ chemolitho- und organoautotrophes Wachstum

Genetische und physiologische Grundlagen des fakultativ lithoautotrophen Stoffwechsels in *R. eutropha* wurden bereits ausführlich untersucht (Kusian & Bowien, 1997; Lenz *et al.*, 2002; Schwartz *et al.*, 2003). *R. eutropha* ist in der Lage, heterotroph auf einer Vielzahl organischer Substrate zu wachsen und diese als Energie- und Kohlenstoffquellen zu nutzen (Wilde, 1962). In Abwesenheit geeigneter organischer Substrate kann *R. eutropha* auf CO₂ wachsen, das über den reduktiven Pentose-Phosphat-Cyclus nach Calvin-Benson-Bassham (CBB-Cyclus) assimiliert wird. Die für die Fixierung von CO₂ benötigte Energie bezieht *R. eutropha* aus der Oxidation von Formiat (organoautotrophes Wachstum) oder molekularem Wasserstoff (lithoautotrophes Wachstum). In *R. eutropha* ist der Enzymapparat des CBB-Cyclus in zwei funktionellen *cbb*-Operons auf dem Megaplasmid pHG1 und auf dem Chromosom kodiert, die bereits ausführlich untersucht und beschrieben wurden (Bowien & Kusian, 2002). Beide Operons werden über einen gemeinsamen Regulator, *cbbR*, kontrolliert, der chromosomal kodiert ist (Windhövel & Bowien, 1991). Das plasmidäre *cbbR*'-Gen ist aufgrund mehrerer kurzer Deletionen defekt. Entscheidende genetische Informationen für das lithoautotrophe Wachstum von *R. eutropha* sind auf dem Megaplasmid pHG1 kodiert. Innerhalb eines Lithoautotrophie-Clusters auf pHG1 kodieren 41 Gene für die Biosynthese von insgesamt vier Hydrogenasen. Im selben Cluster liegt das plasmidäre *cbb*-Operon. Die zweite *cbb*-Region, die das aktive Regulatorgen *cbbR* (H16_B1396) enthält, wurde auf Chr.2

lokalisiert. Außerhalb von *cbb_c* enthält Chr.2 ein weiteres, verkürztes *cbbR*-Gen (H16_B1429), dessen potentiell Produkt das im N-Terminus kodierte DNA-bindende Helix-Turn-Helix-Motiv fehlt. Von diesem Gen wurde gezeigt, daß es ebenfalls für keinen funktionellen Regulator kodiert (B. Kusian, persönliche Mitteilung).

Die Oxidation von Formiat als Grundlage organoautotrophen Wachstums wird in *R. eutropha* von zwei verschiedenen Typen der Formiat-Dehydrogenase katalysiert, einem NAD⁺-abhängigen Enzym (S-FDH, EC 1.2.1.2) und einem Membran-gebundenen Enzym (Friedrich *et al.*, 1979). Das NAD⁺-abhängige Enzym ist im *fdsGBACD*-Operon kodiert (Oh & Bowien, 1998) und befindet sich auf Chr.1 (Tab. 3.2).

Tab. 3.2: Gen-Cluster für potentielle Formiat-Dehydrogenasen auf Chr.1.

Nr.	CDS	Gen	Annotation
1	H16_A0639	<i>fdsR</i>	transcriptional regulator
2	H16_A0640	<i>fdsG</i>	NAD-dependent formate dehydrogenase gamma subunit
3	H16_A0641	<i>fdsB</i>	NAD-dependent formate dehydrogenase beta subunit
4	H16_A0642	<i>fdsA</i>	NAD-dependent formate dehydrogenase alpha subunit
5	H16_A0643	<i>fdsC</i>	FdsC protein
6	H16_A0644	<i>fdsD</i>	NAD-dependent formate dehydrogenase delta subunit
1	H16_A2942		conserved hypothetical protein
2	H16_A2941		conserved hypothetical protein
3	H16_A2940		ferredoxin
4	H16_A2939		formate dehydrogenase-specific chaperone
5	H16_A2938		hypothetical protein
6	H16_A2937		formate dehydrogenase alpha subunit
7	H16_A2936		formate dehydrogenase iron-sulfur subunit
8	H16_A2935		hypothetical membrane-associated protein
9	H16_A2934		formate dehydrogenase cytochrome b subunit
10	H16_A2933		hypothetical membrane-associated protein
11	H16_A2932		formate dehydrogenase accessory protein
1	H16_A3292		formate dehydrogenase alpha chain

Auf Chr.1 befindet sich ein weiterer Cluster aus 12 Genen, der möglicherweise für eine zweite, membrangebundene Formiat-Dehydrogenase kodiert. Verschiedene Gene dieses Clusters kodieren nach TMpred-Vorhersage für Proteine, welche Transmembran-Bereiche enthalten. Untersuchungen zur Cofaktor-Abhängigkeit der membrangebundenen Formiat-Dehydrogenase deuten allerdings auf die Existenz von wenigstens drei verschiedenen Enzymen hin (Burgdorf *et al.*, 2001). Tab. 3.3 gibt einen Überblick über weitere Cluster auf Chr.2, die für potentielle Formiat-Dehydrogenasen unterschiedlichen Typs bzw. deren Untereinheiten kodieren.

Tab. 3.3: Gen-Cluster für potentielle Formiat-Dehydrogenasen auf Chr.2.

Nr	CDS	Annotation
1	H16_B1702	oxalate:formate antiporter
2	H16_B1701	tungsten-containing formate dehydrogenase beta subunit
3	H16_B1700	tungsten-containing formate dehydrogenase alpha subunit
1	H16_B1452	nitrate-inducible formate dehydrogenase alpha subunit
2	H16_B1453	nitrate-inducible formate dehydrogenase beta subunit
3	H16_B1454	nitrate-inducible formate dehydrogenase gamma subunit
4	H16_B1455	formate dehydrogenase accessory protein
1	H16_B1472	transcriptional regulator
2	H16_B1471	formate dehydrogenase alpha subunit
3	H16_B1470	formate dehydrogenase beta subunit
4	H16_B1702	oxalate:formate antiporter

Interessanterweise wurden sowohl auf Chr.1 als auch auf Chr.2 verschiedene Cluster identifiziert, die unter anderem Ähnlichkeiten zu dem in *Starkeya novella* (vormals *Thiobacillus novellus*) charakterisierten *soxB*CD-Operon (Kappler *et al.*, 2001) aufweisen. Die Fähigkeit zur Oxidation von Schwefel, die wahrscheinlich über das *sox*-Operon vermittelt wird, ermöglicht *Starkeya novella* ein fakultativ chemolithotrophes Wachstum (Starkey, 1935). *R. eutropha* ist dagegen nicht in der Lage, Schwefel zu oxidieren (C.G. Friedrich, persönliche Mitteilung). Ob die im Genom von *R. eutropha* identifizierten *sox*-Gene für funktionelle Proteine kodieren oder welche Enzyme für eine vollständige Schwefeloxidation fehlen, bedarf noch weiterer Klärung.

3.1.4.2 Erweiterung des Substratspektrums

3.1.4.2.1 Alternative Kohlenstoffquellen

3.1.4.2.1.1 Abbau von Kohlenhydraten

Bereits während der ersten Charakterisierung von *R. eutropha* H16 war gezeigt worden, daß dieser Organismus von allen getesteten Zuckersubstraten nur auf Fructose wächst (Bartha, 1962; Wilde, 1962). Kein Wachstum hingegen wurde auf Saccharose, Glucose, Galactose, Mannose, Arabinose, Ribose, Sorbose, Xylose und Xylulose beobachtet. Von den getesteten Säuren, die sich aus Zuckern ableiten, wurde eine Umsetzung nur

für Gluconat (Blackkolb & Schlegel, 1968) und 2-Ketogluconat (Nandadasa *et al.*, 1974) gefunden. Der Abbau beider Substrate erfolgt ausschließlich über den Entner-Doudoroff-Weg (Gottschalk *et al.*, 1964). Für die Schlüsselenzyme des oxidativen Pentose-Phosphat-Weges (6-Phosphogluconat-Dehydrogenase, EC 1.1.1.44) und des Embden-Meyerhof-Parnas-Weges (6-Phosphofruktokinase, EC 2.7.1.11) konnten dagegen keine Aktivitäten festgestellt werden (Gottschalk, 1964). Diese Ergebnisse wurden nun durch die Genomanalyse bestätigt, bei der weder für 6-Phosphogluconat-Dehydrogenasen noch für 6-Phosphofruktokinasen potentiell kodierende Gene identifiziert wurden. Daß *R. eutropha* nicht in der Lage ist, Glucose umzusetzen, obwohl die nötigen Enzyme für den Abbau von Glucose über den ED-Weg vorhanden sind (Gottschalk *et al.*, 1964), wird mit dem Fehlen eines aktiven Transportsystems begründet (König *et al.*, 1969).

Der Entner-Doudoroff-Weg stellt gegenüber dem Embden-Meyerhof-Parnas-Weg (EMP) einen alternativen Abbauweg für Kohlenhydrate dar, welcher eine niedrigere ATP-Ausbeute aufweist und deshalb als ursprünglicher angesehen wird. Weit verbreitet ist der ED-Weg in aeroben oder anaeroben, respiratorischen Organismen, die ATP nicht nur aus der Substratstufen- sondern auch aus der Elektronentransport-Phosphorylierung beziehen. Im ED-Weg wird Glucose-6-Phosphat zu Pyruvat und Glycerinaldehyd-3-Phosphat oxidiert und dabei ein NAD(P)^+ reduziert. Bei der folgenden Oxidation von Glycerinaldehyd-3-Phosphat zu Pyruvat wird ein weiteres NAD^+ reduziert und netto ein ATP gewonnen. Im Vergleich zur Glykolyse und dem oxidativen Pentose-Phosphat-Weg beinhaltet der ED-Weg nur zwei zusätzliche Reaktionen:

1. 6-Phosphogluconat wird durch die 6-Phosphogluconat-Dehydratase zu 2-Keto-3-Desoxygluconat-6-Phosphat (KDPG) dehydratisiert.
2. KDPG wird durch die KDPG-Aldolase in Pyruvat und Glycerinaldehyd-3-Phosphat gespalten.

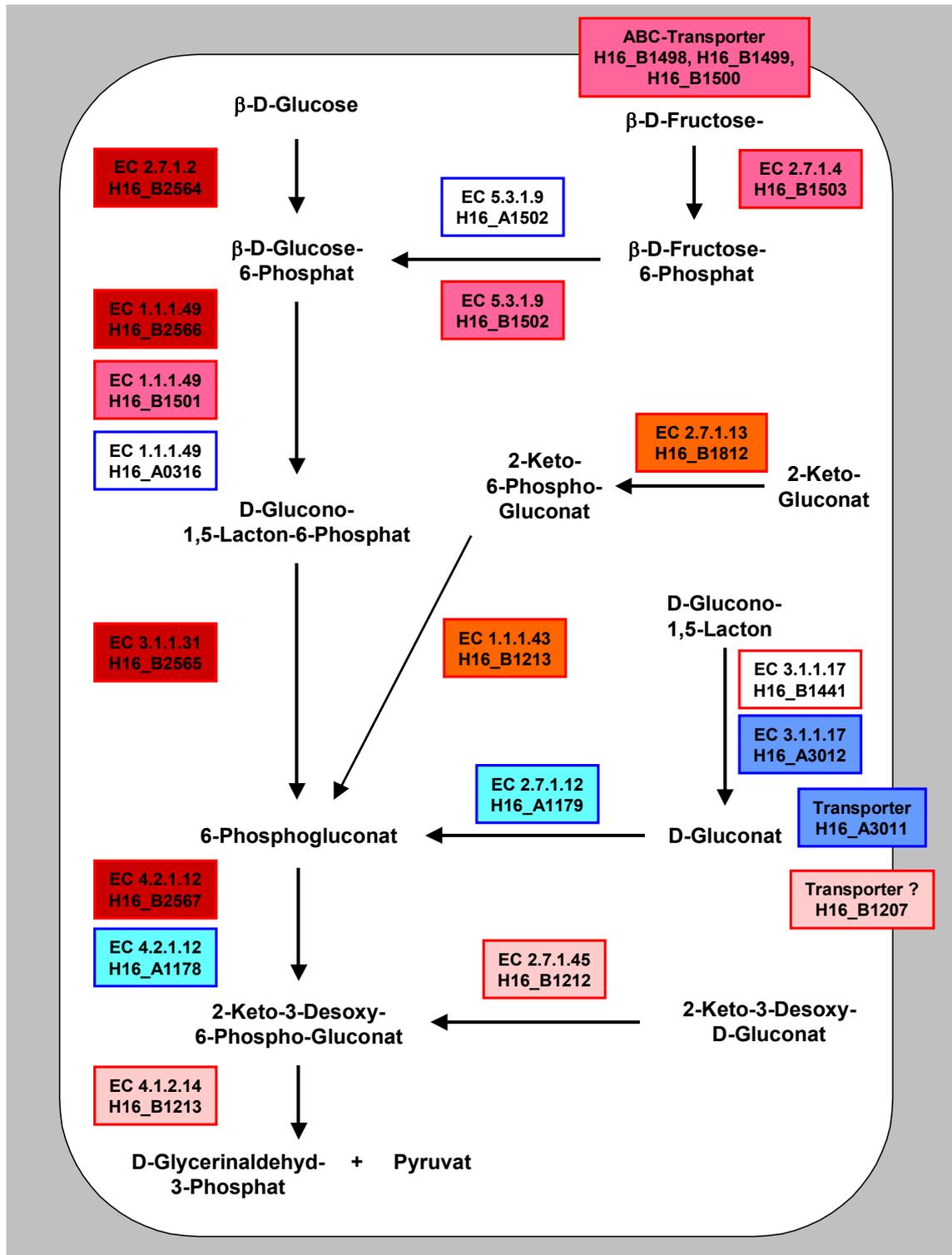


Abb. 3.16: Schematische Darstellung des Abbaus verschiedener Substrate in *R. eutropha* über den Entner-Doudoroff-Weg nach KEGG (*Kyoto Encyclopedia of Genes and Genomes*). Proteine sind - soweit möglich - durch EC-Nummern und durch die Bezeichnung der kodierenden Sequenz gekennzeichnet. Blau umrandet, Proteine, die auf Chr.1 kodiert sind; rot umrandet, Proteine, die auf Chr.2 kodiert sind. CDSs zu Proteinen gleicher Hintergrundfarbe sind in einem gemeinsamen Cluster lokalisiert.

Was den Enzymapparat des ED-Weges in *R. eutropha* angeht, so ist eine klare Zuordnung verschiedener Bereiche zu den beiden großen Replikons möglich:

- Die Aufnahme und Umsetzung von Fructose und die Umsetzung von Glucose über den ED-Weg sind vollständig auf Chr.2 und in wesentlichen Bestandteilen auch ausschließlich auf Chr.2 kodiert (Abb. 3.16). Zu den Genen, die nur auf Chr.2 vorliegen, gehören das Aufnahme-System für Fructose, Fructokinase (EC 2.7.1.4) und Glucokinase (EC 2.7.1.2), 6-Phospho-Gluconolactonase (EC 3.1.1.31) und KDPG-Aldolase (EC 4.1.2.14).
- Für den Abbau von Gluconat über den ED-Weg wird dagegen die Gluconat-Kinase (EC 2.7.1.12) benötigt, die ausschließlich auf Chr.1 kodiert sind, ebenso wie zwei weitere Enzyme, die die Umsetzung von 2-Ketogluconat zu 6-Phosphogluconat katalysieren (Abb. 3.16). Verschiedene andere Enzyme sind sowohl auf Chr.1 als auch auf Chr.2 kodiert.

Die Enzyme des ED-Weges sind in *R. eutropha* im wesentlichen in fünf Clustern kodiert, von denen sich vier auf Chr.2 befinden:

Cluster 1 (Tab. 3.4) weist starke Ähnlichkeit zu dem in *Sinorhizobium meliloti* charakterisierten *frc*-Operon auf (Lambert *et al.*, 2001). Cluster 1 kodiert für ein potentiell Fructose-ABC-Transportsystem (H16_B1498-1500), eine Fructokinase (EC 2.7.1.4, H16_B1503), eine von zwei Phosphoglucose-Isomerasen (EC 5.3.1.9, H16_B1502) und eine von drei Glucose-6-Phosphat-Dehydrogenasen (EC 1.1.1.49, H16_B1501). Ob es sich bei Cluster 1 tatsächlich um das für Aufnahme und Umsetzung von Fructose verantwortliche Operon handelt, wird zur Zeit untersucht (B. Bowien, persönliche Mitteilung).

Tab. 3.4: Entner-Doudoroff-Weg, Cluster 1 auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_B1497	<i>frcR</i> ^a	transcriptional regulator	
2	H16_B1498	<i>frcA</i> ^a	ABC transporter, ATP-binding protein	
3	H16_B1499	<i>frcC</i> ^a	ABC transporter, putative fructose permease	
4	H16_B1500	<i>frcB</i> ^a	putative fructose binding protein	
5	H16_B1501	<i>zwf2</i>	glucose-6-phosphate 1-dehydrogenase	1.1.1.49
6	H16_B1502	<i>pgi2</i>	glucose-6-phosphate isomerase	5.3.1.9
7	H16_B1503	<i>frc</i>	fructokinase	2.7.1.4

^a aus *Sinorhizobium meliloti* (Lambert *et al.*, 2001).

Cluster 2 (Tab. 3.5) kodiert für eine potentielle Glucokinase (H16_B2564), eine weitere Glucose-6-Phosphat-Dehydrogenase (H16_B2566) sowie die einzige 6-Phospho-Gluconolactonase (H16_B2565) und eine Phosphogluconat-Dehydratase (EC 4.2.1.12, H16_B2567).

Tab. 3.5: Entner-Doudoroff-Weg, Cluster 2 auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_B2567	<i>edd</i>	phosphogluconate dehydratase	4.2.1.12
2	H16_B2566	<i>zwf3</i>	glucose-6-phosphate 1-dehydrogenase	1.1.1.49
3	H16_B2565	<i>pgl</i>	6-phosphogluconolactonase	3.1.1.31
4	H16_B2564	<i>glk</i>	glucokinase	2.7.1.2
5	H16_B2563		two component response regulator	
6	H16_B2562		two component sensor histidine kinase	

Cluster 3 (Tab. 3.6) kodiert für das Schlüsselenzym des ED-Weges, die einzige 2-Keto-3-Desoxy-6-Phosphogluconat-Aldolase im Genom von *R. eutropha* (H16_B1213). Durch die 2-Keto-3-Desoxyglucokinase (EC 2.7.1.45, H16_B1212) könnte mit 2-Keto-3-Desoxygluconat ein zusätzliches Substrat in den ED-Weg eingeschleust werden. Schließlich enthält Cluster 3 eine von zwei Gluconat-Permeasen (H16_B1207).

Tab. 3.6: Entner-Doudoroff-Weg, Cluster 3 auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_B1213	<i>eda</i>	2-keto-3-deoxy-6-phosphogluconate aldolase	4.1.2.14
2	H16_B1212	<i>kdgK</i>	2-keto-3-deoxygluconate kinase	2.7.1.45
3	H16_B1211		amino acid aldolase or racemase	
4	H16_B1210		transcriptional regulator	
5	H16_B1209		N-acyl-D-amino acid deacylase	3.5.1.14
6	H16_B1208		endoribonuclease	
7	H16_B1207	<i>gntP</i>	gluconate permease	

Cluster 4 auf Chr.2 (Tab. 3.7) kodiert für eine 2-Ketogluconat-Kinase (EC 2.7.1.13, H16_B1812) und eine potentielle 2-Keto-6-Phosphogluconat-Reduktase (EC 1.1.1.43, H16_B1813). Aktivitäten für beide Enzyme wurden in *R. eutropha* nachgewiesen wurden (Nandadasa *et al.*, 1974) und könnten für den Abbau von 2-Ketogluconat über den ED-Weg verantwortlich sein.

Tab. 3.7: Entner-Doudoroff-Weg, Cluster 4 auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_B1812	<i>kgk^a</i>	2-Ketogluconatkinase	2.7.1.13
2	H16_B1813	<i>kgf^a</i>	2-keto-6-phosphogluconate reductase	1.1.1.43

^a aus *Pseudomonas putida* (Vicente & Canovas, 1973).

Cluster 5 auf Chr.1 (Tab. 3.8) kodiert für eine zweite Gluconat-Permease (H16_A3011). In diesem Cluster ist außerdem eine von zwei Gluconolactonasen kodiert (EC 3.1.1.17, H16_A3012), mit deren Hilfe ein weiteres potentielles Substrat, das D-Glucono-1,5-Lacton, in den Entner-Doudoroff-Weg eingespeist werden könnte. Die zweite Gluconolactonase ist auf Chr.1 kodiert (Abb. 3.16).

Tab. 3.8: Entner-Doudoroff-Weg, Cluster 5 auf Chr.1.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_A3011	<i>gntP</i>	gluconate permease	
2	H16_A3012	<i>gnt1</i>	gluconolactonase	3.1.1.17

Cluster 6 auf Chr.1 (Tab. 3.9) kodiert für die einzige Gluconatkinase im Genom von *R. eutropha* (EC 2.7.1.12, H16_A1179) sowie eine weitere Phosphogluconat-Dehydratase (H16_A1178).

Tab. 3.9: Entner-Doudoroff-Weg, Cluster 5 auf Chr.1.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_A1177		transcriptional regulator	
2	H16_A1178	<i>edd1</i>	phosphogluconate dehydratase	4.2.1.12
3	H16_A1179	<i>gntV</i>	gluconate kinase	2.7.1.12

Cluster 6 auf Chr.1 kodiert für eine potentielle 2-Ketogluconat-Kinase und eine potentielle 2-Keto-6-Phosphogluconat-Reduktase, deren Aktivitäten in *R. eutropha* nachgewiesen wurden (Nandadasa *et al.*, 1974).

3.1.4.2.1.2 Abbau von aus Zuckern abgeleiteten Säuren

Möglicherweise erweitert Chr.2 das Substratspektrum von *R. eutropha* um zwei weitere von Zuckern abgeleitete Säuren: Glucarat und Galactarat (Abb. 3.17). Drei Gene, die innerhalb eines Clusters von insgesamt sechs Genen, jedoch auf unterschiedlichen Strängen lokalisiert sind, kodieren für die Enzyme des vollständigen Abbaus von D-Glucarat zu 2-Oxoglucarat: Glucarat-Dehydratase (EC 4.2.1.40, H16_B0127), 5-Dehydro-4-Desoxyglucarat-Dehydratase (EC 4.2.1.41, H16_B0131) und 2-Oxo-Glutarat-Semialdehyd-Dehydrogenase (EC 1.2.1.26, H16_B0130). Zwei weitere CDSs innerhalb dieses Clusters (H16_B0128, H16_B0129) weisen Ähnlichkeiten auf Protein-

ebene zu einer Domäne auf, die als *Bordetella uptake gene product* beschrieben wurde (pfam03401, Bug). Dabei handelt es sich wahrscheinlich um ein extraplasmatisches Bindeprotein, das in vielen β -Proteobakterien stark überrepräsentiert ist (Antoine *et al.*, 2003). Im gleichen Cluster ist ein Transkriptionsregulator kodiert (H16_B0126), der möglicherweise an der Regulation des Clusters beteiligt ist. An anderer Stelle auf Chr.2 ist eine potentielle D-Galactarat-Dehydratase kodiert (H16_B0126), die sich keinem Cluster zuordnen läßt. Dieses Enzym könnte mit D-Galactarat ein zusätzliches Substrat in den Glucarat-Abbau einspeisen.

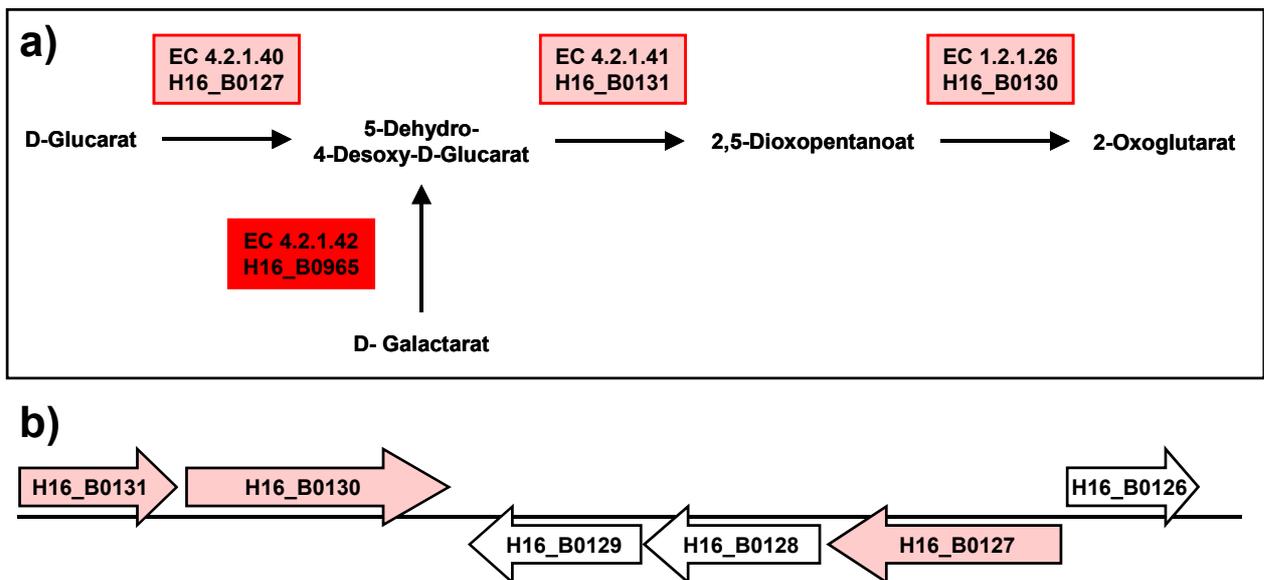


Abb. 3.17: a) Schematische Darstellung des D-Glucarat- und D-Galactarat-Abbaus nach KEGG; b) Cluster kodierender Sequenzen für Enzyme des D-Glucarat-Abbaus.

Der gesamte Enzymapparat des Glucarat- und Galactarat-Abbaus ist in *R. solanacearum* auf dem Megaplasmid kodiert, allerdings innerhalb eines einzigen Operons, zu dem außerdem ein potentieller Glucarat/Galactarat-Transporter gehört.

3.1.4.2.1.3 Abbau aromatischer Kohlenwasserstoffe

Der aerobe Abbau aromatischer Kohlenwasserstoffe gliedert sich im wesentlichen in folgende Schritte (Lengeler *et al.*, 1999):

1. Je nach Anzahl und Position der Substitutionen des aromatischen Ringes werden unterschiedliche Substrate in wenige zentrale Zwischenprodukte überführt, von denen Catechol, Protocatechuat und Gentisat die häufigsten darstellen. Generell

werden 1,2-disubstituierte, nicht-substituierte und viele monosubstituierte Aromaten über Catechol abgebaut, 1,3- und 1,4-disubstituierte und polysubstituierte Aromaten dagegen über Protocatechuat. Tryptophan wird über Anthranilat zu Catechol umgesetzt.

2. Dioxygenasen spalten die aromatischen Ringe von Catechol und Protocatechuat. Je nach Position der oxidativen Spaltung am zentralen aromatischen Zwischenprodukt unterscheidet man zwischen *ortho*- und *meta*-Spaltung.
3. Nicht-zyklische Produkte der oxidativen Spaltung werden weiter abgebaut zu Zwischenprodukten zentraler Stoffwechselwege.

R. eutropha ist in der Lage, verschiedene aromatische Verbindungen als ausschließliche Energie- und Kohlenstoffquellen zu nutzen. Zu diesen Verbindungen gehören Benzoat, davon abgeleitete m- und p-Hydroxy-Derivate, Phenol und Tryptophan (Davis *et al.*, 1969; Johnson & Stanier, 1971). Enzymatische Untersuchungen konnten zeigen, daß Benzoat und p-Hydroxybenzoat über den β -Ketoacid-Weg, m-Hydroxybenzoat über den Gentisat-Weg und Phenol über den *meta*-Spaltungs-Weg abgebaut werden (Johnson & Stanier, 1971). Die Komponenten des aeroben Abbaus aromatischer Kohlenwasserstoffe verteilen sich im Genom von *R. eutropha* auf eine Reihe größerer und kleinerer Cluster auf Chr.1 und Chr.2 (Tabellen 3.9-12; Abb. 3.18, Abb. 3.19). Im wesentlichen lassen sich die kodierenden Eigenschaften beider Replikons folgendermaßen zusammenfassen:

- Chr.1 kodiert für den vollständigen Abbau aromatischer Kohlenwasserstoffe, die über das Zwischenprodukt Catechol und den β -Ketoacid-Weg umgesetzt werden (Tab. 3.10, Abb. 3.18) und für den Abbau von Tryptophan. Einzelne Schritte anderer Abbau-Wege sind ebenfalls auf Chr.1 kodiert (Abb. 3.19).
- Chr.2 erweitert einerseits den β -Ketoacid-Weg für Substrate, die über das Zwischenprodukt Protocatechuat umgesetzt werden (Tab. 3.11, Abb. 3.18) und kodiert andererseits für den vollständigen Abbau phenolischer Substrate über den *meta*-Spaltungs-Weg (Tab. 3.12, Abb. 3.19). Weitere Cluster auf Chr.2 kodieren für einen Teil des auf Chr.1 kodierten β -Ketoacid-Weges und für weitere Abbau-Wege, die auf Basis der Sequenzdaten nicht eindeutig annotiert werden können (Tab. 3.13).

Tab. 3.10: Abbau aromatischer Kohlenwasserstoffe, CDSs auf Chr.1.

Nr.	CDS	Gen	Annotation	EC-Nummer
1. Abbau von Benzoat und Dihydroxybenzoat über β-Ketoacidat-Weg				
1	H16_A1965	<i>benM</i>	benzoate operon transcriptional regulator	
2	H16_A1964	<i>catA</i>	catechol 1,2-dioxygenase	1.13.11.1
3	H16_A1963	<i>benA</i>	benzoate 1,2-dioxygenase	1.14.12.10
4	H16_A1962	<i>benB</i>	hydroxylase component, alpha subunit benzoate 1,2-dioxygenase	1.14.12.10
5	H16_A1961	<i>benC</i>	hydroxylase component, beta subunit benzoate 1,2-dioxygenase	1.14.12.10
6	H16_A1960	<i>benD</i>	electron transfer component cis-1,2-dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase	1.3.1.55
7	H16_A1959		predicted hydrolase or acyltransferase	
8	H16_A1966	<i>catB1</i>	muconate cyclo-isomerase	5.5.1.1
9	H16_A1967	<i>catC</i>	muconolactone delta-isomerase	5.3.3.4
10	H16_A1968	<i>catD1</i>	3-oxoadipate enol-lactone hydrolase	3.1.1.24
11	H16_A1969		conserved hypothetical protein	
2. Abbau von Tryptophan zu Antranilat				
1	H16_A2814		predicted metal-dependent hydrolase	
2	H16_A2815		L-kynurenine hydrolase	3.7.1.3
3	H16_A2816		tryptophan 2,3-dioxygenase	1.13.11.11
3. Komponenten aus dem Aromaten-Abbau über meta-Spaltung				
1	H16_A1806		acetaldehyde dehydrogenase oxidoreductase	1.2.1.10
2	H16_A1807		4-hydroxy-2-oxovalerate aldolase	4.1.3.-
4. Einzelne Gene				
	H16_A2672		benzoate transporter	
	H16_A2422		3-carboxy-cis,cis-muconate cyclo-isomerase	5.5.1.2

Tab. 3.11: Abbau aromatischer Kohlenwasserstoffe über den β -Ketoacidat-Weg, CDSs auf Chr.2.

Nr.	CDS	Annotation	EC-Nummer
1. Abbau von 4-Hydroxybenzoat zu 3-Oxoadipat			
1	H16_B2293	4-hydroxybenzoate transporter	
2	H16_B2292	transcriptional regulator	
3	H16_B2291	protocatechuate 3,4-dioxygenase beta chain	1.13.11.3
4	H16_B2290	protocatechuate 3,4-dioxygenase alpha chain	1.13.11.3
5	H16_B2289	3-carboxy-cis,cis-muconate cycloisomerase	5.5.1.2
6	H16_B2288	4-carboxymuconolactone decarboxylase	4.1.1.44
7	H16_B2287	transcriptional regulator	
8	H16_B2286	p-hydroxybenzoate hydroxylase	1.14.13.2
2. Abbau von 3-Oxoadipat			
1	H16_B0198	3-oxoadipate CoA-transferase subunit A	2.8.3.6
2	H16_B0199	3-oxoadipate CoA-transferase subunit B	2.8.3.6
3	H16_B0200	beta-ketoacidyl CoA thiolase	2.3.1.16
4	H16_B0201	short chain dehydrogenase	1.-.-.-
5	H16_B0202	putative exported protein	
3. Weitere einzelne Gene aus dem β-Ketoacidat-Weg			
	H16_B1584	muconolactone delta-isomerase 2	5.3.3.4
	H16_B1583	3-oxoadipate enol-lactone hydrolase	3.1.1.24
	H16_B0536	muconate cyclo-isomerase	5.5.1.1

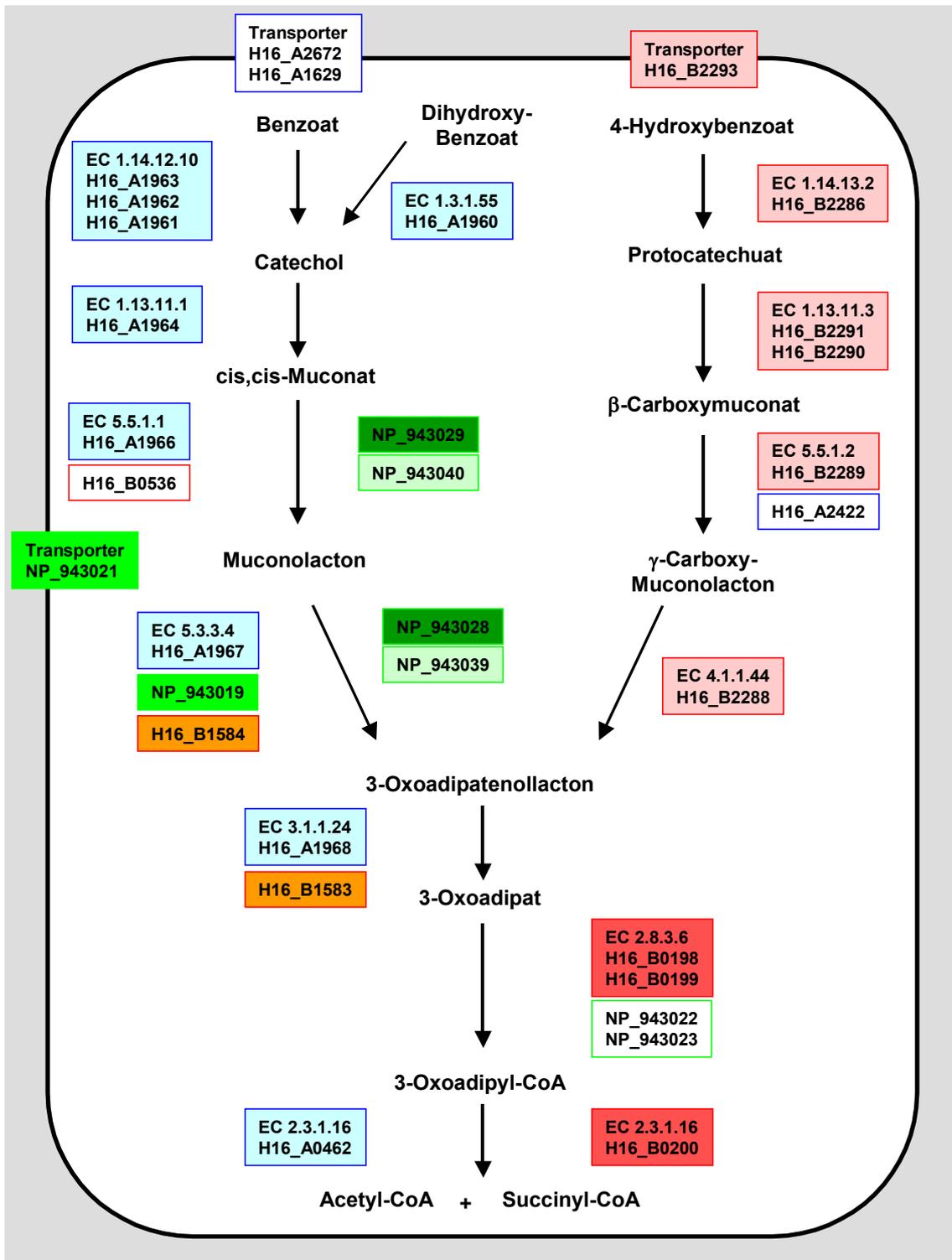


Abb. 3.18: Schematische Darstellung des Abbaus aromatischer Kohlenwasserstoffe in *R. eutropha* über den β -Ketoadipat-Weg nach KEGG (*ortho*-Spaltung). Blau umrandet, Proteine, die auf Chr.1 kodiert sind; rot umrandet, Proteine, die auf Chr.2 kodiert sind; grün umrandet, Proteine, die auf pHG1 kodiert sind. CDSs zu Proteinen gleicher Hintergrundfarbe sind in einem gemeinsamen Cluster lokalisiert.

Der vollständige Abbau von Phenol über Catechol zu Pyruvat und Acetyl-CoA durch den *meta*-Spaltungsweg ist in einem einzigen Cluster kodiert (Tab. 3.12, Abb. 3.19), der starke Ähnlichkeit zum *dmp*-Operon aufweist, das in diesem Zusammenhang auf dem Megaplasmid pVI150 in *Pseudomonas sp.* CF600 charakterisiert wurde (Shingler *et al.*, 1992). Ein weiterer Cluster kodiert möglicherweise für die Enzyme zum Abbau von Biphenyl über den *meta*-Spaltungsweg. Einige CDSs weisen Ähnlichkeiten zu Genen des in *Burkholderia xenovorans* LB400 charakterisierten *bph*-Operons auf (Hofer *et al.*, 1994).

Tab. 3.12: Abbau aromatischer Kohlenwasserstoffe über den *meta*-Spaltungsweg, CDSs auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1. Abbau von Phenol über meta-Spaltung				
1	H16_B0535		transcriptional regulator, LysR family	
2	H16_B0536	<i>catB2</i>	muconate cyclo-isomerase	5.5.1.1
3	H16_B0537		putative exported protein	
4	H16_B0538	<i>dmpK^b</i>	activator of phenol-degradative genes	
5	H16_B0539	<i>dmpL^b</i>	phenol hydroxylase P0 protein	1.14.13.7
6	H16_B0540	<i>dmpM^b</i>	phenol hydroxylase P1 protein	1.14.13.7
7	H16_B0541	<i>dmpN^b</i>	phenol hydroxylase P2 protein	1.14.13.7
8	H16_B0542	<i>dmpO^b</i>	phenol hydroxylase P3 protein	1.14.13.7
9	H16_B0543	<i>dmpP^b</i>	phenol hydroxylase P4 protein	1.14.13.7
10	H16_B0544	<i>dmpQ^b</i>	phenol hydroxylase P5 protein	1.14.13.7
11	H16_B0545	<i>dmpB^b</i>	2Fe-2S ferredoxin	
12	H16_B0546	<i>dmpC^b</i>	catechol 2,3-dioxygenase	1.13.11.2
13	H16_B0547	<i>dmpE^b</i>	2-hydroxymuconic semialdehyde dehydrogenase	1.2.1.32
14	H16_B0548	<i>dmpH^b</i>	2-hydroxypent-2,4-dienoate hydratase	4.2.1.80
15	H16_B0549	<i>dmpI^b</i>	4-oxalocrotonate decarboxylase	4.1.1.77
16	H16_B0550	<i>dmpF^b</i>	4-oxalocrotonate tautomerase	5.3.2.-
17	H16_B0551	<i>dmpG^b</i>	acetaldehyde dehydrogenase	1.2.1.10
18	H16_B0552	<i>dmpK^b</i>	4-hydroxy-2-oxovalerate aldolase	4.1.3.-
2. wahrscheinlich Abbau von Biphenyl über meta-Spaltung				
1	H16_B0600	<i>bphD</i>	2,6-dioxo-6-phenylhexa-3-enoate hydrolase	3.7.1.8
2	H16_B0599		oxidoreductase short-chain dehydrogenase	
3	H16_B0598		hypothetical membrane associated protein	
4	H16_B0597	<i>bphH</i>	2-oxopent-4-enoate hydratase	4.2.1.80
5	H16_B0596		acetaldehyde dehydrogenase	1.2.1.10
6	H16_B0595	<i>bphI</i>	4-hydroxy-2-oxovalerate aldolase	4.1.3.-
3. Einzelnes Gen aus dem Abbau von Biphenyl über meta-Spaltung				
	H16_B0654	<i>bphC</i>	biphenyl-2,3-diol 1,2-dioxygenase	1.13.11.39

^a aus *Acinetobacter lwoffii* K24 (Kim *et al.*, 1998).

^b aus *Pseudomonas sp.* CF600 (Shingler *et al.*, 1992).

^c aus *Burkholderia xenovorans* (Hofer *et al.*, 1994).

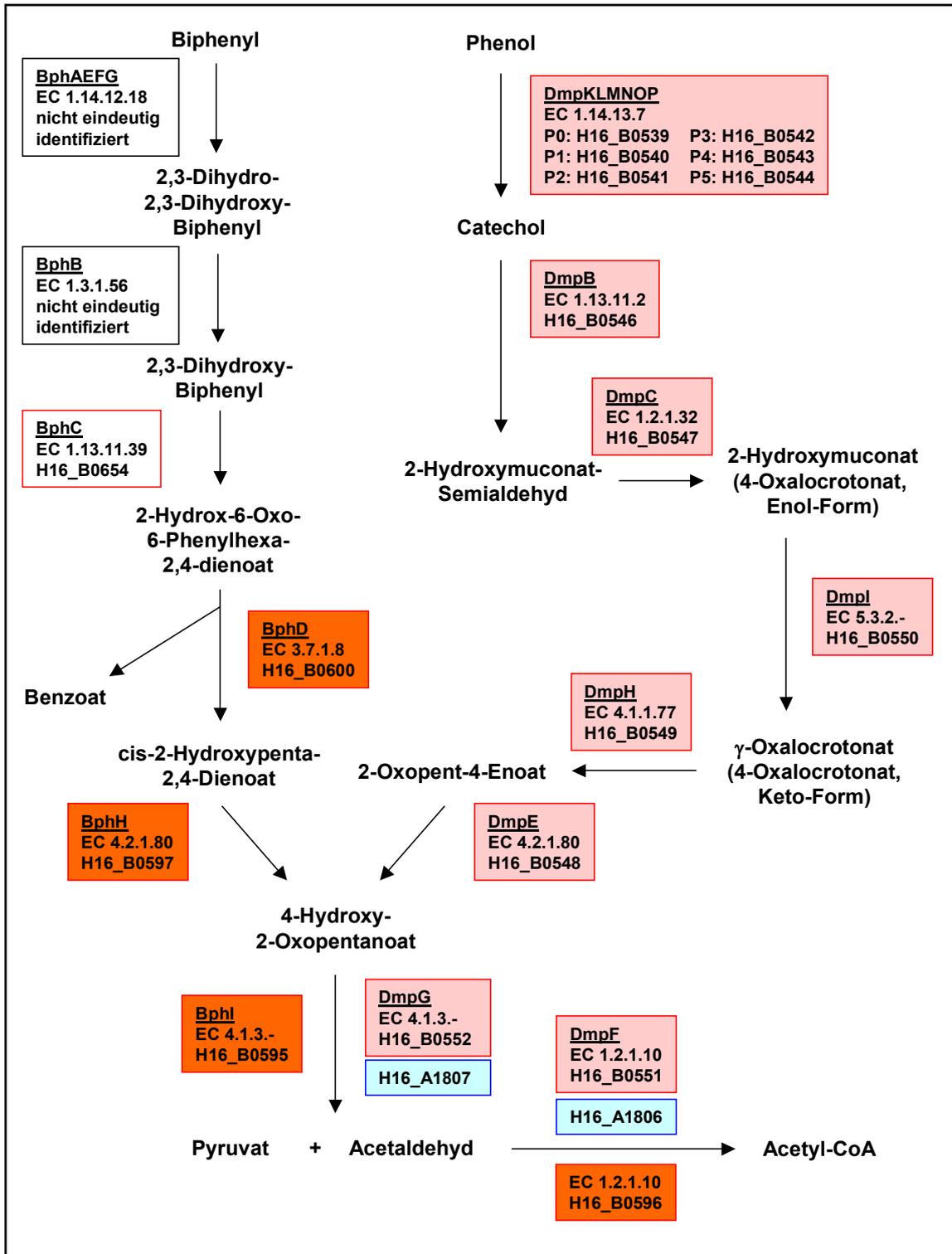


Abb. 3.19: Schematische Darstellung des Abbaus aromatische Kohlenwasserstoffe in *R. eutropha* über den *meta*-Spaltungs-Weg nach KEGG. Blau umrandet, Proteine, die auf Chr.1 kodiert sind; rot umrandet, Proteine, die auf Chr.2 kodiert sind; schwarz umrandet, Proteine, die nicht eindeutig identifiziert werden konnten. CDSs zu Proteinen gleicher Hintergrundfarbe sind in einem gemeinsamen Cluster lokalisiert. Die Bezeichnungen der Proteine beziehen sich auf Shingler *et al.* (1992) (Dmp-Proteine) bzw. Hofer *et al.* (1994) (Bph-Proteine).

Diejenigen Cluster, die auf Basis der Sequenzdaten nicht zuverlässig annotiert werden können bzw. unvollständig vorliegen, kodieren möglicherweise für Enzyme, die den Abbau aromatischer Kohlenwasserstoffe über Gentisat und Homogentisat und den Abbau chlorierter Aromaten über Chlorocatechol katalysieren (Tab. 3.13).

Tab. 3.13: Abbau aromatischer Kohlenwasserstoffe über weitere Wege, CDSs auf Chr.2.

Nr.	CDS	Annotation	EC-Nummer
1. Abbau über Homogentisat			
1	H16_B1671	transcriptional regulator, LysR family	1.13.11.5
2	H16_B1670	fumarylacetoacetase	3.7.1.2
2. Abbau über Gentisat			
1	H16_B0872	transcriptional regulator, LysR-family	
2	H16_B0873	gentisate 1,2-dioxygenase	1.13.11.4
3	H16_B0874	fumarylpyruvate hydrolase	3.7.1.5
4	H16_B0875	maleylpyruvate isomerase	5.2.1.4
5	H16_B0876	3-Hydroxybenzoate 6-hydroxylase	1.14.13.24
6	H16_B0877	4-hydroxybenzoate transporter	
3. Potntiell am Abbau von Chlorocatechol beteiligt			
1	H16_B0967	transcriptional regulator, LysR-family	
2	H16_B0968	6-chlorohydroxyquinol-1,2-dioxygenase	1.13.11.
3	H16_B0969	conserved hypothetical protein	
4	H16_B0970	maleylacetate reductase	1.3.1.32
5	H16_B0971	FAD-dependent monooxygenase	1.14.13.1-
6	H16_B0971	putative dipeptidylaminopeptidase	

3.1.4.2.1.4 Acetoin-Abbau

R. eutropha wächst auf Acetoin als einziger Kohlenstoffquelle (Steinbüchel *et al.*, 1987). Biochemische Hintergründe des Acetoin-Stoffwechsels in *R. eutropha* wurden bereits gut untersucht und veröffentlicht (Frund *et al.*, 1989; Priefert *et al.*, 1991). Die genetischen Informationen für die Umsetzung von Acetoin sind in *R. eutropha* in einem einzigen Operon kodiert (Priefert *et al.*, 1991), das sich auf Chr.2 befindet (*acoXABC*; Tab. 3.14). In *R. solanacearum* konnte kein entsprechendes Operon lokalisiert werden.

Tab. 3.14: Acetoin-Katabolismus, *acoXABC*-Operon und Regulator *acoR* auf Chr.2.

Nr.	CDS	Gen	Annotation	EC-Nummer
1	H16_B0142	<i>acoR</i>	acetoin catabolism regulatory protein R	3.6.1.-
2	H16_B0143	<i>acoX</i>	acetoin catabolism protein X	2.7.1.23
3	H16_B0144	<i>acoA</i>	acetoin dehydrogenase E1 component alpha-subunit	1.2.4.-
4	H16_B0145	<i>acoB</i>	acetoin dehydrogenase E1 component beta-subunit	1.2.4.1
5	H16_B0146	<i>acoC</i>	acetoin dehydrogenase dihydrolipoamide S-acetyltransferase component	2.3.1.12

3.1.4.2.2 Erschließung alternativer Stickstoff-Quellen

3.1.4.2.2.1 Cyanat-Hydratase

Das *cynTSX*-Operon ermöglicht es *Escherichia coli* K12, auf Cyanat als einziger Stickstoffquelle zu wachsen (Sung & Fuchs, 1988). Es kodiert für eine Cyanat-Hydratase (*cynS*), eine Carboanhydrase (*cynT*) und ein Protein unbekannter Funktion (*cynX*). Durch die Cyanat-Hydratase wird Cyanat mit Bicarbonat zu Ammonium und Kohlendioxid umgesetzt. Für das Wachstum auf Cyanat scheint die Carboanhydrase zur Re-Fixierung des freigesetzten Kohlendioxids eine wichtige Rolle zu spielen, da *cynT*-Mutanten in ihrem Wachstum auf Cyanat stark gehemmt werden (Guillotot *et al.*, 1993).

Ein Gen auf Chr.2 (H16_B0046) zeigt 41 % Identität auf Proteinebene zu *CynS* aus *E. coli* K12 (Blattner *et al.*, 1997) und 74 % Identität zu einem C-terminalen Fragment einer Cyanat-Hydratase (tr|Q8XPL0), das in *R. solanacearum* auf dem Megaplasmid lokalisiert ist. Die Cyanat-Hydratase aus *R. eutropha* ist in einem Cluster mit einem Regulator (H16_B0049), einer potentiellen Nucleosid-Deaminase (H16_B0048) – also einem weiteren Ammonium freisetzenden Enzym – und einem *cynX*-analogen Gen (H16_B0047) kodiert. Eine β -Carboanhydrase mit großer Ähnlichkeit zu *CynT* ist ebenfalls auf Chr.2 lokalisiert (H16_B2270), jedoch in einem ca. 400 kbp entfernten Operon, das unter anderem für die respiratorische Nitrat- und Nitrit-Reduktase auf Chr.2 kodiert.

3.1.4.2.2.2 Formamidase

Das Wachstum auf Formamid als einziger Stickstoffquelle wird in *Aspergillus nidulans* durch die Formamidase vermittelt, ein Enzym, das die hochspezifische Hydrolyse von Formamid zu Ammonium und Formiat katalysiert (Hynes, 1975). In Form einer Metabolit-Repression wird das Gen für die Formamidase (*fmdS*) in Abhängigkeit von der Verfügbarkeit anderer Stickstoffquellen wie Ammonium und Glutamin reguliert (Marzluf, 1997).

Auch für *R. eutropha* wurde die Fähigkeit nachgewiesen, Formamid als Stickstoffquelle über die Aktivität einer Formamidase zu nutzen (Friedrich & Mitrenga, 1981). Eine potentielle Formamidase (H16_B0072) ist in einem Cluster auf Chr.2 kodiert und weist eine Identität auf Proteinebene von 53 % zu der in *A. nidulans* charakterisierten FmdS (gi|12620135; Fraser *et al.*, 2001) und 71 % zu einem in *Methylophilus methylotrophus* charakterisierten Enzym (gi|3913682; Wyborn *et al.*, 1994) auf. Zu diesem Cluster gehören außerdem mehrere ABC-Transportsysteme (H16_B0073-0077) sowie ein Zwei-Komponenten-Regulationssystem aus Sensor-Histidin-Kinase (H16_B0078) und *Response Regulator* (H16_B0079).

3.1.4.2.2.3 Ethanolamin-Ammonium-Lyase

Salmonella typhimurium wächst auf Ethanolamin als einziger Stickstoff- und Kohlenstoffquelle (Chang & Chang, 1975). Das Substrat wird durch die Ethanolamin-Ammonium-Lyase, ein Vitamin B12-abhängiges Enzym, zu Ammonium und Acetaldehyd umgesetzt und durch die Acetaldehyd-Dehydrogenase (EC 1.2.1.10) weiter zu Acetyl-Coenzym A abgebaut (Jones & Turner, 1984). Der Adenosyl-Cobalamin-Cofaktor (Vitamin B12), wird in *R. eutropha* wahrscheinlich aus Zwischenprodukten synthetisiert, die aus dem Medium aufgenommen werden. Die dafür notwendigen Gene sind auf Chr.1 lokalisiert.

Eine potentielle Ethanolamin-Ammonium-Lyase aus *R. eutropha* ist auf Chr.2 in einem Cluster aus drei Genen kodiert. Dieser Cluster setzt sich aus Genen für die große und kleine Untereinheit des Enzyms (H16_B0096-0097) und für ein Membranprotein zusammen (H16_B0095), bei dem es sich möglicherweise um eine Permease handelt. Große und kleine Untereinheit der Ethanolamin-Ammonium-Lyase zeigen 51 % und 36 % Sequenzidentität auf Proteinebene mit den in *S. typhimurium* LT2 charakterisierten Proteinen (sp|P19265, sp|P19264; Faust *et al.*, 1990) und noch stärkere Übereinstimmung zu den Untereinheiten aus *Pseudomonas syringae* DC3000 (tr|Q889M2, sp|Q889M3) und *R. solanacearum* (tr|Q8XUQ9, tr|Q8XUR0).

3.1.4.2.3 Alternative Phosphorquellen

Eine Reihe Gram-positiver und Gram-negativer Bakterien ist in der Lage, Phosphonate abzubauen und als einzige Phosphatquelle zum Wachstum zu nutzen (Harkness, 1966). Phosphonate bezeichnen organische Verbindungen von hoher chemischer Stabilität, die sich durch direkte, kovalente Bindungen zwischen Phosphor- und Kohlenstoffatomen auszeichnen. Phosphonate werden in zunehmendem Maße durch industrielle Prozesse freigesetzt, beispielsweise über Abwässer der Textil-Industrie (Nowack, 1998) und über das Herbizid Glyphosat (*Round up*, Fa. Monsanto, USA), bei dem es sich um eines der meistverbreiteten Herbizide weltweit handelt (Aspelin, 1997). Genauere Daten über Phosphonat-Konzentrationen in der Umwelt liegen bisher nicht vor; aufgrund ihrer chemischen Stabilität findet unter natürlichen Bedingungen aber praktisch kein biologischer Abbau dieser Xenobiotika statt (Nowack, 2003). Deshalb besteht Interesse an den genetischen Grundlagen des Phosphonat-Abbaus, der in verschiedenen Organismen untersucht wurde (Chen *et al.*, 1990; White & Metcalf, 2004).

In *R. eutropha* sind in einem Cluster auf Chr.2 acht Gene als „phosphonate metabolism protein“ annotiert: *phnFGHIJKLM* (H16_B1289-1282). Dieser Cluster weist starke Übereinstimmungen in Sequenz und Reihenfolge der Gene zu einem *phn*-Operon auf dem Megaplasmid pSymB aus *Sinorhizobium meliloti* auf (Finan *et al.*, 2001). In *R. solanacearum* wurde kein *phn*-Operon identifiziert.

3.1.4.3 Synthese von Pyrroloquinolinquinon (PQQ)

Pyrroloquinolinquinon (PQQ) wurde zunächst als Cofaktor einer bakteriellen Alkohol-Dehydrogenase beschrieben (Salisbury *et al.*, 1979). PQQ konnte in menschlicher Muttermilch (Mitchell *et al.*, 1999) und als wichtiger Nahrungsbestandteil in Mäusen nachgewiesen werden (Killgore *et al.*, 1989). Da es darüberhinaus Hinweise auf eine Rolle von PQQ als Cofaktor im enzymatischen Abbau von Lysin in Säugetieren gibt, wird eine Vitaminfunktion von PQQ auch für höhere Organismen diskutiert (Kasahara & Kato, 2003).

In *Klebsiella pneumoniae* wurden sechs Gene identifiziert, die an der Synthese von PQQ beteiligt sind (Meulenberg *et al.*, 1992); die Expression von vier Genen (*pqqACDE*) wurde dagegen als hinreichend für die PQQ-Synthese in *Acinetobacter calcoaceticus* gezeigt. Davon ausgehend wurde spekuliert, daß in *K. pneumoniae* möglicherweise ein anderes Substrat das Ausgangsprodukt der PQQ-Synthese bildet (Goosen *et al.*, 1989).

Auf Chr.2 in *R. eutropha* ist ein Cluster aus vier Genen kodiert (H16_B1052-1049), der starke Übereinstimmungen in der Sequenz und der Organisation zu den *pqqBCDE*-Operons aus *Pseudomonas syringae* (Buell *et al.*, 2003), *R. eutropha* JMP134 und *Burkholderia xenovorans* LB400 (beide DOE Joint Genome Institute) zeigt. Das Gen *pqqA*, das in *R. eutropha* und anderen Organismen fehlt, kodiert für ein kurzes Peptid von 23 Aminosäuren, das wahrscheinlich als Vorläufer für die PQQ-Biosynthese dient (Goosen *et al.*, 1992). Über die Funktion der potentiellen Endopeptidase PqqF aus *Klebsiella pneumoniae* (Meulenberg *et al.*, 1992), dem zweiten fehlenden Gen in *R. eutropha*, gibt es keine weiteren Informationen. Stromabwärts vom *pqq*-Operon, auf dem gegenüberliegenden Strang, liegt ein Cluster aus drei CDSs, der für ein potentielles Cytochrom c (H16_B1048), eine potentielle PQQ-abhängige Alkohol-Dehydrogenase (H16_B1047) und eine Zink-abhängige Metallhydrolase oder Glyoxylase (H16_B1046) kodiert.

3.1.4.4 Kdp-ATPase

Kalium-Ionen spielen eine wichtige Rolle bei der Osmoregulation von Bakterien. Aus diesem Grund besitzen Bakterien eine Reihe von Influx- und Efflux-Systemen zur Kontrolle der K^+ -Konzentration (Stumpe *et al.*, 1996). In *Escherichia coli* wird das *kdpFABC*-Operon unter K^+ -limitierenden Bedingungen und bei hoher Osmolarität des Mediums induziert (Rhoads *et al.*, 1976). Es wird als eine Art Notfallsystem angesehen, das eine K^+ -Versorgung garantiert, wenn eine ausreichende Aufnahme von Kalium-Ionen durch andere Systeme nicht gewährleistet werden kann (Epstein, 2003). Die Expression des *kdpFABC*-Operons wird durch ein Zwei-Komponenten-Regulationssystem kontrolliert (*kdpDE*), welches direkt stromabwärts identifiziert wurde (Walderhaug *et al.*, 1992).

Das *kdpABC*-Operon kodiert für einen Enzymkomplex aus der Familie der P-Typ-ATPasen, deren Mitglieder für den Membrantransport vieler Ionen verantwortlich sind (Okamura *et al.*, 2003). *kdpF* kodiert für ein kurzes Peptid, das wahrscheinlich eine stabilisierende Funktion auf den Enzymkomplex ausübt (Gassel *et al.*, 1999).

In *R. eutropha* konnte ein *kdpABCDE*-Cluster auf Chr.2 identifiziert werden (H16_B0898-0894), das einen hohen Grad an Konservierung in Sequenz und Organisation innerhalb der Bakterien aufweist. Stromaufwärts von *kdpA* ist in *R. eutropha* ein Gen lokalisiert, das zwar keine Sequenzähnlichkeit zu *kdpF* aus *E. coli* aufweist, aber ebenfalls für ein potentiell kurzes und hydrophobes Peptid kodiert. Ähnliche CDSs, die keine Übereinstimmung auf der Ebene der Proteinsequenz besitzen, aber für kurze hypothetische Peptide von 30 bis 90 Aminosäuren Länge kodieren, weisen auch andere *kdp*-Operons auf.

3.1.4.5 Begeißelung

Aufgrund von Unterschieden im Begeißelungstyp wurde ein Teil der *Ralstonia*-Spezies, zu dem auch *R. eutropha* und *R. metallidurans* gehören, in dem neuen Genus *Wautersia* zusammengefaßt (Vanechoutte *et al.*, 2004). Während *R. solanacearum* in unterschiedlichen Quellen als unbeweglich (Coenye *et al.*, 1999; De Baere *et al.*, 2001) oder mit ein bis vier Flagellen polar begeißelt beschrieben wurde (Krieg & Holt, 1984; Tans-Kersten *et al.*, 2001), besitzt *R. eutropha* eine peritriche Begeißelung (Davis *et al.*, 1969; Vanechoutte *et al.*, 2004). Innerhalb der Spezies *Aeromonas hydrophila* wurden unterschiedliche Stämme charakterisiert, die sich ähnlich wie *R. eutropha* und *R. solanacearum* durch unterschiedliche Begeißelungstypen (polar und lateral) unterscheiden. In Mutationsversuchen konnte gezeigt werden, daß die Deletion eines Gens (*flgN*) die Bildung lateraler Flagellen reduziert, aber keinen Einfluß auf die allgemeine Fähigkeit zur Fortbewegung durch polare Geißeln besitzt (Altarriba *et al.*, 2003).

Tab. 3.15: Flagellarapparat, verschiedene Gen-Cluster auf Chr.2.

Nr.	CDS	Gen	Annotation
1	H16_B0252	<i>flhB</i>	flagellar biosynthetic protein FlhB
2	H16_B0253	<i>flhA</i>	flagellar biosynthetic protein FlhA
3	H16_B0254	<i>flhF</i>	flagellar biosynthetic protein FlhF
4	H16_B0255	<i>motR</i>	flagellar biosynthesis regulator
5	H16_B0256	<i>fliA</i>	RNA polymerase sigma factor for flagellar operon (Sigma-28)
6	H16_B0257	<i>flhE</i>	flagellar protein FlhE precursor
7	H16_B0258	<i>flgN</i>	flagellar synthesis protein FlgN
8	H16_B0259	<i>flgM</i>	negative regulator of flagellin synthesis, anti-sigma factor
9	H16_B0260	<i>flgA</i>	flagellar basal-body P-ring formation protein FlgA precursor
10	H16_B0261	<i>flgB</i>	flagellar basal-body rod protein FlgB
11	H16_B0262	<i>flgC</i>	flagellar basal-body rod protein FlgC
12	H16_B0263	<i>flgD</i>	basal-body rod modification protein FlgD
13	H16_B0264	<i>flgE</i>	flagellar hook protein FlgE
14	H16_B0265	<i>flgF</i>	flagellar basal-body rod protein FlgF
15	H16_B0266	<i>flgG</i>	flagellar basal-body rod protein FlgG
16	H16_B0267	<i>flgH</i>	flagellar L-ring protein precursor FlgH
17	H16_B0268	<i>flgI</i>	flagellar P-ring protein precursor FlgI
18	H16_B0269	<i>flgJ</i>	peptidoglycan hydrolase (Muramidase FlgJ)
19	H16_B0270	<i>flgK</i>	flagellar hook-associated protein 1
20	H16_B0271	<i>flgL</i>	flagellar hook-associated protein 3
1	H16_B0567	<i>fliL</i>	flagellar protein FliL
2	H16_B0566	<i>fliM</i>	flagellar motor switch protein FliM
3	H16_B0565	<i>fliN</i>	flagellar motor switch protein FliN
4	H16_B0564	<i>fliO</i>	flagellar protein FliO
5	H16_B0563	<i>fliP</i>	flagellar protein FliP
6	H16_B0562	<i>fliQ</i>	flagellar biosynthetic protein FliQ
7	H16_B0561	<i>fliR</i>	flagellar biosynthetic protein FliR
1	H16_B2359		conserved hypothetical protein
2	H16_B2360	<i>fliC</i>	flagellin FliC
3	H16_B2361	<i>flaG</i>	flagellar protein FlaG
4	H16_B2362	<i>fliD</i>	flagellar hook-associated protein 2 (flagellar cap protein)
5	H16_B2363	<i>fliS</i>	flagellar protein FliS
6	H16_B2364	<i>fliT</i>	flagellar protein FliT
7	H16_B2365	<i>fliK1</i>	flagellar hook-length control protein FliK
8	H16_B2366	<i>flhB2</i>	cytoplasmic domain of flagellar biosynthetic protein FlhB
9	H16_B2367	<i>fliE</i>	flagellar hook-basal body complex protein FliE
10	H16_B2368	<i>fliF</i>	flagellar M-ring protein FliF
11	H16_B2369	<i>fliG</i>	flagellar motor switch protein FliG
12	H16_B2370	<i>fliH</i>	flagellar assembly protein FliH
13	H16_B2371	<i>fliI</i>	flagellum-specific ATP synthase (EC 3.6.3.14)
14	H16_B2372	<i>fliJ</i>	flagellar protein FliJ
15	H16_B2373	<i>fliK2</i>	putative flagellar hook-length control protein FliK

In *R. eutropha* ist der Flagellarapparat im wesentlichen in den *flg*-, *flh*- und *fli*-Operons kodiert, die ausnahmslos auf Chr.2 lokalisiert sind (Tab. 3.15). Zu jedem dieser Operons gibt es Entsprechungen auf dem Megaplasmid von *R. solanacearum*, die allerdings jeweils leichte Unterschiede in ihrer Anordnung aufweisen. Auffallende Abwei-

chungen zwischen beiden Spezies konnten nur innerhalb des *flg*-Operons und dort besonders in Bezug auf *flgN* festgestellt werden: Zwei Gene liegen in *R. solanacearum* fragmentarisch vor (*flgAM*; gi|17548562, gi|17548561), eines weist nur sehr geringe Ähnlichkeit zu bekannten, homologen Genen auf (*flgN*; gi|17548560). FlgM bindet als negativer Regulator den Flagellum-spezifischen Sigma-Faktor σ_{28} und kontrolliert so die Transkription der Flagellin-Gene (Kutsukake *et al.*, 1994). Bei FlgA handelt es sich wahrscheinlich ebenso wie bei FlgN um ein periplasmatisches Chaperon (Fraser *et al.*, 1999; Nambu & Kutsukake, 2000). FlgN aktiviert außerdem die Translation von *flgM* (Karlinsky *et al.*, 2000). Möglicherweise gehen phänotypische Unterschiede in der Form der Begeißelung zwischen *R. eutropha* und *R. solanacearum* im wesentlichen auf ein einziges fehlendes Gen – *flgN* – in *R. solanacearum* zurück.

3.1.4.6 Synthese von Sekundärmetaboliten

Bei der Annotation von Chr.2 aus *R. eutropha* konnten im Gegensatz zu Chr.1 einige auffallend lange CDSs identifiziert werden, die überwiegend in Clustern organisiert vorliegen und starke Ähnlichkeit zu Genen aus der Gruppe der pathogenen Proteobakterien aufweisen. Während eine genaue Annotation dieser Cluster aufgrund teilweise sehr schwacher Konservierung der einzelnen CDSs erschwert wird, deuten Ähnlichkeiten einzelner Gene auf Funktionen für den Sekundärmetabolismus hin, die beispielsweise in der Biosynthese von Antibiotika und Toxinen liegen könnten.

3.1.4.6.1 Nicht-ribosomale Peptid-Synthetasen (NRPSs)

Nicht-ribosomale Peptid-Synthetasen (NRPSs) versetzen Mikroorganismen in die Lage, strukturell vielfältige Sekundärmetabolite zu synthetisieren und dabei auch auf nicht-proteinogene Aminosäuren zurückzugreifen. Nicht durch Ribosomen synthetisierte Peptide können cyclische Peptid-Grundgerüste enthalten und sind als Antibiotika, Immunsuppressiva und Anti-Tumor-Reagenzien von großem pharmazeutischen Interesse (Schwarzer & Marahiel, 2001). NRPSs werden modulartig aus repetitiven katalytischen Einheiten aufgebaut, die innerhalb einer CDS oder auf mehrere Gene verteilt vorliegen können (Guenzi *et al.*, 1998). Die Reihenfolge der katalytischen Einheiten hat dabei

entscheidenden Einfluß auf die Spezifität und Reihenfolge der verknüpften Peptide (Stein *et al.*, 1996).

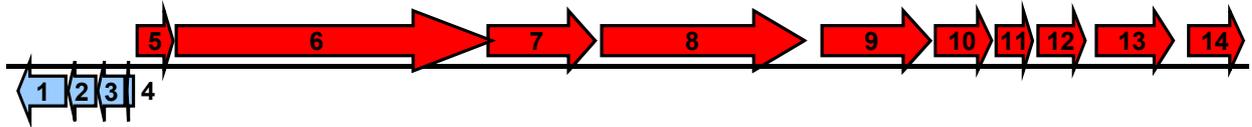


Abb. 3.20: Cluster einer Nicht-ribosomalen Peptid-Synthetase auf Chr.2. Die Nummerierung der CDSs bezieht sich auf Tab. 3.16.

Tab. 3.16: Zusammenhängender Gen-Cluster für eine Nicht-ribosomale Peptid-Synthetase auf Chr.2. Ähnlichkeiten nach BLAST-Vergleich gegen NCBI.

Nr.	CDS	Länge [AA]	Pathway [P], Annotation [A], Domänen [D]	Ähnlichkeit [NCBI-BLAST]
1	H16_B1692	474	A: putative aminotransferase (EC 2.6.1.-)	<i>Pseudomonas aeruginosa</i> gi 11351184
2	H16_B1691	277	D: COG2091, Sfp, phosphopantetheinyl transferase	<i>Gloeobacter violaceus</i> gi 35212517
3	H16_B1690	272	D: COG3208, GrsT, predicted thioesterase involved in non-ribosomal peptide biosynthesis.	<i>Azotobacter vinelandii</i> gi 23106240
4	H16_B1689	68	D: pfam03621, MbtH, found in the MbtH protein, as well as in antibiotic synthesis protein NIKP1.	<i>Pseudomonas syringae</i> tr Q884F6
5	H16_B1688	353	D: Fragment von COG2175, TauD, probable taurine catabolism dioxygenase.	<i>Pseudomonas putida</i> gi 24985933
6	H16_B1687	3094	D: COG1020, EntF, non-ribosomal peptide synthetase modules and related proteins.	<i>Erwinia carotovora</i> tr Q6D738
7	H16_B1686	1031	D: COG1020, EntF, non-ribosomal peptide synthetase modules and related proteins.	<i>Pseudomonas fluorescens</i> gi 48731888
8	H16_B1685	2008	D: COG1020, EntF, non-ribosomal peptide synthetase modules and related proteins.	<i>Pseudomonas putida</i> gi 82736876
9	H16_B1683	1075	D: COG1020, EntF, non-ribosomal peptide synthetase modules and related proteins.	<i>Nostoc sp.</i> gi 6563399
10	H16_B1682	560	D: COG4615, PvdE, ABC-type siderophore export system, fused ATPase and permease components	<i>Salmonella enterica</i> gi 29141115
11	H16_B1681	366	D: COG1670, RimL, acetyltransferases, including N-acetylases of ribosomal proteins	<i>Rhizobium leguminosarum</i> gi 15072886
12	H16_B1680	466	P: siderophore biosynthesis A: L-ornithine 5-monooxygenase (EC 1.13.12.-)	<i>Pseudomonas fluorescens</i> gi 23062336
13	H16_B1679	760	D: COG4773, FhuE, outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	<i>Pseudomonas aeruginosa</i> gi 15599417
14	H16_B1677	543	D: pfam03929, DUF337, uncharacterized iron-regulated membrane protein	<i>Bordetella bronchiseptica</i> gi 33602579

In *R. eutropha* ist eine potentielle NRPS in einem ca. 36 kbp großen Operon aus 14 Genen auf Chr.2 kodiert (Tab. 3.16; Abb. 3.20). Die einzelnen Gene dieses Operons zeigen unterschiedlich starke Ähnlichkeit zu Datenbankeinträgen für Untereinheiten von NRPSs aus verschiedenen Organismen, insbesondere aus *Pseudomonas*-Spezies. Ausschließlich auf Basis von Sequenzvergleichen ist eine Bestimmung der Spezifität einer NRPS zur Zeit noch nicht möglich. Auffallend ist aber die Ähnlichkeit einzelner CDSs zu Genen, die im Zusammenhang mit der Biosynthese von Siderophoren charakterisiert wurden:

1. H16_B1692 zeigt Ähnlichkeit zu einem Gen auf dem Megaplasmid pSymA aus *Sinorhizobium meliloti*, das im Zusammenhang mit der Synthese von Rhizobactin 1021 charakterisiert wurde (Lynch *et al.*, 2001).
2. H16_B1679 weist auf Proteinebene Übereinstimmungen zu einem spezifischen Rezeptor für das Siderophor Pseudobactin m114 aus *Pseudomonas sp.* M114 auf (Morris *et al.*, 1994).

3.1.4.6.2 Potentielle Toxine

Auf Chr.2 von *R. eutropha* sind sechs weitere Cluster lokalisiert, deren Annotation auf eine mögliche toxische Funktion hinweist. Fünf Cluster, die in Zusammensetzung und Länge der enthaltenen Gene variieren, lassen sich der porenbildenden Familie der RTX-Toxine zuordnen (siehe 3.1.4.6.2.1). Zu dieser Familie gehören unter anderem das RTX-Toxin aus *Vibrio cholerae* (Lin *et al.*, 1999) und das Hämolyysin aus *Escherichia coli* (Felmlee *et al.*, 1985). Ein weiterer Cluster, zu dem die längste überhaupt identifizierte, potentiell kodierende Sequenz in *R. eutropha* gehört, umfaßt drei Gene aus der *toxin complex* (tc) - Familie (siehe 3.1.4.6.2.2). Gene dieser Familie kodieren für insektizidale Toxine und wurden in *Photobacterium luminescens* charakterisiert (Bowen *et al.*, 1998).

3.1.4.6.2.1 RTX-Toxine

Exotoxine aus der Familie der RTX-Toxine [*repeats-in-toxin*] spielen eine wichtige Rolle bei der Virulenz vieler menschlicher und tierischer Gram-negativer Pathogene (Welch *et al.*, 1992; Welch *et al.*, 1995). Von diesen Organismen werden sie durch spe-

zifische Sekretionssysteme abgegeben und bilden Poren in eukaryotischen Zellen, die auf diesem Weg lysiert werden. Bei dem bestuntersuchten RTX-Toxin handelt es sich um ein Hämolysin aus *E. coli* (Hly), das mit extraintestinalen Infektionen in Verbindung gebracht wird (Felmlee *et al.*, 1985). Gemeinsamkeiten aller RTX-Toxine sind die post-translationale Reifung, eine C-terminale Ca²⁺-Bindedomäne, die sich aus Glycinreichen, sauren Nonapeptid-Wiederholungen zusammensetzt und ein Export via Typ I Sekretionssystem (TISS) (Welch, 2001). Typischerweise sind RTX-Operons in der Form *rtxCABD* organisiert, in der *rtxC* für das Aktivatorprotein, *rtxA* für das Toxin selbst und *rtxBD* für Komponenten des Sekretionssystems kodieren (Felmlee *et al.*, 1985; Glaser *et al.*, 1988). Eine weitere Komponente dieses spezifischen Sekretionssystems ist in *tolC* kodiert, das jedoch in den meisten Fällen nicht im gleichen Operon lokalisiert ist wie die *rtx*-Gene (Wandersman & Delepelaire, 1990). Im übrigen besitzt nicht jedes Toxin ein eigenes Sekretionssystem, sondern nutzt gegebenenfalls Komponenten, die in anderen Operons kodiert sind (Chang *et al.*, 1991).

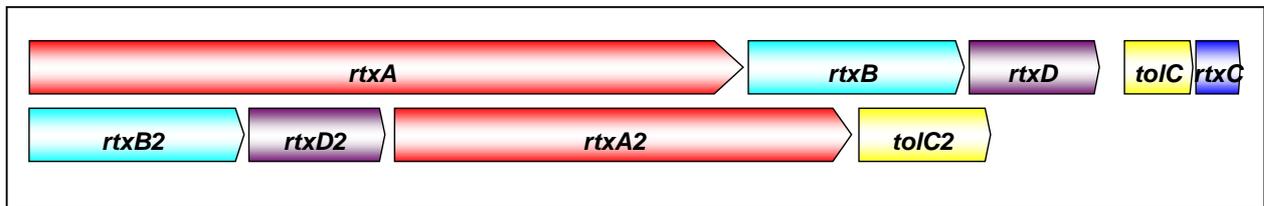


Abb. 3.21: Verschiedene potentielle Gen-Cluster aus der RTX-Familie auf Chr.2. Nummerierungen der Gene verweisen auf Tab. 3.17.

Das Genom von *R. eutropha* enthält zwei potentielle Cluster, die für Proteine aus der Familie der RTX-Toxine kodieren. Darunter befindet sich jedoch wahrscheinlich kein funktionelles *rtx*-Operon, da kein intaktes Gen für das Aktivatorprotein *rtxC* identifiziert werden konnte (Abb. 3.21; Tab. 3.17). Das einzige vorhandene *rtxC*-Analog (H16_B2309, Cluster 1) enthält einen Frameshift und liegt deshalb als unvollständiges Fragment vor. Von zwei potentiellen *rtxA*-Genen, weist nur H16_B0692 (Cluster 2) die für RTX-Toxine typischen Nonapeptid-Wiederholungen auf, mittels derer RTX-Toxine im allgemeinen charakterisiert werden (Welch *et al.*, 1995). Hinsichtlich der Reihenfolge der Gene weichen beide *rtx*-Operons sowohl voneinander als auch von der üblicherweise beschriebenen Reihenfolge der *rtx*-Gene (*rtxCABD*) ab. Interessanterweise sind in *R. eutropha* alle identifizierten *tolC*-Analoge innerhalb der

rtx-Operons kodiert – eine Organisationsform, die bisher nur für *Bordetella pertussis* beschrieben wurde (Laoide & Ullmann, 1990).

Tab. 3.17: Potentielle RTX-Toxine, verschiedene Gen-Cluster auf Chr.2.

Nr.	CDS	Länge [AA]	Gen	Annotation	Besonderheit
Cluster 1:					
1	H16_B2313	2426	<i>rtxA</i>	RTX toxin exported protein	
2	H16_B2312	731	<i>rtxB</i>	ABC-type RTX toxin transporter, ATPase and permease components	
3	H16_B2311	445	<i>rtxD</i>	RTX toxin membrane fusion protein	Frameshift
4	H16_B2310	320	<i>tolC</i>	outer membrane protein TolC precursor	Frameshift
5	H16_B2309	150	<i>rtxC</i>	RTX toxin activating acyltransferase	Frameshift
Cluster 2:					
6	H16_B0690	730	<i>rtxB2</i>	ABC-type RTX toxin transporter, ATPase and permease components	
7	H16_B0691	462	<i>rtxD2</i>	RTX toxin membrane fusion protein	
8	H16_B0692	1552	<i>rtxA2</i>	RTX toxin exported protein	
9	H16_B0693	450	<i>tolC2</i>	outer membrane protein TolC precursor	

3.1.4.6.2.2 Tc-Toxine

Die *toxin complex* (*tc*)-Gene aus *Photobacterium luminescens* kodieren für hochmolekulare Toxine mit oralem Wirkmechanismus gegen Insekten (Bowen *et al.*, 1998). Durch Expression von *tcdA* aus *Photobacterium luminescens* ssp. *akhurstii* W14 konnte transgenen *Arabidopsis*-Pflanzen eine Resistenz gegen Raupen des Modellinsekts *Manduca sexta* vermittelt werden (Liu *et al.*, 2003). In *E. coli* konnte die Toxizität von TcdA durch Koexpression von *tcdA1* mit *tcdB1* und *tccC1* gesteigert werden. Diese sogenannten Potentioren verleihen außerdem in *E. coli* einem weiteren Tc-Protein Toxizität, das sonst keine eigene toxische Wirkung zeigt (TcaAB; Waterfield *et al.*, 2001).

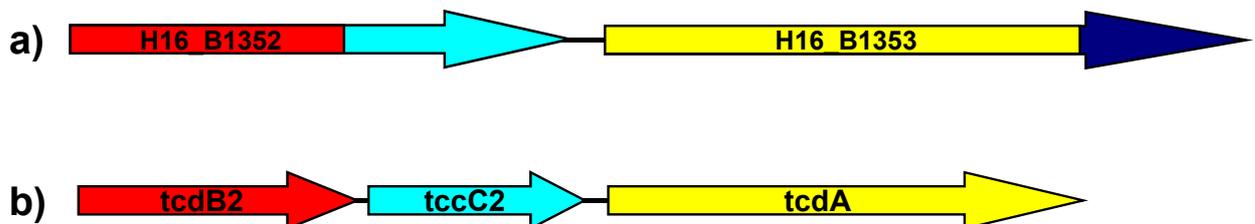


Abb. 3.22: Tc-Toxin-Cluster auf Chr.2. a) Cluster aus drei Genen von Chr.2. b) Cluster aus *Photobacterium luminescens* (Waterfield *et al.*, 2001). Übereinstimmende Sequenzbereiche sind durch gleiche Farben gekennzeichnet.

Zu den Genen für das Tc-Toxin TcdA1 und die Potentioren TcdB1 und TccC1 konnten analoge CDSs in einem Cluster auf Chr.2 identifiziert werden (Abb. 3.22). Im Gegensatz zu *Photobacterium luminescens* sind TcdB und TccC in *R. eutropha* aber innerhalb einer einzigen CDS kodiert (H16_B1352), die N-terminale Ähnlichkeit zu *tcdB* und C-terminale Ähnlichkeit zu *tccC* aufweist. Daß es sich bei dem in *R. eutropha* identifizierten Gen aller Wahrscheinlichkeit nach um eine echte kodierende Sequenz handelt, wird durch die Existenz einer CDS in *Pseudomonas syringae* CD3000 bestätigt, die zwar eine geringe Identität auf Proteinebene (29 %), aber eine ähnliche Länge wie H16_B1352 zeigt (tr|Q87X46; Buell *et al.*, 2003). Bei *tcdA* (H16_B1353), dem zweiten potentiellen *tc*-Gen aus *R. eutropha*, handelt es sich um die größte in *R. eutropha* identifizierte CDS überhaupt (3395 AA). Dieses Gen weicht in Sequenz und Länge insbesondere des N-Terminus von allen bisher gefundenen *tc*-Genen ab. Obwohl inzwischen gezeigt wurde, daß *tc*-Gene unter Gram-negativen Bakterien weit verbreitet sind (Waterfield *et al.*, 2001), existieren in den Datenbanken keine Einträge anderer β -Proteobakterien mit signifikanter Ähnlichkeit zu den potentiellen *tc*-Genen aus *R. eutropha*.

3.II Das *Methanosphaera stadtmanae*-Genomprojekt

Der zweite Teil dieser Arbeit beschäftigt sich mit dem *Methanosphaera stadtmanae*-Genomprojekt. Methanogene Archaeen gehören einer Domäne des Lebens an, die sich sowohl auf physiologischer als auch auf genetischer Ebene von den Bakterien unterscheidet, zu denen *Ralstonia eutropha* H16 gehört. Trotzdem stellte die Bearbeitung des *Msp. stadtmanae*-Genomprojektes keine grundsätzlich veränderten Anforderungen, so daß – wie im folgenden gezeigt werden soll – die im Laufe des ersten Teils dieser Arbeit gewonnenen Erfahrungen für die Optimierung der Sequenzierung und Sequenzanalyse des *Msp. stadtmanae*-Genoms genutzt werden konnten. Der Vergleich der beiden Projekte zeigt zudem, daß ähnliche Methoden der Bearbeitung und Analyse eines Genoms auf völlig unterschiedliche Organismen angewandt werden können und dabei zu sehr charakteristischen Ergebnissen führen.

3.II.1 Sequenzierung und Annotation

3.II.1.1 Strategieanpassung

Für die Sequenzierung von *Msp. stadtmanae* wurde die Strategie aus dem *R. eutropha*-Genomprojekt in zwei wesentlichen Punkten verändert:

1. Das Projekt wurde auf eine Plasmid-Genbank mit kurzer Insertlänge beschränkt. Durch die geringe Größe des Genoms (1.8 Mbp), das sich zudem auf ein einziges Replikon beschränkt, konnte auf die strukturierende Funktion einer Cosmid-Genbank verzichtet werden.
2. Durch eine im Verhältnis zur Genomgröße erhöhte Anzahl an Sequenzläufen wurde die Sequenzabdeckung des Genoms schon in der Rohsequenzierungsphase gesteigert. Damit konnte einerseits die Contiganzahl gesenkt und andererseits die Qualität des Datensatzes verbessert werden, so daß der Zeitaufwand für Lückenschluß und Editierung der Consensus-Sequenz minimiert wurde.

3.II.1.2 Rohsequenzierung

Rechnerisch ergibt sich bei einer Genomgröße von 1.8 Mbp, einer angestrebten Sequenzabdeckung von sieben und einer durchschnittlichen Sequenzlänge von 700 bp ein Bedarf von 18 000 Sequenzläufen (siehe 3.I.1.2). Lander & Waterman (1988) haben eine Formel vorgeschlagen, die die Wahrscheinlichkeit (P) angibt, mit der eine beliebige Base des Genoms nicht sequenziert wurde – in Abhängigkeit von der Genomgröße (L), der Anzahl der Sequenzläufe (N) und der durchschnittlichen Sequenzlänge (w):

$$P = e^{-\frac{N \cdot w}{L}}$$

Daraus berechnet sich die Anzahl der Contigs in Abhängigkeit von der Anzahl der Sequenzläufe durch $P \cdot N$.

Aus dem Vergleich der Lander-Waterman-Vorhersage mit der tatsächlichen Anzahl von Contigs bzw. Sequenzlücken, die nach verschiedenen Zwischen-Assemblierungen erreicht wurden, konnte der Verlauf der Rohsequenzierung bewertet werden (Abb. 3.23). Die Anzahl der Sequenzlücken verringerte sich dabei zunächst deutlich langsamer als berechnet und erreichte erst ab ca. 18 000 Sequenzläufen ein Minimum. Die Ursache hierfür liegt wahrscheinlich darin begründet, daß es während einer Genbankherstellung niemals zu einer vollkommenen, statistischen Verteilung des Genoms auf die rekombinanten Plasmide kommt, sondern einzelne Bereiche des Genoms über- oder unterrepräsentiert werden. Zudem neigen homologe Sequenzabschnitte bzw. Sequenzläufe schlechter Qualität während der automatischen Rohassemblierung zur Bildung von Fehlassemblierungen. Bei einer Contiganzahl von 195, die sich aus insgesamt 20 186 assemblierten Sequenzläufen zusammensetzten, wurde die Rohsequenzierungsphase beendet und mit der Editierung der Consensus-Sequenz und dem Lückenschluß begonnen.

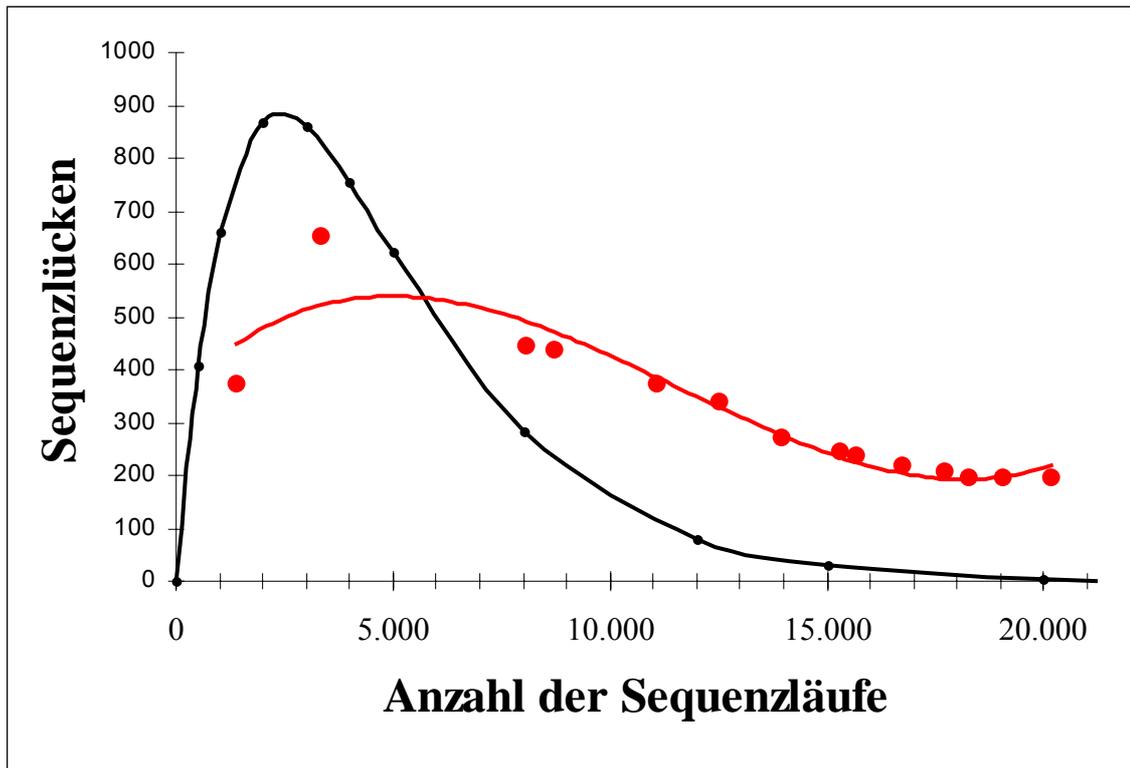


Abb. 3.23: Verlauf der Rohsequenzierungsphase und Vorhersage nach Lander & Waterman (1988). Schwarz Linie, Vorhersage der Sequenzlücken in Abhängigkeit von der Anzahl der Sequenzläufe nach Lander-Waterman; rote Punkte, tatsächliche Anzahl der Sequenzlücken; rote Linie, Interpolation.

3.II.1.3 Editierung des Rohdatensatzes

Während der ersten Editierung, die von Dr. A. Henne (Göttingen Genomics Laboratory) durchgeführt wurde, konnte von der hohen Abdeckung (8.34) des Genoms durch Sequenzläufe von sehr guter Qualität und hoher Durchschnittslänge (737 bp) profitiert werden. Vor Beginn der Editierung wurden mit der Gap4-Funktion *Find Internal Joins* alle Contig-Enden miteinander abgeglichen und bereits einige Sequenzlücken geschlossen. Während der Editierung wurden problematische Bereiche durch Wiederholung einzelner Sequenzläufe und durch Primerwalkings auf Plasmid-Inserts abgesichert. An den Enden der Contigs wurden hierzu spezifische Primer abgeleitet, um durch Walkings auf Plasmidbrücken verbliebene Sequenzlücken zu schließen (siehe 2.11.1).

Parallel dazu wurden auffällige Fehlassemblierungen aufgelöst bzw. zur späteren Bearbeitung markiert (rRNA-Operons; siehe 3.II.1.5). Auf diese Weise konnte die Zu-

verlässigkeit des Datensatzes innerhalb kurzer Zeit erhöht und die Anzahl der Contigs größer 3 kbp von 117 auf 53 verringert werden.

3.II.1.4 Lückenschluß

Nach der ersten Editierung des Rohdatensatzes wurde eine genomweite Vorhersage der kodierenden offenen Leseraster (CDSs) mit YACOP durchgeführt (siehe 2.13.1). Die resultierenden CDSs wurden in ERGO annotiert (siehe 2.13.3). Aus dem Vergleich der Annotationen von CDSs der Contig-Enden von *Msp. stadtmanae* mit den Annotationen bereits vollständig sequenzierter Organismen konnten zusätzliche Informationen über die Reihenfolge der Contigs im Genom gesammelt werden (siehe 2.11.2.2). Erwartungsgemäß wies *Msp. stadtmanae* dabei die größten Übereinstimmungen in Sequenz und Anordnung der Gene zu *Methanothermobacter thermautotrophicus* str. Delta H (Smith *et al.*, 1997) auf. Mit den verbliebenen Contigs wurden Multiplex-PCRs durchgeführt, d.h. verschiedene Primer in mehreren Ansätzen vereinigt und in allen möglichen Varianten miteinander kombiniert (siehe 2.11.3).

3.II.1.5 Identifikation von zwei Varianten eines ribosomalen Operons

Das Genom von *Msp. stadtmanae* enthält ein ribosomales RNA-Operon, dessen Sequenz zunächst nicht eindeutig aufgelöst werden konnte. Unklarheit herrschte über den C-Terminus der 23S rRNA dieses Operons, der je nach Sequenzlauf entweder vollständig oder um ca. 900 bp verkürzt vorlag (Abb. 3.24). Beide Versionen wurden sowohl durch Sequenzläufe aus der Rohsequenzierung der Genbank als auch durch Primerwalkings auf PCR-Produkten gestützt, die mit insgesamt fünf verschiedenen, spezifischen Primerkombinationen amplifiziert worden waren. Deshalb wurde untersucht, ob es sich bei der genomischen DNA, die für die Genbank-Herstellung und die PCR-Amplifikationen verwendet wurde, möglicherweise um ein uneinheitliches Gemisch aus genotypisch unterschiedlichen *Msp. stadtmanae*-Zellen handelte. Hierzu wurde chromosomale DNA aus verschiedenen Kulturen von *Msp. stadtmanae* isoliert, die zu unterschiedlichen Zeitpunkten zwischen 2002 und 2005 aus jeweils frisch bezogenen Zellen der Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH, Braun-

schweig (DSMZ) angezogen worden waren. Mit Hilfe von PCR-Amplifikationen konnte gezeigt werden, daß die kürzere Variante der 23S rRNA erst ab 2003 innerhalb der Zellkulturen der DSMZ auftauchte und seitdem beide Varianten in einem Gemisch vorlagen. Da durch die deutliche Verkürzung der 23S rRNA von einem Funktionsverlust des Gens ausgegangen werden muß, wurde die kürzere Variante als Deletionsmutante gewertet und das vollständige 23S rRNA-Gen der längeren Variante dem Wildtyp-Stamm zugewiesen.

3.II.1.6 Vorhersage und Annotation kodierender Sequenzen

Aufgrund des niedrigen G+C-Gehaltes des *Msp. stadtmanae*-Genoms (siehe 3.II.2.1) liegen die A/T-reichen STOP-Codons (TAA, TAG, TGA) statistisch gehäuft vor. Dadurch wird einerseits die Anzahl zufälliger, artifizieller offener Leserrahmen (ORFs) drastisch reduziert und andererseits durch die stark begrenzte Länge der kodierenden ORFs eine zuverlässigere Vorhersage der START-Codons möglich. Die CDS-Vorhersage wurde mit YACOP durchgeführt (siehe 2.13.1). Aufgrund der geringen Größe und der relativ homogenen Struktur des *Msp. stadtmanae*-Genoms lieferte diese Vorhersage eine gute Grundlage für die anschließende manuelle Annotation in ERGO, die am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg von H. Seedorf durchgeführt wurde. Als besonders hilfreich für die Annotation erwies sich wiederum der Sequenzvergleich mit *Methanothermobacter thermautotrophicus* str. Delta H (Smith *et al.*, 1997).

3.II.2 Sequenzanalyse

3.II.2.1 Allgemeine Eigenschaften

Methanosphaera stadtmanae besitzt ein relatives kleines Genom, das in einem einzigen Replikon von 1 767 405 bp Länge organisiert ist (Abb. 3.25). Tab. 3.18 gibt eine Übersicht über einige allgemeine Eigenschaften des Genoms im Vergleich zu *Mth. thermautotrophicus* str. Delta H (Smith *et al.*, 1997). Alle Angaben, die im folgenden zu diesem Organismus gemacht werden, beziehen sich, soweit nicht anders vermerkt, auf diese Veröffentlichung.

Tab. 3.18: Allgemeine Eigenschaften des *Msp. stadtmanae*-Genoms im Vergleich zu *Mth. thermautotrophicus*.

Eigenschaft	<i>Msp. stadtmanae</i>	<i>Mth. therm.</i>
Länge [bp]	1 767 405	1 751 377
G+C-Gehalt [%]	28	49
Anteil der kodierenden Sequenz [%]	84	90
CDSs	1 545	1 855
ribosomale RNA-Operons	4	2
tRNAs (+ potentielle tRNAs)	38 (+ 2)	39 (+ 2)
Proteine unbekannter Funktion [%]	34	k.A.
Hypothetische Proteine [%]	26	27
Konservierte hypothetische Proteine [%]	12	k.A.

Im Genom von *Msp. stadtmanae* konnten 1545 CDSs identifiziert werden, die einen Anteil von ca. 84 % an der Gesamtsequenz einnehmen. Davon konnte für mehr als ein Drittel (34 %) keine direkte Funktion abgeleitet werden, da in den Vergleichsdatenbanken bisher keine Proteine ähnlicher Sequenz veröffentlicht wurden ("hypothetical proteins": 26 %) bzw. über ähnliche Proteine keine weiteren Informationen vorliegen ("conserved hypothetical proteins": 12 %).

Msp. stadtmanae kodiert für vier ribosomale RNA-Operons, von denen die Operons 2, 3 und 4 starke Übereinstimmungen zeigen, aber deutlich von Operon 1 abweichen (Abb. 3.24). Die drei ähnlichen Operons enthalten etwa gleich lange rRNA-Gene und kodieren zwischen der 16S rRNA und 23S rRNA für eine Alanin-tRNA. Demgegenüber liegen in Operon 1 die 16S rRNA und 5S rRNA um ca. 20 % verkürzt vor; die Alanin-tRNA im intergenischen Bereich zwischen 16S rRNA und 23S rRNA fehlt.

Ein Teil der genomischen DNA, die aus Zellkulturen der DSMZ, Braunschweig isoliert wurde, enthält ein verkürztes Operon 2, dessen 23S rRNA eine C-terminale Deletion von 919 bp aufweist (siehe 3.II.1.5).

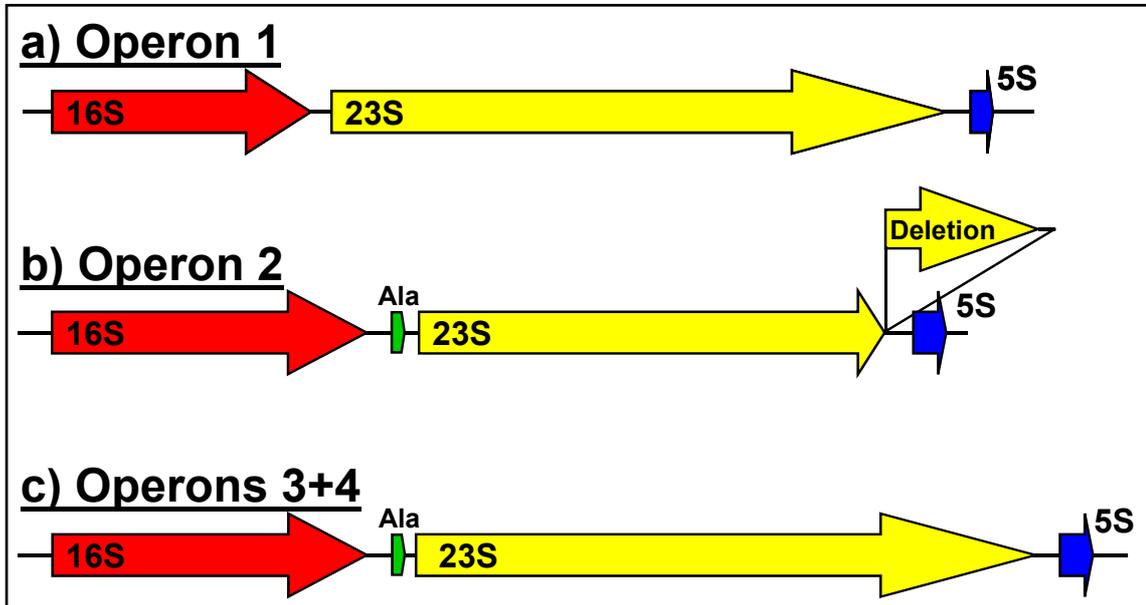


Abb. 3.24: Ribosomale RNA-Operons im Genom von *Msp. stadtmannae*. Die Nummerierung der erfolgte ausgehend vom Replikationsursprung im Uhrzeigersinn. Rot, 16S rRNA; gelb, 23S rRNA; blau, 5S rRNA; grün, Alanin-tRNA.

Mit 28 % liegt der G+C-Gehalt in *Msp. stadtmannae* deutlich niedriger als in *Mth. thermotrophicus* (41 %). Wie in Abb. 3.25 kommt es dabei innerhalb der Genomsequenz von *Msp. stadtmannae* zu starken Schwankungen, die von den kodierenden Eigenschaften der Sequenz beeinflusst werden:

1. Der G+C-Gehalt der kodierenden Sequenz (29.2 %) ist in *Msp. stadtmannae* höher als der G+C-Gehalt der Gesamtsequenz (28 %).
2. Stromaufwärts von einem *cdc6*-Gen, das mit dem potentiellen Replikationsursprung assoziiert wird (siehe 3.II.2.2), weist der G+C-Gehalt ein lokales Minimum auf.
3. Die vier ribosomalen RNA-Operons zeigen mit ca. 43 % G+C-Gehalt die mit Abstand höchsten Werte des gesamten Genoms.
4. Bereiche von abweichendem, insbesondere erhöhtem G+C-Gehalt fallen mit solchen Gen-Clustern zusammen, die anhand einer SIGI-Vorhersage als potentiell hochexprimierte Gene identifiziert wurden (siehe 3.II.2.3).

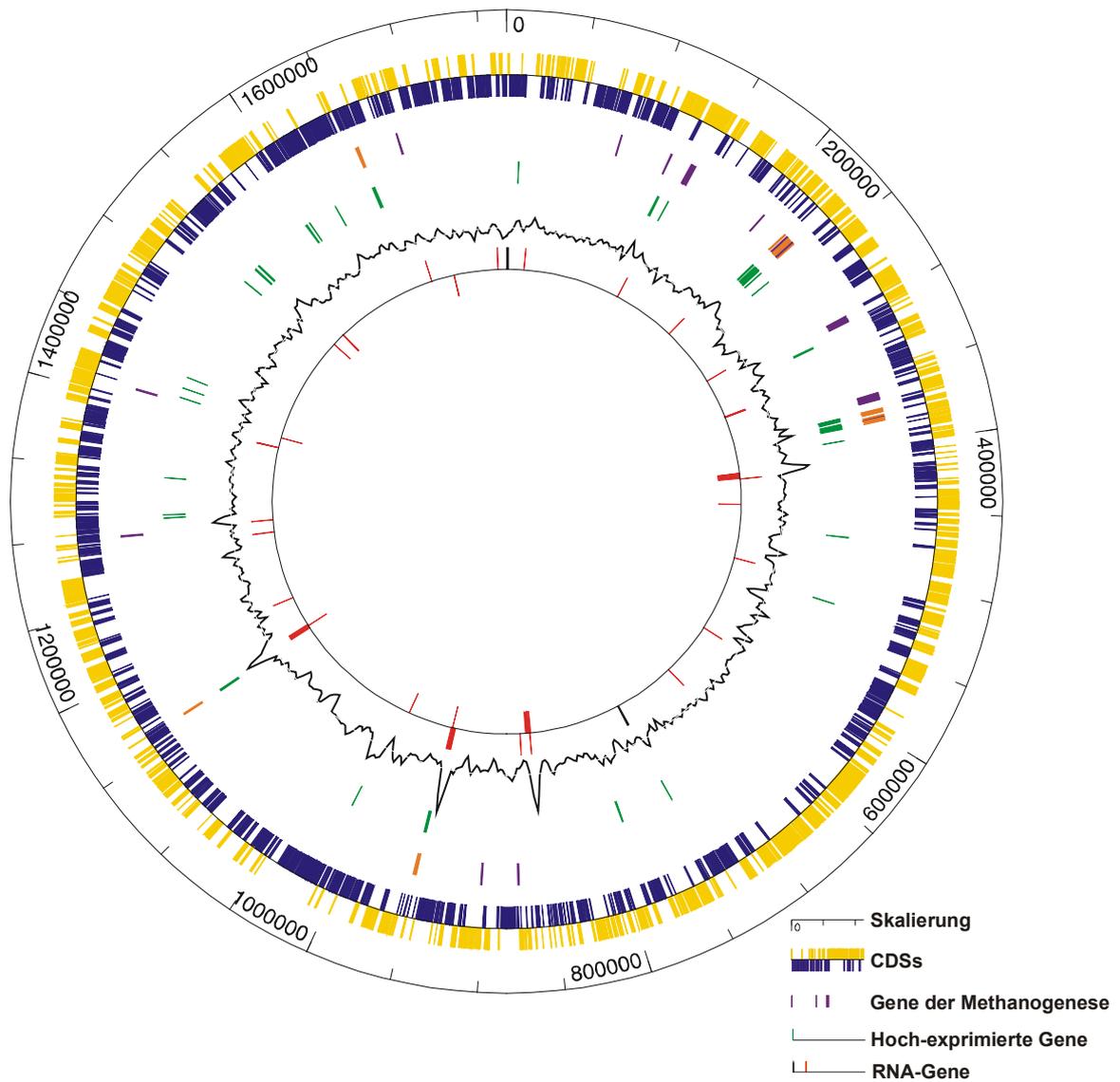


Abb. 3.25: Schematische Darstellung des Genoms von *Msp. stadtmanae*. Von innen nach außen werden folgende Eigenschaften dargestellt: 1. Rote Balken, RNA-Gene, d.h. rRNA-Operons und tRNA-Gene; schwarze Balken, *cdc6*-Gene. Von der Linie aus nach außen weisende Balken stehen für Gene, die in Richtung des Uhrzeigersinns kodiert sind. 2. Schwarze Linie, G+C-Gehalt. Bereiche niedrigen G+C-Gehaltes weichen vom Durchschnitt nach Innen ab. 3. Grüne Balken, potentiell hochexprimierte Gene nach SIGI-Vorhersage (siehe 3.II.2.3). Strangunabhängige Darstellung. 4. Violette Balken, CDSs für Enzyme der Methanogenese; orange Balken, CDSs für Enzyme der Methanogenese aus Methanol. Strangunabhängige Darstellung. 5. Blaue Balken, CDSs in Richtung gegen den Uhrzeigersinn; gelbe Balken, CDSs in Richtung des Uhrzeigersinns. 6. Skalierung ausgehend vom Replikationsursprung in Basenpaaren.

3.II.2.2 Replikationsursprung

Wenig ist bisher über Mechanismen bekannt, die zur Initiation der Replikation in Archaeen führen. Allerdings konnte über komparative Genomanalysen eine grundsätzliche Ähnlichkeit in Prozessen der Informationsverwaltung zwischen archaeellen und eukaryotischen Systemen nachgewiesen werden (Olsen & Woese, 1997). Im Gegensatz zu Prokaryoten mit einem singulären und Eukaryoten mit mehreren Replikationsursprüngen scheint die Verteilung der Anzahl der Replikationsursprünge innerhalb der Archaeen uneinheitlich zu sein. Während verschiedene Untersuchungen – unter anderem über die Analyse des kumulativen *GC-Skews* – einen einzigen Replikationsursprung in *Mth. thermotrophicus* und *Pyrococcus horikoshii* nahelegen (Lopez *et al.*, 1999; Rocha *et al.*, 1999), wurden in zwei *Sulfolobus*-Spezies drei chromosomale Replikationsursprünge *in vivo* nachgewiesen (Lundgren *et al.*, 2004). Lundgren *et al.* zeigten zudem, daß zwei der Replikationsursprünge in direkter Nachbarschaft von *cdc6*-Genen liegen. Cdc6-Faktoren wird in Archaeen ähnlich wie in Eukaryoten eine Funktion bei der Initiation der Replikation durch Bindung des *minichromosome maintenance complex* (MCM) zugeschrieben (De Felice *et al.*, 2004).

Eine kumulative *GC-Skew*-Analyse des Genoms von *Msp. stadtmanae* weist mit einem relativ klar begrenzten globalen Minimum auf nur einen Replikationsursprung hin, der in einen intergenischen Bereich neben der CDS für ein potientielles *cdc6*-Gen fällt (Msp_0001; Abb. 3.25, Abb. 3.26). In unmittelbarer Nachbarschaft davon liegen Gene für die kleine Untereinheit der DNA-Polymerase II (EC 2.7.7.7; Msp_1584), zwei Helikasen (Msp_0005, Msp_0007) und zwei Gene aus der Cobalamin-Biosynthese (Msp_1587, Msp_1588; *cbiBG*; Roth *et al.*, 1993). Ein weiteres *cdc6*-Homolog, in dessen Nachbarschaft sich jedoch keine längeren intergenischen Bereiche befinden, ist etwa 800 kbp entfernt auf dem Chromosom kodiert (Abb. 3.25). Da sowohl die kumulative *GC-Skew*-Analyse als auch die Nachbarschaft des zweiten *cdc6*-Homologs keine Anzeichen für einen weiteren Replikationsursprung beinhalten, wird nach gegenwärtigem Kenntnisstand von einem singulären Replikationsursprung im Genom von *Msp. stadtmanae* ausgegangen.

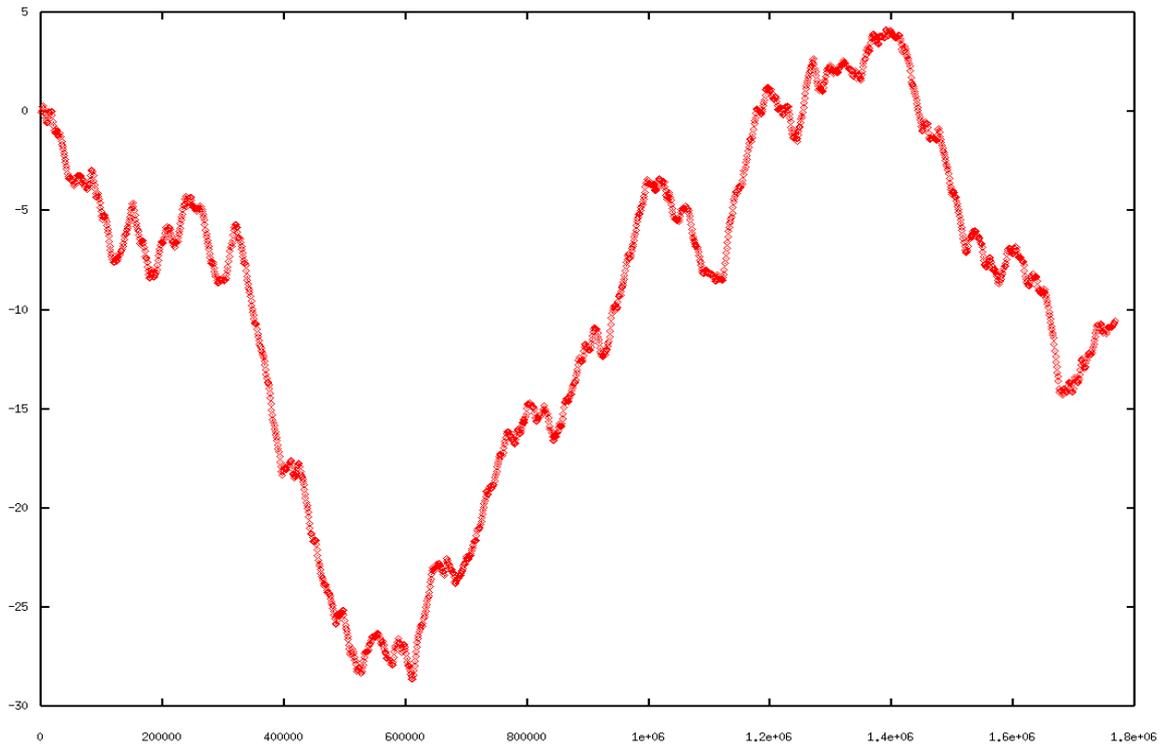


Abb. 3.26: Kumulative GC-Skew-Analyse des *Msp. stadtmannae*-Genoms nach Grigoriev (1998).

3.II.2.3 Potentiell hochexprimierte Gene

Das Programm SIGI (Merkl, 2004) identifiziert mit Hilfe der *Codon Usage* nicht nur sogenannte Fremdgene (siehe 3.I.3.2), sondern auch potentiell hochexprimierte Gene (siehe 2.15). Beide Gruppen von Genen werden durch Abweichungen von der durchschnittlichen *Codon Usage* charakterisiert (Merkl, 2003). Über Untersuchungen der *Codon Usage* in prokaryotischen Genomen werden typischerweise Gene als hochexprimiert charakterisiert, die für ribosomale Proteine, Transkriptions- und Translationsfaktoren, Chaperone und Enzyme des Basis-Energiestoffwechsels kodieren (Karlín & Mrazek, 2000).

Durch eine SIGI-Vorhersage auf Grundlage des *Msp. stadtmannae*-Datensatzes wurden 39 CDSs identifiziert, deren *Codon Usage* eine überdurchschnittlich gute Adaptation aufweist (Tab. 3.19; Abb. 3.25). Gemessen daran, daß ca. ein Drittel aller CDSs in *Msp. stadtmannae* nicht funktionell annotiert werden konnte (siehe 3.II.2.1), kodieren wenig potentiell hochexprimierte CDSs für hypothetische Proteine unbekannter Funktion (< 10 %).

Tab. 3.19: Potentiell hochexprimierte Gene im Genom von *Msp. stadtmanae* nach SIGI-Vorhersage. Positionsangaben ausgehend vom Replikationsursprung im Uhrzeigersinn.

CDS	Position [kbp]	Annotation
Msp_0114	129	thermosome, subunit beta (<i>thsB</i>)
Msp_0122	138	archaeal histone
Msp_0182	226	methylcobalamin:coenzyme M methyltransferase 4 (<i>mtaA4</i>)
Msp_0184	228	methanol:cobalamin methyltransferase 3 (<i>mtaB3</i>)
Msp_0185	229	methanol corrinoid protein 3 (<i>mtaC3</i>)
Msp_0186	231	methanol:cobalamin methyltransferase 2 (<i>mtaB2</i>)
Msp_0187	232	methanol corrinoid protein 2 (<i>mtaC2</i>)
Msp_0189	235	methanol corrinoid protein 1 (<i>mtaC1</i>)
Msp_0198	245	DNA/RNA-binding protein AlbA (<i>albA</i>)
Msp_0251	310	putative thiamine biosynthesis protein 2 (<i>thiC2</i>)
Msp_0314	372	F420-non reducing hydrogenase, subunit delta (<i>mvhD1</i>)
Msp_0315	373	F420-non reducing hydrogenase, subunit gamma (<i>mvhG</i>)
Msp_0316	374	F420-non reducing hydrogenase, subunit alpha (<i>mvhA</i>)
Msp_0318	378	methyl-coenzyme M reductase II, subunit B (<i>mrtB</i>)
Msp_0320	379	methyl-coenzyme M reductase II, subunit G (<i>mrtG</i>)
Msp_0321	380	methyl-coenzyme M reductase II, subunit A (<i>mrtA</i>)
Msp_0330	391	hypothetical protein
Msp_0409	471	hypothetical conserved protein
Msp_0633	742	50S ribosomal protein L7Ae (<i>rpl7Ae</i>)
Msp_0675	786	conserved hypothetical protein
Msp_0818	950	methanol corrinoid protein 4 (<i>mtaC4</i>)
Msp_0819	951	methanol:cobalamin methyltransferase 4 (<i>mtaB4</i>)
Msp_0886	1014	50S ribosomal protein L15P (<i>rpl15p</i>)
Msp_1013	1160	heterodisulfide reductase, subunit B1 (<i>hdrB1</i>)
Msp_1014	1160	heterodisulfide reductase, subunit C1(<i>hdrC1</i>)
Msp_1139	1311	A1A0 ATPase, subunit K (<i>ahaK</i>)
Msp_1141	1314	A1A0 ATPase, subunit H (<i>ahaH</i>)
Msp_1174	1344	50S ribosomal protein L37e (<i>rpl37e</i>)
Msp_1243	1412	50S ribosomal protein L15e (<i>rpl15e</i>)
Msp_1253	1420	50S ribosomal protein L37Ae (<i>rpl37Ae</i>)
Msp_1263	1430	50S ribosomal protein L12P (<i>rpl12p</i>)
Msp_1352	1523	translation initiation factor 5A (<i>aIF-5A</i>)
Msp_1366	1536	translation elongation factor 1-alpha (EF-Tu) (<i>tuf</i>)
Msp_1368	1540	30S ribosomal protein S7P (<i>rps7p</i>)
Msp_1369	1540	30S ribosomal protein S12P (<i>rps12p</i>)
Msp_1416	1591	glutamate dehydrogenase (<i>gdhA</i>)
Msp_1419	1594	hypothetical protein
Msp_1450	1619	energy-converting hydrogenase B, subunit H (<i>ehbH</i>)
Msp_1476	1654	heterodisulfide reductase, subunit A1 (<i>hdrA1</i>)

Auffallenderweise werden Gene für vier Proteinkomplexe aus insgesamt 17 CDSs als hochexprimiert charakterisiert, die möglicherweise für den gesamten Enzymapparat der Umsetzung von Methanol zu Methan in *Msp. stadtmanae* kodieren (siehe 3.II.3).

Zwei weitere potentiell hochexprimierte Gene kodieren für die K-Untereinheit (*ahaK*; Msp_1139) und H-Untereinheit (*ahaH*; Msp_1141) einer ATPase des A₁A₀-Typs zur H⁺-Translokation. Diese Gene sind Teil eines in methanogenen Archaeen konservierten Clusters, der für insgesamt neun Untereinheiten der A₁A₀-ATPase kodiert

und in *Methanosarcina mazei* (Ruppert *et al.*, 1998) und *Methanothermobacter thermautotrophicus* (Smith *et al.*, 1997) beschrieben wurde (*ahaHIKECFABDG*). Bei der K-Untereinheit des Enzyms handelt es sich wahrscheinlich um das hydrophobe Proteolipid, das direkt an der H⁺-Translokation beteiligt ist und sich aus einer unterschiedlichen Anzahl von Transmembran(TM)-Helices zusammensetzt (Ruppert *et al.*, 1998). Das potentielle Gen *ahaK* aus *Msp. stadtmanae* kodiert für ein Protein, das nach TMpred-Vorhersage ebenso wie die K-Untereinheit aus *Mth. thermautotrophicus* vier TM-Helices enthält. H⁺-transportierende A₁A₀-ATPasen sind bis jetzt in allen methanogenen Archaeen identifiziert worden und werden für die ATP-Synthese unter Ausnutzung eines während der Methanogenese erzeugten Ionengradienten verantwortlich gemacht (Deppenmeier, 2002a). Damit gehört die V-Typ ATPase wahrscheinlich ebenso wie die Enzyme der Methanogenese aus Methanol zum Basis-Energiestoffwechsel in *Msp. stadtmanae*.

3.II.3 Methanogenese in *Msp. stadtmanae*

Msp. stadtmanae wächst ausschließlich auf Methanol und molekularem Wasserstoff, andere methanogene Substrate werden nicht umgesetzt (Miller & Wolin, 1985). Die Reduktion von Methanol zu Methan findet in *Msp. stadtmanae* in analoger Weise statt wie in *Methanosarcina barkeri* (van de Wijngaard *et al.*, 1991). Zunächst überträgt eine lösliche Methyltransferase den Methylrest aus Methanol auf HS-Coenzym M (Ding *et al.*, 2002). In den *Methanosarcina*-Spezies werden die Methylreste aus Methyl-S-Coenzym M im Gegensatz zu *Msp. stadtmanae* über den methylotrophen Weg der Methanogenese zu CH₄ und CO₂ disproportioniert (Deppenmeier, 2004). Dabei wird im oxidativen Zweig des Weges einer von vier Methylresten zu CO₂ oxidiert und die aus dieser Oxidation stammenden Elektronen für die Methanogenese aus den drei übrigen Methylresten im reduktiven Zweig des methylotrophen Weges verwendet (Deppenmeier, 2002b).

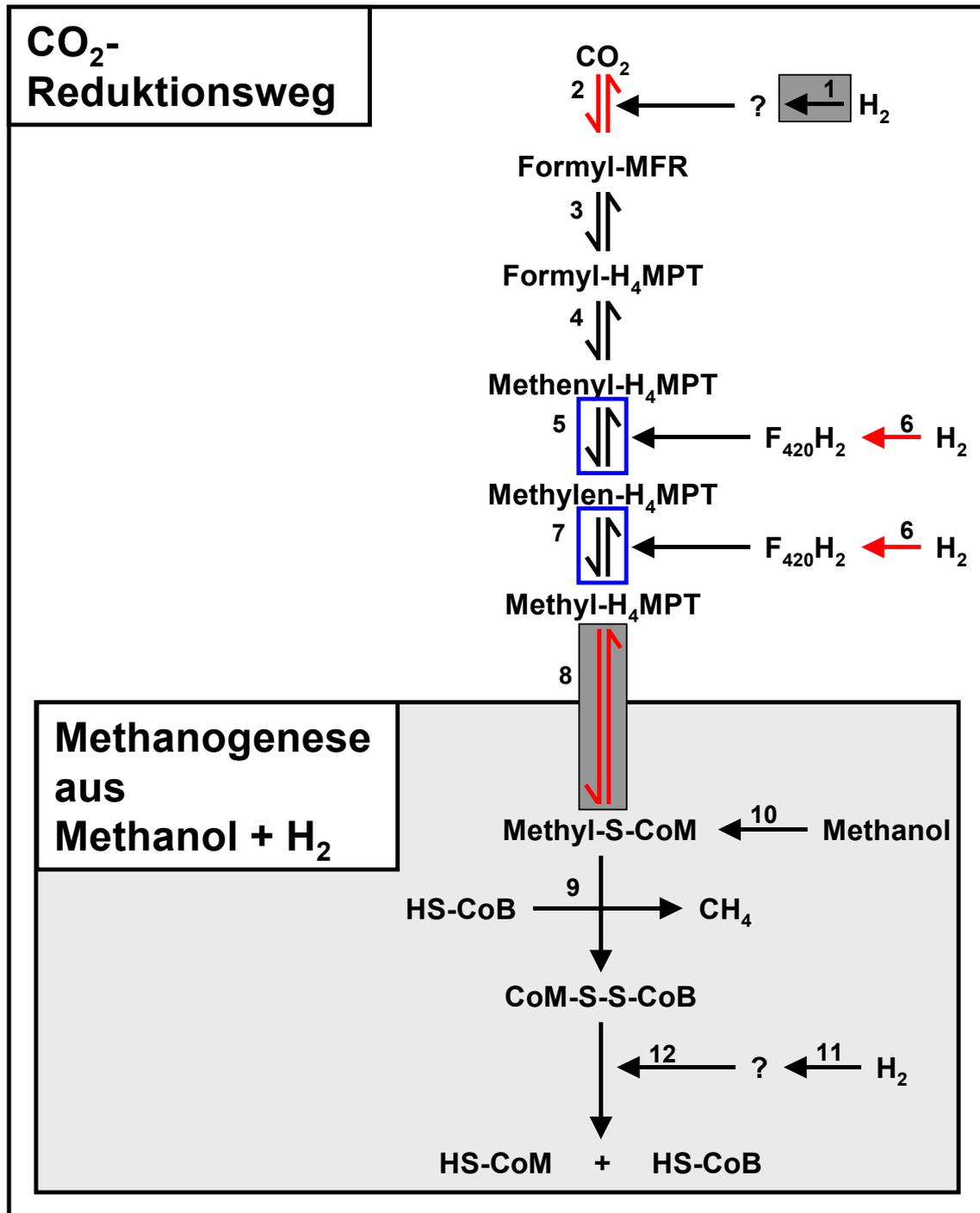


Abb. 3.27: Schematische Darstellung des CO₂-Reduktionsweges und der Methanogenese aus Methanol und H₂. Zu allen Reaktionsschritten, die von nummerierten Enzymen katalysiert werden, wurden in *Msp. stadtmanae* analoge kodierende Sequenzen identifiziert. Rote Pfeile markieren Reaktionsschritte, zu denen in *Msp. stadtmanae* keine Aktivität nachgewiesen werden konnte (van de Wijnngaard *et al.*, 1991; Wong *et al.*, 1994). Zu den Reaktionsschritten der CO₂-Reduktion in blauen Kästchen wurden Enzymaktivitäten nachgewiesen (van de Wijnngaard *et al.*, 1991). Dunkelgrau hinterlegte Reaktionen werden durch membrangebundene Enzyme katalysiert. Abkürzungen: F₄₂₀H₂, reduzierte Form von Coenzym F₄₂₀; MFR, Methanofuran; H₄MPT, Tetrahydromethanopterin; HS-CoM, Coenzym M; HS-CoB, Coenzym B. Enzyme: 1, Ehb-ähnliche Hydrogenase; 2, Formyl-MFR-Dehydrogenase; 3, Formyl-MFR:H₄MPT-Formyltransferase; 4, Methenyl-H₄MPT-Cyclohydrolase; 5, Methylene-H₄MPT-Dehydrogenase; 6, F₄₂₀-reduzierende Hydrogenase; 7, Methylene-H₄MPT-Reduktase; 8, Methyl-H₄MPT:HS-CoM-Methyltransferase; 9, Methyl-CoM-Reduktase; 10, lösliche Methyltransferase; 11, F₄₂₀-non-reducing-Hydrogenase; 12, Heterodisulfid-Reduktase.

In *Msp. stadtmanae* wird Methanol nur über den reduktiven Zweig des methylo-trophen Weges umgesetzt, deshalb benötigt der Organismus H_2 als Elektronenquelle (Miller & Wolin, 1985): Methyl-S-Coenzym M wird reduktiv gespalten durch die Methyl-CoM-Reduktase. Dabei wird ein Heterodisulfid (CoM-S-S-CoB) gebildet und Methan freigesetzt. Untereinheiten der Methyl-CoM-Reduktase wurden in Zellextrakten von *Msp. stadtmanae* nachgewiesen (Rouviere & Wolfe, 1987). CoM-S-S-CoB wird durch die Heterodisulfid-Reduktase zu HS-CoM und HS-CoB reduziert. Die nötigen Reduktionsäquivalente werden von einem membrangebundenen Elektronentransportsystem bereitgestellt und stammen aus der Oxidation von molekularem Wasserstoff (Ide *et al.*, 1999). Der Elektronentransport ist mit einer Protonen-Translokation über die Cytoplasmamembran verbunden und wird für die Synthese von ATP genutzt (Deppenmeier, 2002a).

Um die genetischen Grundlagen der Methanogenese in *Msp. stadtmanae* zu charakterisieren, wurde das Genom nach kodierenden Sequenzen für die oben beschriebenen Enzyme durchsucht.

1. Die Methanol:CoM-Methyltransferase setzt sich in *Methanosarcina barkeri* aus den drei Untereinheiten MtaA, MtaB und MtaC zusammen (Keltjens & Vogels, 1993). MtaC ist ein Corrinoïd-Protein und bildet einen festen Komplex mit MtaB. Entsprechend sind beide Proteine innerhalb eines *mtaCB*-Operons kodiert (Sauer *et al.*, 1997); MtaA ist an anderer Stelle kodiert und wird monocistronisch exprimiert (Harms & Thauer, 1996). Das Genom von *Msp. stadtmanae* enthält vier potentielle *mtaBC*-Operons (Tab. 3.20). *mtaB1C1*, *mtaB2C2* und *mtaB3C* sind direkt hintereinander auf dem gleichen Strang kodiert, ca. 225 kbp vom Replikationsursprung entfernt, *mtaB4C4* dagegen ca. 950 kbp vom Replikationsursprung entfernt. Von vier möglichen *mtaA*-Genen ist *mtaA4* direkt neben den drei geclusterten *mtaBC*-Operons und auf dem gleichen Strang kodiert, *mtaA1*, *mtaA2* und *mtaA3* liegen dagegen auf dem Chromosom verstreut. Von 12 potentiellen *mta*-Genen in *Msp. stadtmanae* wurden – bis auf *mtaB1* im Cluster und die drei verstreuten *mtaA*-Gene – alle als hochexprimiert charakterisiert (siehe 3.II.2.3). Interessanterweise liegen alle *mtaB*- und *mtaC*-Gene in *Msp. stadtmanae* im Gegensatz zu verschiedenen *Methanosarcina*-Spezies (*Msa. mazei*, *Msa. barkeri* und *Msa. acetivorans*) innerhalb der Cluster in umgekehrter Reihenfolge angeordnet vor (*mtaBC* statt *mtaCB*).

Tab. 3.20: Untereinheiten der Methanol:CoM-Methyltransferase. Positionsangaben ausgehend vom Replikationsursprung im Uhrzeigersinn. Bezeichnung der Gene nach Keltjens & Vogels (1993).

CDS	Position [kbp]	Gen	Besonderheit
Msp_0112	124	<i>mtaA2</i>	
Msp_0182	226	<i>mtaA4</i>	hochexprimiert
Msp_0184	228	<i>mtaB3</i>	hochexprimiert
Msp_0185	229	<i>mtaC3</i>	hochexprimiert
Msp_0186	231	<i>mtaB2</i>	hochexprimiert
Msp_0187	232	<i>mtaC2</i>	hochexprimiert
Msp_0188	233	<i>mtaB1</i>	
Msp_0189	235	<i>mtaC1</i>	hochexprimiert
Msp_0761	874	<i>mtaA1</i>	
Msp_0774	902	<i>mtaA3</i>	
Msp_0818	950	<i>mtaC4</i>	hochexprimiert
Msp_0819	951	<i>mtaB4</i>	hochexprimiert

2. In *Mth. thermautotrophicus* wurden zwei Methyl-CoM-Reduktasen identifiziert, deren Expression wahrscheinlich in Abhängigkeit von unterschiedlichen Wachstumsbedingungen reguliert wird (MCR I, Ellermann *et al.*, 1988; MCR II, Rospert *et al.*, 1990). Während MCR II verstärkt unter exponentiellen Wachstumsbedingungen und bei ausreichender Versorgung mit H₂ und CO₂ exprimiert wird, dominiert MCR I unter H₂- und CO₂-limitierenden Bedingungen. In *Msp. stadtmanae* wurde ein *mrtBDGA*-Operon identifiziert (AJ584650; Krueer & Thauer (2003), unpublished), das große Ähnlichkeit zu demjenigen *mrt*-Operon aus *Mth. thermautotrophicus* aufweist, das für die MCR II kodiert (Msp_0318-0321). Drei der vier Gene in *Msp. stadtmanae* (*mrtBGA*) wurden als potentiell hochexprimierte Gene charakterisiert (siehe 3.II.2.3). Die genaue Funktion des vierten Gens (*mrtD*) ist bisher nicht bekannt. An anderer Stelle im Genom kodiert *Msp. stadtmanae* für eine potentielle C-Untereinheit der MCR I (Msp_0299).

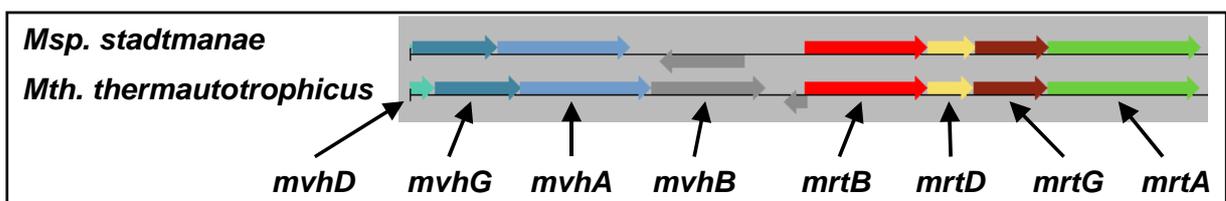


Abb. 3.28: Ähnliche Organisation der für die Methyl-CoM-Reduktase (*mrtBDGA*) und F₄₂₀-non-reducing-Hydrogenase (*mvhDGAB*) kodierenden Operons in *Msp. stadtmanae* und *Mth. thermautotrophicus*.

3. In *Mth. marburgensis* bilden Heterodisulfid-Reduktase und F₄₂₀-non-reducing-Hydrogenase einen membranassoziierten, stabilen H₂-Heterodisulfid-Oxidoreduktase-Komplex (Setzke *et al.*, 1994). Dieser Komplex setzt sich aus drei Untereinheiten der Heterodisulfid-Reduktase und vier Untereinheiten der F₄₂₀-non-reducing-Hydrogenase zusammen. In dem nah verwandten methanogenen Archaeon *Mth. thermautotrophicus* ist eine F₄₂₀-non-reducing-Hydrogenase in einem *mvh*-Operon (methyl viologen-reducing hydrogenase) aus vier Genen kodiert (Reeve *et al.*, 1989). *mvhDGA* kodieren für drei Untereinheiten der Hydrogenase und *mvhB* wahrscheinlich für ein Polyferredoxin, das möglicherweise als primärer Elektronenakzeptor für die Kern-Hydrogenase fungiert. In *Msp. stadtmanae* ist ein *mvhDGA*-Operon (Msp_0314-0316) in unmittelbarer Nachbarschaft vom *mrtBDGA*-Operon der Methyl-CoM-Reduktase (Abb. 3.28) kodiert. Mit Ausnahme des *mvhB*-Gens, das in *Msp. stadtmanae* fehlt, weisen *Msp. stadtmanae* und *Mth. thermautotrophicus* über das gesamte *mvh*- und *mrt*-Operon hinweg eine identische Genom-Organisation auf. Auf Proteinebene zeigen die *mvh*-Gene beider Organismen Identitäten zwischen 65 % und 75 %. Alle drei Untereinheiten der potentiellen F₄₂₀-non-reducing-Hydrogenase gehören zu den hochexprimierten Genen in *Msp. stadtmanae* (siehe 3.II.2.3).

Die Heterodisulfid-Reduktase aus *Mth. thermautotrophicus* (Hdr) besteht aus drei Untereinheiten, die in einem einzelnen *hdrA*-Gen und einem an anderer Stelle im Genom lokalisierten *hdrCB*-Operon kodiert sind (Hedderich *et al.*, 1994). In analoger Weise sind die drei Untereinheiten in *Msp. stadtmanae* in *hdrA1* (Msp_1476) und einem *hdrC1B1*-Operon (Msp_1014, Msp_1013) kodiert. Alle drei Gene wurden als hochexprimiert charakterisiert (siehe 3.II.2.3).

In den *Methanosarcina*-Spezies und in *Archaeoglobus fulgidus* wurden Gene identifiziert, die eine Fusion aus einem *mvhD*-Homolog mit dem 3'-Ende eines *hdrA*-Homologs darstellen (Stojanowic *et al.*, 2003). Daraus wurde geschlossen, daß *mvhD* für ein Elektronentransferprotein der Hydrogenase kodiert, das die Heterodisulfid-Reduktase über die Untereinheit HdrA mit Reduktionsäquivalenten versorgt. Das Genom von *Msp. stadtmanae* verfügt neben dem oben beschriebenen über ein zweites Operon, das für eine potentielle Heterodisulfid-Reduktase kodiert (*hdrC2B2A2*, Msp_0125-0127). *hdrA2* weist starke Ähnlichkeit zu den fusionierten *hdrA*-Homologen aus den *Methanosarcina*-Spezies auf. Es unterscheidet sich von *hdrA1* und homologen Genen durch ein verlängertes 3'-Ende mit Übereinstimmung zu *mvhD*-Genen aus *Archaeoglobus fulgidus* (NP_070203) und *Methanopyrus kandleri* (NP_613912).

Allerdings weist das *hdrC2B2A2*-Operon aus *Msp. stadtmanae* eine andere Organisation auf als vergleichbare Operons aus den *Methanosarcina*-Spezies (Abb. 3.29).

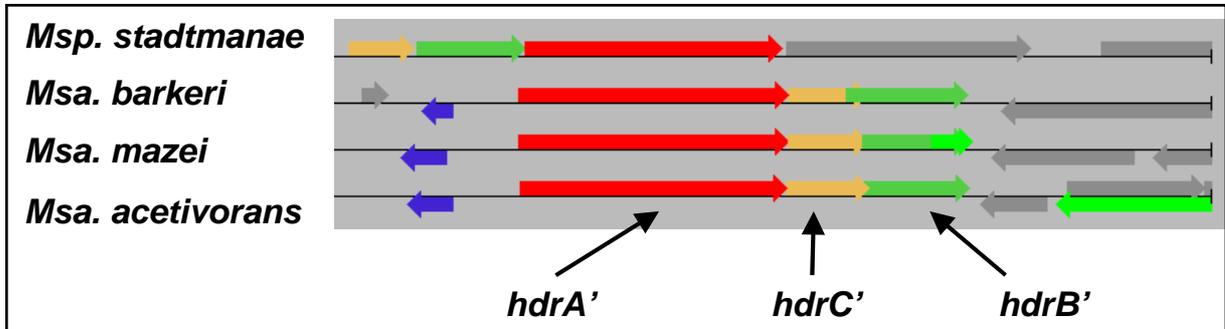


Abb. 3.29: Unterschiede in der Organisation des für eine Heterodisulfid-Reduktase vom *Methanosarcina*-Typ kodierenden *hdr'*-Operons in *Msp. stadtmanae* im Vergleich zu verschiedenen *Methanosarcina*-Spezies.

3.II.4 CO₂-Reduktionsweg

Obwohl *Msp. stadtmanae* CO₂ nicht zu Methan reduzieren kann (Miller & Wolin, 1985), konnten schwache Aktivitäten einzelner Enzyme des CO₂-Reduktionsweges (Abb. 3.27) in Zellextrakten von *Msp. stadtmanae* nachgewiesen werden (van de Wijngaard *et al.*, 1991). Bei diesen Enzymen handelt es sich um die Methylen-H₄MPT-Dehydrogenase und die Methylen-H₄MPT-Reduktase. Dagegen wurde Methanofuran als Cofaktor der Formyl-Methanofuran-Dehydrogenase ebensowenig gefunden wie Aktivitäten einer Methyl-H₄MPT:HS-CoM-Methyltransferase (van de Wijngaard *et al.*, 1991) und einer F₄₂₀-abhängige Hydrogenase (Wong *et al.*, 1994). Überraschenderweise verfügt das Genom von *Msp. stadtmanae* dennoch über kodierende Sequenzen für den gesamten Enzymapparat des CO₂-Reduktionweges.

Die Methanbildung aus CO₂ und H₂ (Deppenmeier, 2004) wird mit der Reduktion von CO₂ zu Formyl-Methanofuran (Formyl-MFR) durch die Formyl-MFR-Dehydrogenase eingeleitet (de Poorter *et al.*, 2003). Diese endergonische Reaktion wird über einen elektrochemischen Ionengradienten angetrieben, der durch eine energiekonservierende [NiFe]-Hydrogenase aufgebaut und für die Reduktion eines Polyferredoxins als Elektronenüberträger genutzt wird (Meuer *et al.*, 2002). In einem zweiten Schritt

wird die resultierende Formyl-Gruppe durch die Formyl-MFR:H₄MPT-Formyl-Transferase auf Tetrahydromethanopterin (H₄MPT) übertragen. Formyl-H₄MPT wird schrittweise reduziert zu Methyl-H₄MPT durch die Methenyl-H₄MPT-Cyclohydrolase, Methylen-H₄MPT-Dehydrogenase und Methylen-H₄MPT-Reduktase (Shima *et al.*, 2002). Die Enzyme dieser Reduktionsschritte verwenden als Elektronenüberträger reduziertes Coenzym F₄₂₀ – ein Deazaflavin-Derivat, das durch Oxidation von H₂ mittels einer F₄₂₀-reduzierende Hydrogenase regeneriert wird (Sorgenfrei *et al.*, 1997). Im nächsten Schritt wird die Methyl-Gruppe aus Methyl-H₄MPT auf HS-CoM übertragen. Diese exergonische Reaktion wird durch die Methyl-H₄MPT:HS-CoM-Methyltransferase katalysiert und ist mit der Bildung eines elektrochemischen Na⁺-Gradienten verbunden (Gottschalk & Thauer, 2001). Die weitere Umsetzung des resultierenden Methyl-S-CoM erfolgt entsprechend der Methanogenese aus Methanol (siehe 3.II.3).

1. In *Mth. marburgensis* konnten zwei Formyl-Methanofuran-Dehydrogenasen identifiziert werden, ein Molybdän-abhängiges (Fmd) und ein Wolfram-abhängiges Enzym (Fwd) (Hochheimer *et al.*, 1996). Während das Wolfram-abhängige Enzym konstitutiv exprimiert wird, findet eine Induktion des Molybdän-abhängigen Enzyms durch Molybdat statt (Hochheimer *et al.*, 1998). Beide Enzyme enthalten einen Pterin-Cofaktor in Form eines Molybdopterin-Guanin-Dinukleotids (Bertram *et al.*, 1994).

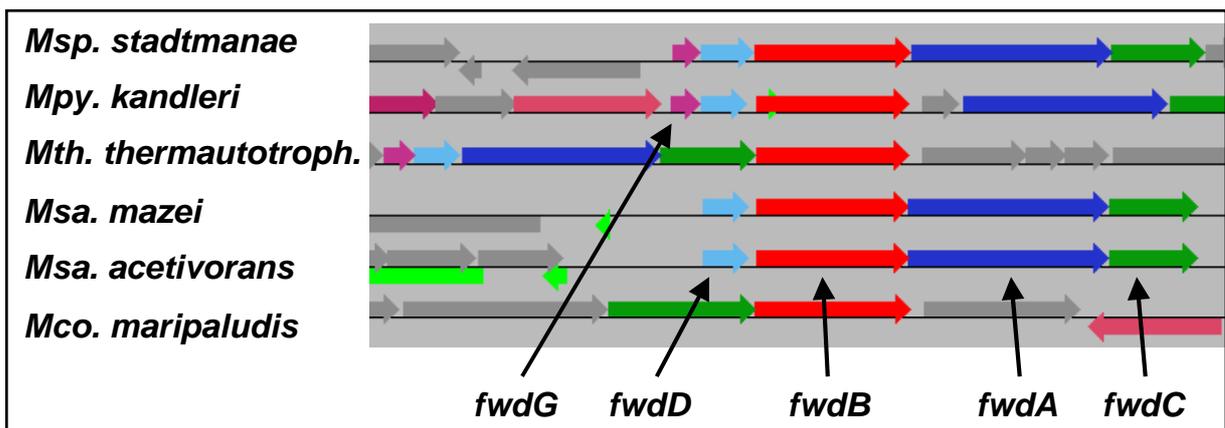


Abb. 3.30: Organisation des für die Formyl-MFR-Dehydrogenase kodierenden *fwd*-Operons in *Msp. stadtmannae* im Vergleich zu verschiedenen *Methanosarcina*-Spezies (*Msa.*), *Methanopyrus* (*Mpy.*) *kandleri* und *Methanococcus* (*Mco.*) *maripaludis*.

In *Msp. stadtmannae* wurde ein Cluster aus fünf Genen identifiziert, der auf Proteinebene stärkere Übereinstimmung mit den Untereinheiten der Wolfram-

abhängigen Formyl-MFR-Dehydrogenase, Isoenzym II, aus *Mth. marburgensis* aufweist als zu dem entsprechenden Isoenzym I. Da sich jedoch beide Isoenzyme strukturell kaum unterscheiden (Johnson *et al.*, 1996), ist eine Bestimmung des Metall-Cofaktors auf Basis der Sequenzdaten allein nicht möglich. Die Reihenfolge der Gene innerhalb des mutmaßlichen *fwdGDBAC*-Operons entspricht der Organisation der Formyl-MFR-Dehydrogenase-kodierenden Gene in *Methanopyrus kandleri* und in den *Methanosarcina*-Spezies (Abb. 3.30). Auf Proteinebene zeigen die einzelnen Gene des potentiellen *fwd*-Clusters unterschiedliche starke Übereinstimmungen zu Untereinheiten der Formyl-MFR-Dehydrogenase verschiedener anderer methanogener Archaeen (Tab. 3.21).

Tab. 3.21: Kodierende Sequenzen der einzelnen Untereinheiten des für die Formyl-MFR-Dehydrogenase kodierenden *fwd*-Operon in *Msp. stadtmanae*. Ähnlichkeiten zu Untereinheiten von Formyl-MFR-Dehydrogenasen anderer methanogener Archaeen nach BLAST-Vergleich.

CDS	Annotation	Gen	Ähnlichkeit
Msp_0241	formylmethanofuran dehydrogenase, subunit G	<i>fwdG</i>	<i>Mth. therm.</i>
Msp_0242	formylmethanofuran dehydrogenase, subunit D	<i>fwdD</i>	<i>Mco. maripaludis</i>
Msp_0243	formylmethanofuran dehydrogenase, subunit B	<i>fwdB</i>	<i>Mpy. kandleri</i>
Msp_0244	formylmethanofuran dehydrogenase, subunit A	<i>fwdA</i>	<i>Mth. therm.</i>
Msp_0245	formylmethanofuran dehydrogenase, subunit C	<i>fwdC</i>	<i>Mca. jannaschii</i>

2. In *Mth. thermautotrophicus* wurden zwei membrangebundene [NiFe]-Hydrogenasen charakterisiert (*energy-converting hydrogenases*, Eha und Ehb), die in Operons aus 20 CDSs (Eha) bzw. 17 CDSs (Ehb) kodiert sind (Tersteegen & Hedderich, 1999). In *Msp. stadtmanae* wurde ein Operon identifiziert (Msp_1457-1442, Msp_1436), das hohe Übereinstimmungen auf Proteinebene und in der Reihenfolge der Gene innerhalb des Operons mit dem *ehb*-Operon aus *Mth. thermautotrophicus* aufweist (*ehbABCDEFGHIJKLMNOQ*). Auf ein Rekombinationsereignis, das wahrscheinlich innerhalb des *ehb*-Operons aus *Msp. stadtmanae* stattgefunden hat, deutet ein kurzer Cluster zwischen *ehbP* und *ehbQ*, der ein IS-Element enthält (ISMst1) und für potentielle Eisen-Transportproteine kodiert (Msp_1441, Msp_1440).

3. *Msp. stadtmanae* enthält zwei Gene, die für potentielle Formyl-MFR:H₄MPT-Formyltransferasen kodieren (Msp_0070, Msp_1502). Eine dieser CDSs zeigt größere Ähnlichkeit zu einem in *Methanothermus fervidus* (Q49168; Lehmacher, 1994), die andere zu einem in *Mth. thermautotrophicus* charakterisierten Gen (Q9P9J4; Tersteegen & Hedderich, 1999).

4. Zu den Enzymen Methenyl-H₄MPT-Cyclohydrolase, Methenyl-H₄MPT-Dehydrogenase und Methenyl-H₄MPT-Reduktase wurden in *Msp. stadtmanae* kodierende Sequenzen gefunden, die auf Proteinebene Übereinstimmungen zwischen 52 % und 61 % mit den entsprechenden Proteinen in *Mth. marburgensis* zeigen (Tab. 3.22).

Tab. 3.22: Kodierende Sequenzen in *Msp. stadtmanae* für verschiedene Enzyme des CO₂-Reduktionsweges. Ähnlichkeiten zu Enzymen aus *Mth. marburgensis* nach BLAST-Vergleich auf Proteinebene.

CDS	Annotation	Ähnlichkeit
Msp_1238	N(5),N(10)-methenyltetrahydromethanopterin cyclohydrolase (EC 3.5.4.27) (<i>mch</i>)	<i>Mth. marburgensis</i> ^a
Msp_0163	F420-dependent methylenetetrahydromethanopterin dehydrogenase (EC 1.5.99.9) (<i>mta</i>)	<i>Mth. marburgensis</i> ^b
Msp_1128	F420-dependent N5,N10-methylenetetrahydro-methanopterin reductase (EC 1.5.99.11) (<i>mer</i>)	<i>Mth. marburgensis</i> ^a

^a (Vaupel & Thauer, 1995); ^b (Mukhopadhyay *et al.*, 1995)

5. In *Mth. marburgensis* wurde eine membranassoziierte Methyl-H₄MPT:HS-CoM-Methyltransferase identifiziert, die aus acht Untereinheiten zusammengesetzt und innerhalb eines einzigen Operons in der Reihenfolge *mtrEDCBAFGH* kodiert ist (Harms *et al.*, 1995). In *Msp. stadtmanae* ist bei ca. 360 kbp ein ähnliches *mtr*-Operon lokalisiert (Msp_0300-0307). Drei Gene können jedoch nur aufgrund ihrer Lokalisation innerhalb dieses Operons als die potentiellen Untereinheiten MtrBFG annotiert werden (Msp_0303, Msp_0305, Msp_0306), da sie auf Proteinebene kaum Ähnlichkeiten zu den entsprechenden Untereinheiten aus *Mth. marburgensis* zeigen.

8. Die F₄₂₀-reduzierende Hydrogenase aus *Mth. thermotrophicus* ist in einem Operon aus vier Genen kodiert (*frhADGB*), von denen drei (*frhAGB*) für die *alpha*-, *beta*- und *gamma*-Untereinheiten des Enzyms kodieren (Alex *et al.*, 1990). Bei FrhD handelt es sich wahrscheinlich um eine Peptidase, die möglicherweise für die Prozessierung der Hydrogenase verantwortlich ist. Ein entsprechendes *frhADGB*-Operon ist in *Msp. stadtmanae* ca. 1.5 Mbp vom Replikationsursprung entfernt kodiert (Msp_1302-1305).

3.II.5 Genomvergleiche

Um Unterschiede in den spezifischen Eigenschaften von *Msp. stadtmanae* zu untersuchen, wurden bidirektionale BLAST-Vergleiche mit allen bisher veröffentlichten Methanogenen sowie mit allen bisher sequenzierten Archaeen überhaupt durchgeführt (siehe 2.12.1). Bei den methanogenen Archaeen handelte es sich um *Methanocaldococcus jannaschii* DSM 2661 (Bult *et al.*, 1996), *Methanococcus maripaludis* S2 (Hendrickson *et al.*, 2004), *Methanopyrus kandleri* AV19 (Slesarev *et al.*, 2002), *Methanosarcina acetivorans* C2A (Galagan *et al.*, 2002), *Methanosarcina mazei* Go1 (Deppenmeier *et al.*, 2002) und *Mth. thermautotrophicus* str. Delta H (Smith *et al.*, 1997). In die Vergleiche mit allen bisher veröffentlichten Archaeen wurden außerdem *Aeropyrum pernix* K1 (Kawarabayasi *et al.*, 1999), *Archaeoglobus fulgidus* DSM 4304 (Klenk *et al.*, 1997), *Haloarcula marismortui* ATCC 43049 (Baliga *et al.*, 2004), *Halo bacterium* sp. NRC-1 (Ng *et al.*, 2000), *Picrophilus torridus* DSM 9790 (Futterer *et al.*, 2004), *Pyrobaculum aerophilum* str. IM2 (Fitz-Gibbon *et al.*, 2002), *Pyrococcus abyssi* GE5 (Cohen *et al.*, 2003), *Pyrococcus furiosus* DSM 3638 (Robb *et al.*, 2001), *Pyrococcus horikoshii* OT3 (Kawarabayasi *et al.*, 1998), *Sulfolobus solfataricus* P2 (She *et al.*, 2001), *Sulfolobus tokodaii* str. 7 (Kawarabayasi *et al.*, 2001), *Thermoplasma acidophilum* DSM 1728 (Ruepp *et al.*, 2000) und *Thermoplasma volcanium* GSS1 (Kawashima *et al.*, 2000) miteinbezogen.

Anhand dieser Vergleiche sollten einerseits die genetischen Ursachen des reduzierten methanogenen Stoffwechsels in *Msp. stadtmanae* untersucht und andererseits Hinweise auf spezifische genetische Faktoren gesammelt werden, die *Msp. stadtmanae* das kommensale Wachstum im Enddarm ermöglichen.

3.II.5.1 Nicht identifizierte Proteine

Diejenigen Gene aus *Mth. thermautotrophicus*, die bidirektionale BLAST-Treffer in jedem der bisher veröffentlichten Genome methanogener Archaeen aber nicht in *Msp. stadtmanae* aufweisen, sind in Tab. 3.23 dargestellt.

Tab. 3.23: Kodierende Sequenzen, die mit Hilfe bidirektionaler BLAST-Vergleiche in allen methanogenen Archaeen mit Ausnahme von *Msp. stadtmanae* nachgewiesen wurden. RefSeq-Bezeichnung und Annotation nach Smith *et al.* (1997).

RefSeq accession	Annotation nach Smith <i>et al.</i> (1997) (Annotation nach ERGO)
NP_276820	carbon monoxide dehydrogenase, alpha subunit (CO dehydrogenase/acetyl-CoA synthase complex, subunit alpha)
NP_276822	carbon monoxide dehydrogenase, alpha subunit (CO dehydrogenase/acetyl-CoA synthase complex, subunit beta)
NP_276825	corrinoid/iron-sulfur protein, large subunit (CO dehydrogenase/acetyl-CoA synthase complex, subunit gamma)
NP_27682	corrinoid/iron-sulfur protein, small subunit (CO dehydrogenase/acetyl-CoA synthase complex, subunit delta)
NP_276823	nitrogenase reductase related protein (putative CO dehydrogenase accessory protein)
NP_276248	nifH / minD related protein (CO dehydrogenase maturation factor)
NP_276663	molybdenum cofactor biosynthesis MoaA
NP_276967	molybdenum cofactor biosynthesis MoaB
NP_275948	molybdenum cofactor biosynthesis MoaC
NP_276667	tungsten formylmethanofuran dehydrogenase, subunit F
NP_275541	formate hydrogenlyase, subunit 5 (COG3261, HycE, Ni,Fe-hydrogenase III large subunit)
NP_276615	phenylalanyl-tRNA synthetase alpha subunit
NP_276113	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (COG1012, PutA, NAD-dependent aldehyde dehydrogenases)
NP_275992	inosine-5'-monophosphate dehydrogenase related protein VIII
NP_275964	L-isoaspartyl protein carboxyl methyltransferase
NP_275991	ferredoxin
NP_275744	conserved protein (contains ferredoxin domain)
NP_276299	phosphonoacetaldehyde methylase
NP_276732	thiamine biosynthetic enzyme
NP_276058	anion transport system permease protein
NP_277018	possible protein methyltransferase
NP_276195	acetyltransferase
NP_276442	hypothetical protein
NP_275780	hypothetical protein
NP_275367	hypothetical protein
NP_275838	hypothetical protein
NP_276973	hypothetical protein
NP_276349	hypothetical protein
NP_275386	hypothetical protein
NP_276747	hypothetical protein
NP_276390	hypothetical protein
NP_276421	hypothetical protein
NP_276472	hypothetical protein
NP_276473	hypothetical protein
NP_277013	hypothetical protein
NP_275999	hypothetical protein
NP_275703	hypothetical protein
NP_276239	hypothetical protein
NP_276433	hypothetical protein
NP_277015	hypothetical protein
NP_275894	hypothetical protein

Für diesen Vergleich wurden zunächst die gemeinsamen Proteine aller methanogenen Archaeen einschließlich *Msp. stadtmanae* bestimmt und in einem zweiten Schritt mit den gemeinsamen Proteinen aller methanogenen Archaeen ohne *Msp. stadtmanae* verglichen. Als Referenz beider Gruppen diene der veröffentlichte Datensatz von *Mth. thermotrophicus*. Aufgrund des relativ strikten Grenzwertes von e^{-15} wurden weitere Proteine, die eine geringe Länge oder wenig konservierte Sequenz aufweisen, in diesen BLAST-Vergleichen möglicherweise nicht erfaßt. Tab. 3.23 erlaubt deshalb bestenfalls eine konservative Abschätzung möglicher fehlender Funktionen in *Msp. stadtmanae*.

Von 41 CDSs, die in allen methanogenen Archaeen mit Ausnahme von *Msp. stadtmanae* identifiziert wurden, kodiert etwa die Hälfte (20 CDSs) für hypothetische Proteine unbekannter Funktion. Innerhalb der verbliebenen 21 CDSs können zwei funktionelle Gruppen erkannt werden:

1. Insgesamt sechs CDSs, die in *Msp. stadtmanae* fehlen, kodieren in *Mth. thermotrophicus* für Untereinheiten des CO-Dehydrogenase/Acetyl-CoA-Synthase-Komplexes (CO-Dehydrogenase). In autotrophen methanogenen Archaeen dient die CO-Dehydrogenase der assimilatorischen CO₂-Fixierung (Ferry, 1995). Eine Methylgruppe aus der methanogenen Reduktion von CO₂ wird mit einer Carbonylgruppe aus der Reduktion von CO₂ durch die CO-Dehydrogenase zur Synthese von Acetyl-CoA verknüpft (Simpson & Whitman, 1993). *Methanosarcina*-Spezies nutzen die CO-Dehydrogenase außerdem zur Methanogenese aus Acetat (Ferry, 1997). Dabei wird in einer Umkehrung der assimilatorischen CO₂-Fixierung Acetyl-CoA gespalten und der dabei entstehende Methylrest zu Methan reduziert. Aus *Methanosarcina barkeri* wurde ein Enzymkomplex aus CO-Dehydrogenase und Corrinoid-Protein isoliert, der sich aus fünf Untereinheiten zusammensetzt ($\alpha\beta\gamma\delta\epsilon$) (Grahame, 1991). Das aktive Zentrum für die C–C-Spaltung konnte in der β -Untereinheit lokalisiert werden und setzt sich aus einem Nickel-Eisen-Schwefel-Komplex (Ni(2)[Fe(4)S(4)]) zusammen, dem sogenannten A-Cluster (Gencic & Grahame, 2003). In *Msp. stadtmanae* konnten im Gegensatz zu allen vollständig sequenzierten methanogenen Archaeen die α -, β -, γ - und δ -Untereinheiten der CO-Dehydrogenase, ein assoziierter Maturations-Faktor sowie ein potentiell akzessorisches Protein nicht identifiziert werden. Die ϵ -Untereinheiten verschiedener methanogener Archaeen weisen im bidirektionalen BLAST-Vergleich Ähnlichkeiten zueinander auf, die unterhalb des Grenzwertes von e^{-15} liegen. Deshalb taucht diese Untereinheit in Tab. 3.23 nicht auf.

2. Aus der Molybdän-Cofaktor-Biosynthese fehlen in *Msp. stadtmanae* die drei Gene *moaABC*. Das Übergangsmetall Molybdän wird in methanogenen Archaeen hauptsächlich für die Synthese des Molybdopterin-Cofaktors verwendet (Vorholt & Thauer, 1997), der eine essentielle Funktion für die Formyl-MFR-Dehydrogenase spielt – das Eingangsenzym der Methanogenese aus CO₂ und H₂ (siehe 3.II.4).

In *E. coli* wurden zwei Loci identifiziert, die für die Biosynthese von Molybdopterin benötigt werden: *moa* und *moe* (Johnson & Rajagopalan, 1987; Pitterle & Rajagopalan, 1989). Dem *moa*-Locus wurde ein Operon aus fünf Genen (*moaABCDE*) zugeordnet (Rivers *et al.*, 1993): *moaABC* kodieren für die Enzyme, die den ersten spezifischen Schritt der Molybdopterin-Biosynthese katalysieren, die Bildung des Zwischenproduktes Z (Wuebbens & Rajagopalan, 1995); *moaDE* kodieren für die beiden Untereinheiten der Molybdopterin-Synthase, die den letzten Schritt der Biosynthese katalysiert, die Bindung von Dithiolen an das Zwischenprodukt Z (Rudolph *et al.*, 2001). *moaDE*-analoge Gene konnten im Annotationsprogramm ERGO weder in *Msp. stadtmanae* noch in den übrigen methanogenen Archaeen eindeutig identifiziert werden. Ein Aufnahmesystem für Molybdän, das typischerweise aus einem ABC-Transporter besteht, der im *mod*-Locus kodiert ist (Grunden & Shanmugam, 1997), wurde ebenfalls nicht gefunden.

3.II.5.2 *Msp. stadtmanae*-spezifische Gene

Eine Auswahl potentieller Gene, die keine signifikanten bidirektionalen BLAST-Treffer ($> e^{-15}$) zu irgendeinem der bisher veröffentlichten Genome eines Archaeons aufweisen, ist in Tab. 3.24, Tab. 3.25 und Tab. 3.27 dargestellt. Aus einem automatisierten, strikten bidirektionalen Vergleich fallen diejenigen Proteine heraus, die trotz Funktionsgleichheit starke Sequenzunterschiede zeigen oder aber in einem Organismus in mehreren Kopien vorliegen. Aus diesem Grund wurden alle aufgeführten CDSs in ERGO auf ihre tatsächliche Spezifität für *Msp. stadtmanae* hin überprüft.

3.II.5.2.1 Enzyme der Zellwandsynthese

Unter denjenigen CDSs, die keine Ähnlichkeit zu bisher identifizierten Proteinen anderer Archaeen aufweisen, befinden sich eine Reihe potentieller Enzyme der Zellwandsynthese (Tab. 3.24). *rfb*-Cluster wurden im Zusammenhang mit der Synthese der O-Antigen-Komponente der Lipopolysaccharide (LPS) Gram-negativer Pathogene charakterisiert und zeichnen sich für die hohe antigenische Variabilität dieser Oberflächenstrukturen verantwortlich (Reeves, 1993). Vier potentielle *rfb*-Gene in *Msp. stadtmanae* (Msp_0223, Msp_0531-0533) zeigen Übereinstimmungen (41 % bis 56 % Identität auf Proteinebene) zu Enzymen, die in *Salmonella typhimurium* und *Klebsiella pneumoniae* mit der Synthese der Oligosaccharid-Anteile des O-Antigens in Verbindung gebracht werden (Jiang *et al.*, 1991; Kelly & Whitfield, 1996). Acht Glycosyltransferasen sind möglicherweise an einer spezifischen Assemblierung dieser Zuckerreste im O-Antigen beteiligt. Ein weiteres *Msp. stadtmanae*-spezifisches Gen (Msp_0060) weist schwache Identität auf Proteinebene (52 % Identität über ca. 30 % der Sequenz) zu einem Enzym (LicD) aus der Zellwandsynthese des Gram-negativen Humanpathogens *Haemophilus influenzae* auf (Weiser *et al.*, 1997). LicD katalysiert die Übertragung von Phosphoryl-Cholin (ChoP) auf exponierte Hexosereste der LPS (Risberg *et al.*, 1997). Von verschiedenen *licD*-Genen kodierte Enzyme können ChoP auf unterschiedliche Hexosereste übertragen und damit über antigenische Eigenschaften der Zellwand entscheiden (Lysenko *et al.*, 2000). In *H. influenzae* werden vier Gene (*licABCD*) für die Aufnahme von Cholin aus dem Nährmedium, die Phosphorylierung und die Übertragung auf die LPS verantwortlich gemacht (Weiser *et al.*, 1997). Von diesen vier Genen kann nur zu *licD* eine ähnliche CDS in *Msp. stadtmanae* identifiziert werden.

Die gesamte Gruppe der Enzyme aus der Zellwandsynthese in *Msp. stadtmanae*, zu der neben den in Tab. 3.24 dargestellten noch weitere, weniger spezifische Enzyme gehören, ist in ca. einem Drittel des Genoms kodiert, das sich in unmittelbarer Nachbarschaft des Replikationsursprungs befindet und mit dem Bereich des Genoms zusammenfällt, in dem auch die meisten Gene für die Enzyme des methanogenen Stoffwechsels lokalisiert sind (Abb. 3.25).

Tab. 3.24: Kodierende Sequenzen für potentielle Enzyme der Zellwandsynthese aus *Msp. stadtmanae* Gene, die keine bidirektionalen BLAST-Treffer zu anderen Archaeen-Genomen aufweisen. Positionsangaben ausgehend vom Replikationsursprung im Uhrzeigersinn. Ähnlichkeit nach FASTA-Vergleich in ERGO.

CDS	Position [kbp]	Annotation	Ähnlichkeit
Msp_0060	75	putative lipooligosaccharide cholinephosphotransferase	<i>H. influenzae</i> , LicD, sp P14184 ^a : Phosphoryl-Cholin-Transfer (LPS)
Msp_0207	255	predicted glycosyltransferase	
Msp_0215	265	predicted glycosyltransferase	
Msp_0218	270	predicted glycosyltransferase	
Msp_0220	275	predicted glycosyltransferase	
Msp_0223	280	predicted UDP-galactopyranose mutase (EC 5.4.99.9)	<i>K. pneumoniae</i> , RfbD sp Q48481 ^b : O-Antigen-Biosynthese (LPS).
Msp_0442	521	predicted glycosyltransferase	
Msp_0495	584	predicted glycosyltransferase	
Msp_0500	594	predicted glycosyltransferase	
Msp_0531	626	predicted nucleoside-diphosphate-sugar pyrophosphorylase	<i>S. typhimurium</i> , RfbF, sp P26398 ^c : CDP-Abequose-Biosynthese innerhalb der O-Antigen-Biosynthese (LPS).
Msp_0532	627	predicted nucleoside-diphosphate-sugar epimerase	<i>S. typhimurium</i> , RfbG, sp P26398 ^c : CDP-Abequose-Biosynthese innerhalb der O-Antigen-Biosynthese (LPS).
Msp_0533	628	predicted pyridoxal-phosphate-dependent aminotransferase	<i>S. typhimurium</i> , RfbH, sp P26398 ^c : O-Antigen-Biosynthese (LPS).
Msp_0538	633	predicted glycosyltransferase	
Msp_0581	673	predicted UDP-N-acetylmuramyl tripeptide synthase	
Msp_0697	808	predicted methyladenine glycosylase	

^a (Weiser *et al.*, 1997); ^b (Kelly & Whitfield, 1996); ^c (Jiang *et al.*, 1991)

3.II.5.2.2 Überlange offene Leserahmen

Weitere *Msp. stadtmanae*-spezifische CDSs, die keine bidirektionalen BLAST-Treffer in anderen Archaeen-Genomen aufweisen, wurden in einer umfangreichen Gruppe überdurchschnittlich langer offener Leserahmen identifiziert, die kaum Ähnlichkeit zu bekannten Proteinsequenzen, dafür aber starke Übereinstimmungen untereinander zeigen (Tab. 3.25). Zu den veröffentlichten Datenbank-Einträgen, die auf Proteinebene schwache Ähnlichkeit zu einzelnen Bereichen dieser "überlangen ORFs" zeigen, gehö-

ren hypothetische Proteine aus *Parachlamydia* sp., *Mannheimia succiniproducens* MBEL55E, *Escherichia coli* O157:H7, *Plasmodium falciparum* 3D7, *Chloroflexus aurantiacus* und *Erwinia carotovora* subsp. *atroseptica* SCRI1043. Diese Proteine, die mehrheitlich der Zellwand zugeordnet werden, weisen wie die überlangen ORFs eine überdurchschnittliche und uneinheitliche Länge auf. Innerhalb der abgeleiteten Proteinsequenzen der überlangen ORFs konnten keine konservierten Domänen ausgemacht werden.

Tab. 3.25: Gruppe der überlangen ORFs aus *Msp. stadtmanae*. Positionsangaben ausgehend vom Replikationsursprung im Uhrzeigersinn.

Nr.	CDS	Position [kbp]	Länge [AA]	Nr.	CDS	Position [kbp]	Länge [AA]
1	Msp_0043	46	816	20	Msp_0912	1035	2037
2	Msp_0046	51	881	21	Msp_0913	1041	1519
3	Msp_0048	56	1909	22	Msp_0922	1059	973
4	Msp_0141	160	2120	23	Msp_0963	1102	1104
5	Msp_0145	171	2774	24	Msp_0964	1106	769
6	Msp_0154	190	2036	25	Msp_0980	1123	1063
7	Msp_0432	502	2004	26	Msp_0983	1128	1311
8	Msp_0440	515	1552	27	Msp_1091	1234	735
9	Msp_0590	683	2979	28	Msp_1103	1252	857
10	Msp_0606	703	2459	29	Msp_1104	1255	1140
11	Msp_0619	726	2468	30	Msp_1108	1261	1921
12	Msp_0713	820	2151	31	Msp_1109	1267	853
13	Msp_0747	861	706	32	Msp_1115	1281	1896
14	Msp_0762	876	3356	33	Msp_1117	1517	834
15	Msp_0765	890	1222	34	Msp_1399	1568	1447
16	Msp_0776	905	421	35	Msp_1401	1574	1324
17	Msp_0777	907	1875	36	Msp_1465	1632	2214
18	Msp_0836	970	1790	37	Msp_1572	1749	843
19	Msp_0911	1030	1030				

Die überlangen ORFs aus *Msp. stadtmanae* verteilen sich gleichmäßig über das gesamte Genom, treten an einigen Stellen gehäuft, aber nicht geclustert und in keiner erkennbaren, konservierten Umgebung auf. Von insgesamt 37 hypothetischen Proteinen, die mit einer Ausnahme alle länger als 639 Aminosäuren sind, weisen 26 eine Länge von mehr als 1000, 11 sogar von mehr als 2000 Aminosäuren auf. Wie beschrieben (siehe 3.II.1.6) werden durch den niedrigen G+C-Gehalt von 27.6 % des Gesamtgenoms von *Msp. stadtmanae* A/T-reiche STOP-Codons statistisch begünstigt, so daß es sich bei den überlangen ORFs kaum um Artefakte der CDS-Vorhersage handeln dürfte. Zudem konnten weder über automatische Vorhersagen, noch über Sequenzvergleiche alternative, kürzere CDSs in anderen Leserastern innerhalb der überlangen ORFs identifiziert werden.

3.II.5.2.2.1 Zusammensetzung der überlangen ORFs und Proteine

Den überlangen ORFs aus *Msp. stadtmanae* können weitere spezifische Eigenschaften zugeordnet werden, die sie von den restlichen CDSs des Genoms unterscheiden. Aus einem Vergleich aller kodierenden Sequenzen mit einer Länge von mehr als 1000 Aminosäuren mit allen übrigen CDSs konnten folgende Besonderheiten abgeleitet werden:

1. Überlange ORFs weisen mit 27.6 % einen deutlich niedrigeren G+C-Gehalt auf als der Rest der kodierenden Sequenzen (29.4 %).
2. Überlange Proteine enthalten reduzierte Anteile der meisten Aminosäuren zugunsten von Asparagin, Threonin und Serin (Tab. 3.26). Diese Unterschiede weisen keine Korrelation mit dem A+T-Gehalt der jeweiligen Codons auf, die für diese Aminosäuren kodieren.

Tab. 3.26: Aminosäure (AA)-Zusammensetzung aller CDSs mit einer Länge von mehr als 1000 Aminosäuren im Vergleich zu allen übrigen CDSs.

Aminosäure	Anteil (< 1000 AA) [%]	Anteil (> 1000 AA) [%]	Veränderung [%]
Gly	5,9	6,3	+ 7
Asn	6,7	13,4	+ 101
Ile	10,2	11,0	+ 7
Lys	8,3	7,0	- 16
Asp	6,2	5,2	- 16
Gln	2,6	2,2	- 15
Ala	5,3	4,2	- 22
Arg	3,2	1,7	- 47
His	1,8	0,9	- 50
Cys	1,3	0,6	- 52
Thr	6,0	11,8	+ 96
Val	6,3	6,4	+ 2
Trp	0,6	0,4	- 25
Tyr	4,2	5,0	+ 17
Glu	7,1	4,3	- 39
Pro	3,3	2,0	- 39
Leu	8,7	6,1	- 30
Phe	3,7	2,8	- 24
Ser	6,1	7,4	+ 21
Met	2,4	1,2	- 48

3.II.5.2.2.2 Aufbau der überlangen Proteine

Innerhalb der überlangen Proteine konnten weder bekannte konservierte Domänen noch bekannte Leader-Sequenzen oder Schnittstellen identifiziert werden. Den Asparagin-, Threonin- und Serin-Resten innerhalb der überlangen Proteine konnten keine der bekannten Umgebungen zugeschrieben werden, die im Zusammenhang mit posttranslationalen Modifikationen beschrieben worden sind. Auf Basis von Sequenz-Alignments mit verschiedenen überlangen Proteinen konnten allerdings einige gemeinsame strukturelle Merkmale dieser Gruppe ausgemacht werden (Abb. 3.31).

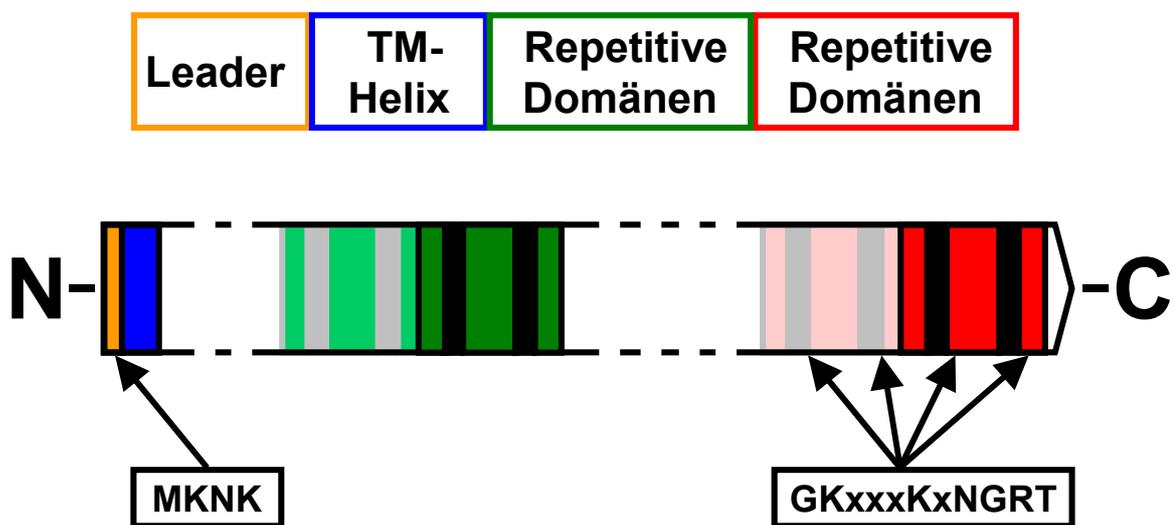


Abb. 3.31: Schematischer Aufbau der überlangen Proteine aus *Msp. stadtmannae*. Abkürzungen: TM-Helix, Transmembran-Helix; N, N-Terminus; C, C-Terminus; M, Methionin; K, Lysin; N, Asparagin; G, Glycin; T, Threonin; R, Arginin; x, beliebige Aminosäure.

- Alle überlangen ORFs kodieren für eine N-terminale potentielle Leader-Sequenz von 4, seltener 5 Aminosäuren, die bevorzugt Lysin (K) und Asparagin (N) enthält und meistens folgende Sequenz aufweist: [Methionin-K-N-K].
- An diese Leader-Sequenz schließt sich eine potentielle Transmembran(TM)-Helix aus 20 bis 21 Aminosäuren an (nach TMpred-Vorhersage; siehe 2.13.3). Dieser Transmembranbereich ist strukturell konserviert, d.h. als Folge hydrophober Aminosäuren, deren Sequenz sich zwischen verschiedenen Proteinen unterscheiden kann. Im Anschluß an die TM-Helix kann keine konservierte potentielle Schnittstelle identifiziert werden, die auf eine posttranslationale Prozessierung hinweisen würde.

- Überlange Proteine weisen eine in Domänen organisierte Architektur auf, in der kurze Sequenzmotive von wenigen Aminosäuren Länge im Abstand von ca. 100 Aminosäuren in unterschiedlicher Anzahl wiederholt werden.
- Alle überlangen Proteine enthalten am C-Terminus in leicht variierender Form mindestens zwei Wiederholungen eines Glycin-reichen kurzen Sequenzmotives [G-K-x-x-K-x-N-G-R-T] (G = Glycin, R = Arginin, T = Threonin, x = beliebige Aminosäure).

3.II.5.2.3 Restriktions-Modifikations-System

Drei Untereinheiten eines potentiellen Typ III Restriktions-Modifikations-Systems (R-M-Systeme) gehören zu denjenigen Genen, die im Genom von *Msp. stadtmanae* im Gegensatz zu allen anderen bisher sequenzierten Archaeen-Genomen identifiziert werden konnten (Tab. 3.27). Diesen Systemen, die in Bakterien weit verbreitet sind, wird im allgemeinen eine Schutzfunktion des Organismus vor wirtsfremder DNA zugeschrieben (Bickle & Kruger, 1993). Darüberhinaus wurden in einer Reihe pathogener Bakterien Methyl-Transferasen aus Typ III R-M-Systemen charakterisiert, die Muster einer phasenvariablen Expression (*phase variation*) zeigen (De Bolle *et al.*, 2000; Saunders *et al.*, 2000). Phasenvariation bezeichnet in diesem Fall die reversible Aktivierung und Deaktivierung der Expression, die durch Mutationen in Sequenzwiederholungen innerhalb des Gens oder im Promotorbereich vor dem Gen vermittelt wird (Weiser *et al.*, 1990). In *Haemophilus influenzae* kann durch die phasenvariable Aktivierung eines Typ III R-M-Systems eine zufällige, aber koordinierte Expression einer ganzen Gruppe von Genen vermittelt werden (Srihanta *et al.*, 2005). Eindeutige Sequenzwiederholungen, durch die auf eine phasenvariable Expression des Typ III R-M-Systems in *Msp. stadtmanae* geschlossen werden könnte, wurden nicht gefunden.

Tab. 3.27: Potentielles *Msp. stadtmanae*-spezifisches Restriktions-Modifikations-System.

CDS	Annotation
Msp_0470	partially conserved hypothetical protein, related to type III restriction-modification system, methylation subunit
Msp_0472	partially conserved hypothetical protein, related to type III restriction-modification system, methylation subunit
Msp_0473	type III restriction-modification system, restriction subunit

4 Diskussion

4.1 Das *Ralstonia eutropha* H16-Genomprojekt

4.1.1 Das *R. eutropha* H16-Genom

Das Genom von *R. eutropha* H16 besteht aus drei Replikons, von denen die beiden größeren strukturell stärkere Ähnlichkeit zueinander zeigen als zu dem Megaplasmid pHG1. Sie unterscheiden sich nur unwesentlich im Anteil der kodierenden Sequenz (ca. 89 %), im G+C-Gehalt (ca. 66 %) und in der Dichte genomischer Inseln (ca. 14 %). Insgesamt sind mehr als 90 % aller Gene von *R. eutropha* H16 auf Chr.1 (4.1 Mbp) und Chr.2 (2.9 Mbp) kodiert. Dennoch gibt es große Unterschiede in den kodierenden Eigenschaften beider Replikons. Nahezu der gesamte Enzymapparat für die genetische Informationsverwaltung und den elementaren Energiestoffwechsel ist auf Chr.1 kodiert. Dazu gehören neben DNA- und RNA-Polymerasen auch Topoisomerasen, Restriktions-Endonukleasen, Transkriptions- und Elongations-Faktoren, eine Gyrase und eine Primase, Cluster für ribosomale Proteine, ein vollständiger Satz tRNAs und drei ribosomale RNA-Operons. Aus dem elementaren Energiestoffwechsel sind auf Chr.1 die Enzyme des Citrat-Cyklus, der meisten Synthese- und Abbau-Wege von Aminosäuren, der Atmungskette und der ATP-Synthese über Elektronentransport-Phosphorylierung kodiert. Ein kleiner Teil dieser Gene liegt in Kopie auf Chr.2, zum Beispiel einzelne Untereinheiten der DNA-Polymerase III, mehrere tRNAs, zwei ribosomale RNA-Operons, ein Elongations-Faktor und mehrere Enzyme des Citrat-Cyklus. Von einem einzigen Enzym aus der Aminosäuresynthese – der Argininosuccinat-Synthase (ArgG; H16_B2531) – wurde bisher gezeigt, das es ausschließlich von Chr.2 kodiert wird. Während also der überwiegende Anteil essentieller, Wachstums-unabhängiger Funktionen in den sogenannten Haushaltsgenen (*housekeeping genes*) auf Chr.1 kodiert ist, steuert Chr.2 in Verbindung mit pHG1 zusätzliche Eigenschaften bei, die *R. eutropha* H16 gegenüber verwandten *Ralstonia*-Stämmen auszeichnen.

4.1.2 Charakterisierung von Chromosom 2 aus *R. eutropha* H16

Hinsichtlich des fakultativ anaeroben und fakultativ lithoautotrophen Stoffwechsel von *R. eutropha* H16 wurde der kodierende Anteil von Chr.2 in Verbindung mit dem Megaplasmid pHG1 bereits ausführlich untersucht (Pfitzner & Schlegel, 1973; Kusian & Bowien, 1997; Lenz *et al.*, 2002; Schwartz *et al.*, 2003). Zusammenfassend und vereinfacht dargestellt würde *R. eutropha* H16 ohne Chr.2 und das Megaplasmid pHG1 die Fähigkeit zu autotrophem (*cbb*-Gene auf Chr.2 und pHG1), zu lithotrophem (Hydrogenasen auf pHG1) und zu anaerobem Wachstum verlieren (Nitrat-, Nitrit- und NO-Reduktase auf Chr.2, Nitrat-, NO- und N₂O-Reduktase auf pHG1).

Ein weiterer beträchtlicher Anteil der Anpassungsfähigkeit von *R. eutropha* H16, der in dieser Arbeit untersucht wurde, geht auf eine Erweiterung des Substratspektrums um zusätzliche Kohlenstoff-, Stickstoff- und Phosphorquellen zurück. Die enzymatischen Grundlagen dieser Funktionen sind auf Chr.2 kodiert. In der Regel werden dazu Stoffwechselwege ausgenutzt und erweitert, die in den Haushaltsgenen auf Chr.1 kodiert sind und zum Basis-Stoffwechsel gehören. Anders verhält es sich beim Kohlenhydrat-Abbau, der in *R. eutropha* H16 ausschließlich über den Entner-Doudoroff-Weg erfolgt und in wesentlichen Bestandteilen von Enzymen katalysiert wird, die nur auf Chr.2 lokalisiert sind (siehe 3.I.4.2.1.1). In diesem Fall wird der auf Chr.2 kodierte Abbau-Weg für Fructose und für Glucose (die allerdings nicht aktiv aufgenommen wird (König *et al.*, 1969)) durch Chr.1 um die Zuckersäuren Gluconat und 2-Keto-Gluconat erweitert. Eine weitere Kohlenstoffquelle, die durch Chr.2 erschlossen wird, ist Acetoin. Außerdem sind die zur Umsetzung von Glucarat und Galactarat nötigen Enzyme, die als Substrate für das Wachstum von *R. eutropha* H16 bisher nicht getestet wurden, auf Chr.2 kodiert (siehe 3.I.4.2.1.2).

R. eutropha H16 ist darüberhinaus in der Lage, eine Vielzahl aromatischer Verbindungen als ausschließliche Kohlenstoff- und Energiequelle zu nutzen und über verschiedene Abbau-Wege umzusetzen (Davis *et al.*, 1969; Johnson & Stanier, 1971). Diese Fähigkeit beruht auf einander ergänzenden kodierenden Eigenschaften der beiden großen Replikons (siehe 3.I.4.2.1.3). Auf Chr.1 ist der Enzymapparat für den Abbau bestimmter Aromaten lokalisiert, die über das Zwischenprodukt Catechol und den β -Keto adipat-Weg umgesetzt werden. Durch die auf Chr.2 kodierten Enzyme wird einerseits der β -Keto adipat-Weg um eine Reihe von Substraten erweitert, die über das Zwi-

schenprodukt Protocatechuat abgebaut werden; andererseits werden mit dem *meta*-Spaltungs- und wahrscheinlich auch dem Gentisat-Weg zusätzliche Umsetzungen für weitere aromatische Verbindungen zur Verfügung gestellt.

Zu den potentiellen Stickstoffquellen, deren Abbau Ammonium freisetzt und von Enzymen katalysiert wird, die auf Chr.2 lokalisiert sind, gehören Cyanat, Formamid und Ethanolamin (siehe 3.I.4.2.2). Zumindest der Cyanat-Dehydratase dürfte dabei neben der Erschließung alternativer Stickstoffquellen eine detoxifizierende Funktion zukommen, da Cyanat aufgrund starker Reaktivität mit nucleophilen Gruppen von Proteinen eine toxische Wirkung ausübt (Carepo *et al.*, 2004). Eine zusätzliche Phosphorquelle wird schließlich durch ein potentielles Operon auf Chr.2 erschlossen, daß *R. eutropha* H16 in die Lage versetzt, chemisch stabile Phosphonate abzubauen (siehe 3.I.4.2.3).

Weitere Eigenschaften, die auf Chr.2 kodiert sind und teilweise zur taxonomischen Unterscheidung von *R. eutropha* H16 und anderen *Ralstonia*-Spezies verwendet werden (Vanechoutte *et al.*, 2004), sind die Fähigkeit zur Fortbewegung über peritriche Begeißelung (siehe 3.I.4.5), die Synthese eines Pyrroloquinolinquinon-Cofaktors (siehe 3.I.4.3) und die Kontrolle der K^+ -Ionenkonzentration über eine P-Typ-ATPase (siehe 3.I.4.4).

Obwohl über eine toxische Wirkung von *R. eutropha* H16 bis heute nichts bekannt ist, konnten auf Chr.2 eine Reihe von Clustern lokalisiert werden, die Ähnlichkeit zu RTX-Operons aufweisen, denen eine Funktion bei der Synthese und Sekretion von Hämolyysin oder ähnlicher RTX-Toxine in verschiedenen pathogenen Organismen zugeschrieben werden (siehe 3.I.4.6.2.1). Allerdings konnte eine hämolytische Wirkung von *R. eutropha* H16 experimentell bisher nicht nachgewiesen werden (A. Pohlmann, persönliche Mitteilung). Wahrscheinlich handelt es sich also bei allen identifizierten RTX-Clustern um kryptische Operons, in denen einzelne Gene für die Synthese und den Transport funktioneller Toxine fehlen bzw. zerstört sind. Auf Chr.2 sind außerdem weitere Enzyme des Sekundärmetabolismus kodiert, beispielsweise für die Synthese eines potentiellen Insekten-Toxins der Tc-Familie, die als Alternative zu *Bacillus thuringiensis*-Toxinen (bt-Toxine) für die Transformation zur Herstellung schädlingsresistenter Nutzpflanzen diskutiert werden (3.I.4.6.2.2). Eine potentielle nicht-ribosomalen Peptid-Synthetase, deren Untereinheiten auf Chr.2 lokalisiert sind, ist möglicherweise an der Expression von Siderophoren beteiligt (siehe 3.I.4.6.1).

4.1.3 Genomorganisation in *R. eutropha* H16

Vergleichende Analysen der Replikons aus *R. eutropha* H16 untereinander geben Grund zu der Annahme, daß Chr.2 sinnvollerweise als Megaplasmid bezeichnet werden sollte. Wie bereits erwähnt ist die überwiegende Mehrheit aller Haushaltsgene auf Chr.1 kodiert (siehe 4.I.1), während Chr.2 die genetischen Informationen für eine Anpassung an speziellere Wachstumsbedingungen enthält. Wie im vorigen Kapitel gezeigt bestehen diese Anpassungen im Fall von *R. eutropha* H16 in der Fähigkeit zu lithoautotrophem und anaerobem Wachstum sowie im wesentlichen einer Erweiterung des Substratspektrums.

Bei der Identifizierung des Replikationsursprungs in *R. eutropha* H16 wurden außerdem Hinweise auf unterschiedliche Mechanismen der Replikationsinitiation gefunden, die stärkere Übereinstimmungen zwischen Chr.2 und pHG1 zeigen als zwischen Chr.1 und Chr.2 (siehe 3.I.2.2). In allen drei Replikons können die *Origins of replication* mittels *GC-Skew*-Analyse eindeutig bestimmt werden. Für Chr.2 und pHG1 fallen diese in die Nähe von potentiellen *repA* und *parAB1*-Genen, die mit der plasmidären Initiation der Replikation bzw. einem plasmidären Partitionierungssystem in Verbindung gebracht werden. Dagegen fällt der Replikationsursprung auf Chr.1 mit einem *dnaA*-Gen zusammen, welches klassischerweise mit der Initiation der chromosomalen Replikation verbunden wird (Messer, 2002).

Bis zum jetzigen Zeitpunkt ist es nicht gelungen, Chr.2 in *R. eutropha* H16 zu deletieren (A. Pohlmann, persönliche Mitteilung). Ein in dieser Weise veränderter Stamm verhielte sich möglicherweise auxotroph in Bezug auf einige Metabolite – eine Vorhersage, die auch für *R. solanacearum* GMI1000 getroffen wurde (Salanoubat *et al.*, 2002).

4.1.4 Replikonvergleiche innerhalb der *Burkholderiales*

Aus der Familie der *Burkholderiaceae* wurden neben *R. eutropha* H16 bisher drei Genome vollständig sequenziert: *R. solanacearum* GMI1000, *Burkholderia mallei* ATCC 23344 (Nierman *et al.*, 2004) und *Bu. pseudomallei* K96243 (Holden *et al.*, 2004). Eine Reihe weiterer Genome, zu denen unter anderem *R. eutropha* JMP134, *R. metallidurans* CH34, *Bu. cepacia* R1808 und *Bu. xenovorans* LB400 gehören, befin-

det sich zur Zeit in einem fortgeschrittenen Stadium der Sequenzierung am DOE Joint Genome Institute, Kalifornien und ist dort öffentlich zugänglich (<http://www.jgi.doe.gov>). Die nächsten vollständig sequenzierten Verwandten von *R. eutropha* H16 außerhalb der Familie der *Burkholderiaceae* gehören zur Gattung *Bordetella* und umfassen *Bo. bronchiseptica* RB50, *Bo. parapertussis* 12833 und *Bo. pertussis* Tohama I (Parkhill *et al.*, 2003). *Cupriavidus necator* N1 wurde kürzlich als naher Verwandter von *R. eutropha* H16 beschrieben, der als Typ-Stamm der Spezies *R. eutropha* fungieren müsse (Vandamme & Coenye, 2004) (siehe 4.I.9). Tab. 4.1 gibt eine Übersicht über die Genomorganisation der genannten *Burkholderiales*-Stämme.

Tab. 4.1: Genomorganisation einiger vollständig sequenzierter und in der Sequenzierung begriffener Stämme aus der Ordnung *Burkholderiales*.

	Genom [Mbp]	1. Replikon [Mbp]	2. Replikon [Mbp]	3. Replikon, etc. [Mbp]
<i>R. eutropha</i> H16	7.4	4.1 (Chr.1)	2.9 (Chr.2)	0.452 (pHG1)
<i>C. necator</i> N1	~ 7.9	~ 3.9	~ 2.7	~ 1.3
<i>R. eutropha</i> JMP134 ^a	7.2	?	?	0.088 (pJP4)
<i>R. metallidurans</i> CH34 ^a	6.8	?	?	pMOL30, pMOL28
<i>R. solanacearum</i> GMI1000 ^b	5.8	3.7 (Chr.)	2.1 (Mpl.)	
<i>Bu. pseudomallei</i> K96243 ^c	7.2	4.1 (Chr.1)	3.2 (Chr.2)	
<i>Bu. mallei</i> ATCC 23344 ^d	5.8	3.5 (Chr.1)	2.3 (Chr.2)	
<i>Bu. cepacia</i> R1808 ^a	8.4	?		
<i>Bu. xenovorans</i> LB400 ^a	9.7	?		
<i>Bo. bronchiseptica</i> RB50 ^e	5.3	5.3		
<i>Bo. parapertussis</i> 12822 ^e	4.8	4.8		
<i>Bo. pertussis</i> Tohama I ^e	4.1	4.1		

R. = *Ralstonia*; *C.* = *Cupriavidus*; *Bu.* = *Burkholderia*; *Bo.* = *Bordetella*

Chr. = Chromosom; Mpl. = Megaplasmid

^a DOE Joint Genome Institute, California, USA; ^b Salanoubat *et al.*, 2002; ^c Holden *et al.*, 2004;

^d Nierman *et al.*, 2004; ^e Parkhill *et al.*, 2003

Gegenüber *R. solanacearum* GMI1000, dem nächsten vollständig sequenzierten Verwandten von *R. eutropha* H16 konnten über bidirektionale BLAST-Vergleiche hohe Übereinstimmungen in den kodierenden Eigenschaften der beiden großen Replikons festgestellt werden (siehe 3.I.3.1). Zu mehr als der Hälfte aller CDSs auf Chr.1 aus *R. eutropha* gibt es bidirektionale BLAST-Treffer auf dem Chromosom aus *R. solanacearum*. Von diesen Treffern liegen nahezu 90 % auf beiden Replikons in konservierter Umgebung vor, also in gleicher oder umgekehrter Reihenfolge der Gene. Übereinstimmende Beobachtungen konnten in einer Ausweitung dieser Vergleiche für alle bisher sequenzierten Genome der Ordnung *Burkholderiales* gemacht werden (Tab. 4.2; Abb. 4.1).

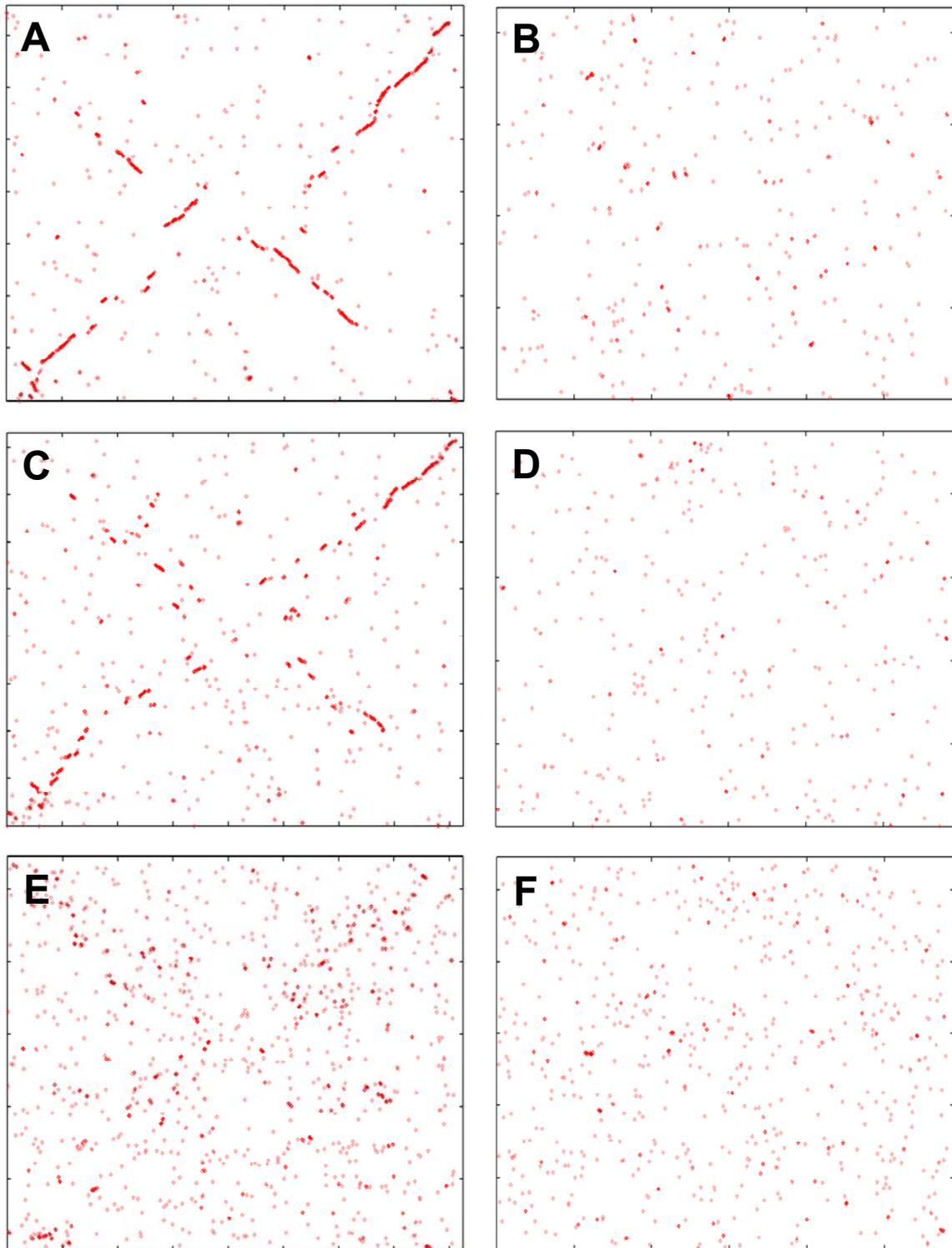


Abb. 4.1: Bidirektionale BLAST-Vergleiche auf Proteinebene ($> e^{-15}$) zwischen Chr.1 und 2 und den Replikons verschiedener *Burkholderiales*-Genome. Abkürzungen: Chr., Chromosom; Mpl., Megaplasmid. **A**, Chr.1 aus *R. eutropha* H16 mit Chr. aus *R. solanacearum* GMI1000; **B**, Chr.2 aus *R. eutropha* H16 mit Mpl. aus *R. solanacearum* GMI1000; **C**, Chr.1 aus *R. eutropha* H16 mit Chr.1 aus *Bu. pseudomallei* K96243; **D**, Chr.2 aus *R. eutropha* H16 mit Chr.2 aus *Bu. mallei* ATCC 23344; **E**, Chr.1 aus *R. eutropha* H16 mit *Bo. bronchiseptica* RB50; **F**, Chr.2 aus *R. eutropha* H16 mit *Bo. bronchiseptica*, RB50. Die Replikons von *R. eutropha* H16 stellen die jeweiligen x-Achse dar.

Tab. 4.2: Bidirektionale BLAST-Vergleiche auf Proteinebene ($> e^{-15}$) zwischen Chr.1 und 2 aus *R. eutropha* H16 und den Replikons verschiedener *Burkholderiales*-Genome.

		Anzahl CDSs	<i>Ralstonia eutropha</i> H16	
			Chr.1	Chr.2
<i>R. solanacearum</i> GMI1000	Chr.	3442	2008	551
	Mpl.	1678	365	465
<i>Bu. pseudomallei</i> K96243	Chr.1	3460	1698	565
	Chr.2	2393	269	247
<i>Bu. mallei</i> ATCC 23344	Chr.1	2996	1541	642
	Chr.2	1768	467	413
<i>Bo. bronchiseptica</i> RB50		4994	1810	861
<i>Bo. parapertussis</i> 12822		4185	1739	785
<i>Bo. pertussis</i> Tohama I		3436	1552	651

Alle verglichenen Genome zeigen eine hohe Zahl bidirektionaler BLAST-Treffer der Chromosomen untereinander und eine erkennbare Konservierung der Abfolge der Gene auf dem Chromosom. Für die mutmaßlichen Megaplasmide können vergleichbaren Übereinstimmungen nicht gefunden werden. In Übereinstimmung mit den taxonomischen Verwandtschaftsverhältnissen nimmt die Anzahl der bidirektionalen BLAST-Treffer zwischen den beiden großen Replikons von *R. eutropha* H16 und den jeweiligen Replikons von *R. solanacearum* über die *Burkholderia*- zu den *Bordetella*-Spezies hin ab. Wie die Vergleiche zeigen, bleibt die Anzahl gemeinsamer Proteine für alle Chromosomen auch zwischen den entfernter verwandten Spezies relativ hoch und ist nicht direkt abhängig von der Anzahl der kodierenden Sequenzen auf dem Chromosom. Interessanterweise enthält das mutmaßliche Megaplasmid aus *Bu. mallei* ATCC 23344 deutlich mehr gemeinsame CDSs mit Chr.2 aus *R. eutropha* H16 als das entsprechende Replikon aus *Bu. pseudomallei* K96243 und als der Gesamtzahl der CDSs nach zu erwarten wäre.

4.1.5 Genomorganisation innerhalb der *Burkholderiaceae*

Alle bisher sequenzierten Spezies der *Burkholderiaceae* weisen Gemeinsamkeiten in der genomischen Organisation auf, die in Analogie zu *R. eutropha* H16 eine Unterscheidung der verschiedenen Replikons nach Chromosom und Megaplasmid erlauben:

1. In *R. solanacearum* GMI1000, *Bu. mallei* ATCC 23344 und *Bu. pseudomallei* K96243 wurde die Mehrheit aller Haushaltsgene auf dem größten Replikon identifiziert, während auf dem kleineren Replikon Eigenschaften für spezifische Anpassungen der Organismen an unterschiedliche Habitats kodiert sind (Salanoubat *et al.*, 2002; Holden *et al.*, 2004; Nierman *et al.*, 2004). Innerhalb des *Burkholderia cepacia*-Komplexes, zu dem *Bu. xenovorans* LB400 und *Bu. cepacia* R1808 gehören, wird eine ähnliche Funktionsaufteilung zwischen mehreren Replikons angenommen (Mahenthiralingam *et al.*, 2005).
2. In jedem der drei vollständig sequenzierten Genome wurden außerdem Hinweise auf unterschiedliche, replikonspezifische Mechanismen der Replikation gefunden: *dnaA*-homologe Gene auf dem großen Replikon, *repA*- und *parAB1*-homologe Gene auf dem kleineren Replikon.

In *Bu. mallei* ATCC 233344 und *Bu. pseudomallei* K96243 wurden allerdings ebenso wie in *R. eutropha* H16 Gene aus dem Basis-Stoffwechsels auf dem kleineren Replikon identifiziert.

4.1.6 Verwandtschaftsverhältnisse mutmaßlicher Megaplasme

Um Hinweise auf mögliche Verwandtschaftsverhältnisse zwischen den mutmaßlichen Megaplasmen der *Burkholderiaceae*-Genome zu sammeln, wurden Sequenzvergleiche zwischen verschiedenen potentiellen RepA-Proteinen durchgeführt, die den vollständig annotierten Datensätzen entnommen bzw. durch BLAST-Vergleiche in den unvollständigen Datensätzen identifiziert wurden. Von insgesamt 16 RepA-analogen Proteinen, die über Sequenz-Alignments miteinander verglichen wurden (Abb. 4.2), stammen 13 aus acht unterschiedlichen *Burkholderiaceae*-Stämmen: je ein *repA*-Gen von insgesamt sieben mutmaßlichen Megaplasmen aus *R. eutropha* H16 (Chr.2, pHG1), *R. metallidurans* CH34 (pMOL28, pMOL30), *R. solanacearum* GMI1000 (Mpl.), *Bu. mallei* ATCC 233344 (Chr.2) und *Bu. pseudomallei* K96243 (Chr.2). pMOL30 und pMOL28 wurden in *R. metallidurans* CH34 als Schwermetallresistenz vermittelnde Megaplasme charakterisiert (Mergeay *et al.*, 1985). Fünf weitere RepA-analoga Proteine wurden in den unvollständigen Datensätzen von *R. eutropha* JMP134, *Bu. cepacia* R1808, *Bu. xenovorans* LB400 (je ein *repA*-Gen) und *R. metallidurans*

CH34 (zwei *repA*-Gene) identifiziert. In den Vergleich miteinbezogen wurden außerdem RepA-analoge Proteine von zwei Plasmiden mit breitem Wirtsspektrum (pSB102, pIP02T), ein RepA-Protein aus einem *E. coli*-Plasmid (R388) und ein weiteres RepA-analoges Protein, das in *Bu. pseudomallei* K96243 auf dem großen Replikon identifiziert worden ist und möglicherweise auf ein ins Chromosom inseriertes Plasmid hinweist.

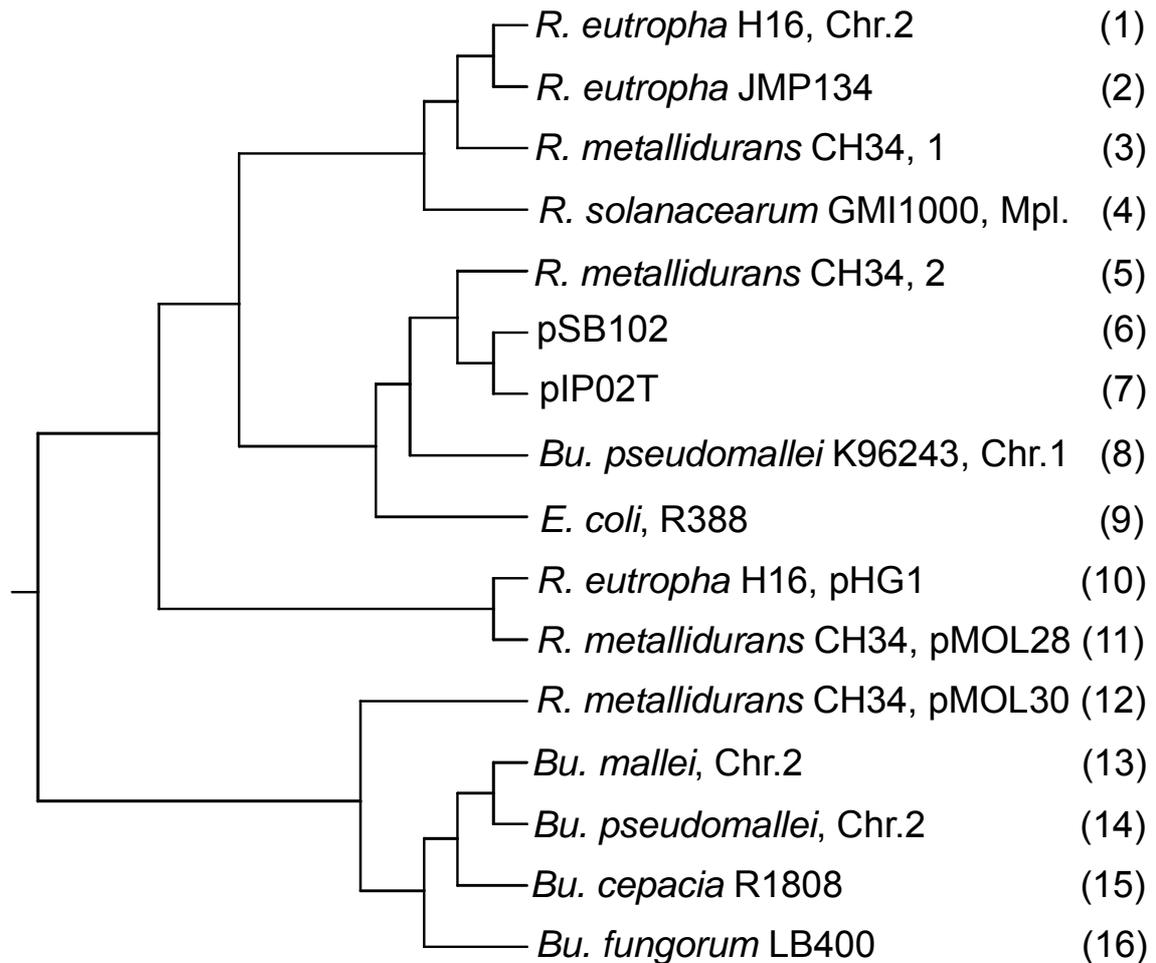


Abb. 4.2: Bewurzelter phylogenetischer Stammbaum basierend auf Sequenzvergleichen mutmaßlicher RepA-Proteine aus Genomen verschiedener Stämme der *Burkholderiaceae*, aus einem *E. coli*-Plasmid und aus unspezifischen Plasmiden mit breitem Wirtsspektrum (pSB102 und pIP02T). (1) H16_B0001; (2) gi|53761021; (3) gi|48768367; (4) gi|17548222; (5) unbenannte CDS; (6) gi|15722254; (7) gi|16610030; (8) gi|53720880; (9) gi|419569; (10) gi|38638033; (11) gi|56410265; (12) gi|56130630; (13) gi|53716049; (14) gi|53723369; (15) gi|46321161; (16) gi|48782664. Sequenzvergleiche wurden mit DIALIGN 2 durchgeführt (Morgenstern, 1999) und mit DRAWGRAM aus dem Phylip-Paket dargestellt (Felsenstein, 1989). Die Länge der Äste im Diagramm ist nicht gleichzusetzen mit dem Verwandtschaftsgrad der verbundenen Sequenzen.

Wie dieser Vergleich belegt, enthalten alle *Burkholderiaceae*-Genome neben dem Chromosom ein mutmaßliches Megaplasmid, dessen Verwandtschaftsgrade – anhand von RepA-Sequenzvergleichen bestimmt – gute Übereinstimmungen mit der taxonomischen Einteilung der Genome in die *Ralstonia*- und *Burkholderia*-Spezies zeigen. Dementsprechend können diese Megaplasmitide, zu denen auch Chr.2 aus *R. eutropha* H16 gehört, als gattungsspezifisch bezeichnet werden. Daneben enthalten *R. eutropha* H16 und *R. metallidurans* CH34 mit pHG1 und pMOL28 weitere Megaplasmitide, die nach RepA-Vergleich nah miteinander verwandt sind, aber wenig Ähnlichkeit zu Chr.2, pMOL30 und den übrigen Megaplasmitiden anderer *Burkholderiaceae* zeigen. Das Genom von *R. metallidurans* CH34 setzt sich demnach aus mindestens vier Megaplasmitiden zusammen: einem gattungsspezifischen Megaplasmitid, pMOL28, das mit pHG1 verwandt ist, pMOL30, das Ähnlichkeit mit den Megaplasmitiden der *Burkholderia*-Spezies aufweist und einem weiteren Megaplasmitid, das verwandt ist mit einer Gruppe um das *E. coli*-Plasmitid R388.

4.1.7 Gemeinsame Abstammung verschiedener Megaplasmitide

Auf jedem der gattungsspezifischen Megaplasmitide der vier in Tab. 4.1 dargestellten *Ralstonia*-Spezies (bzw. auf dem Contig des dazugehörigen *repA*-Gens) wurden Gen-Cluster identifiziert, die für die Eingangsenzyme des Kohlenhydrat-Abbaus über den Entner-Doudoroff-Weg kodieren (Abb. 4.3). Diese Cluster weisen einen Grad an Konservierung auf, der sie von vergleichbaren Gen-Clustern in den Genomen der *Burkholderia*-Spezies unterscheidet. Demzufolge kann auf ein gemeinsames Ursprungsplasmitid der gattungsspezifischen Megaplasmitide der *Ralstonia*-Spezies geschlossen werden, auf dem bereits die genetischen Grundlagen für den Entner-Doudoroff-Weg angelegt waren.

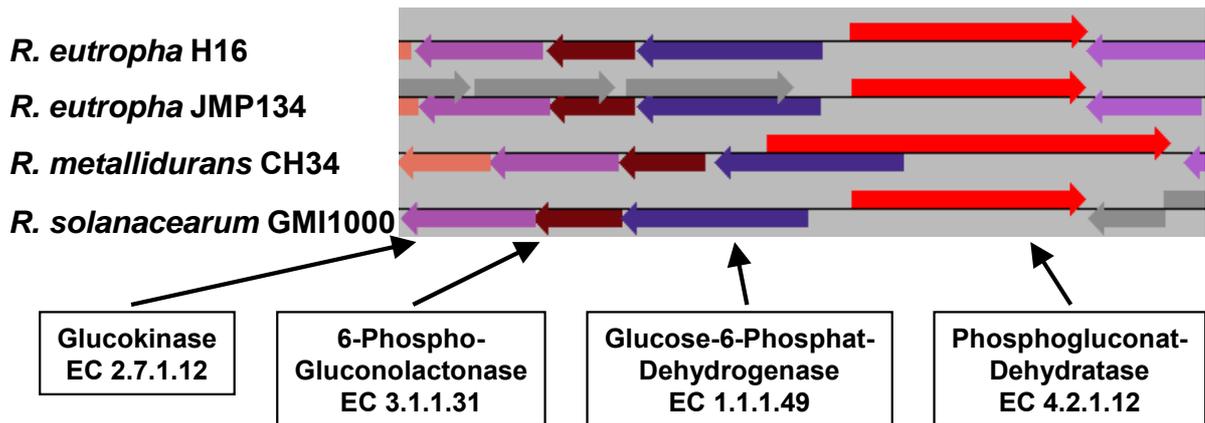


Abb. 4.3: Vergleich der potentiellen Gene für Enzyme des Entner-Doudoroff-Weges in einem konservierten Cluster auf den Megaplasmiden verschiedener *Ralstonia*-Spezies.

4.1.8 Hypothetische Entwicklung der *Burkholderiales*-Genome

Aus dem Vergleich verschiedener Genome aus der Familie der *Burkholderiaceae* kann ein allgemeines Organisationsprinzip abgeleitet werden, das möglicherweise Hinweise auf die evolutionsgeschichtliche Entwicklung dieser Familie gibt. Alle *Burkholderia*-Genome enthalten ein Chromosom, das für die Grundlagen der Informationsverwaltung und den Basis-Stoffwechsel kodiert und entsprechend hohe Übereinstimmungen in der kodierenden Sequenz und deren Verteilung auf dem Chromosom zwischen verschiedenen Spezies aufweist. Dieses Replikon könnte vom Chromosom eines gemeinsamen Vorläuferbakteriums aller *Burkholderiaceae* abstammen. Aus dem Genom dieses ursprünglichen Bakteriums, das keine weiteren Replikons enthielt, könnten sich die einzelnen Gattungen durch Aufnahme unterschiedlicher, kleiner Plasmide differenziert haben. Auf diesen Plasmiden könnten die genetischen Grundlagen einer Spezialisierung auf bestimmte Wachstumsbedingungen angelegt gewesen sein, die heute zum Beispiel den *Ralstonia*-Spezies den Abbau von Kohlenhydraten über den Entner-Doudoroff-Weg ermöglichen (siehe 4.1.7). Daß Sequenzvergleiche von RepA-Proteinen der großen mutmaßlichen Megaplasmide noch immer Gattungszugehörigkeiten widerspiegeln (Abb. 4.2), spricht für eine entwicklungsgeschichtlich lange zurückliegende Aufnahme dieser Plasmide in das Ursprungsgenom. Für einen Vorläuferorganismus der *Burkholderiaceae* hätte es sich als nützlich erwiesen, eine ausgeprägte genomische Plastizität zu entwickeln, um das Plasmid als Experimentierfeld zur Aufnahme und Überprüfung zu-

sätzlicher Eigenschaften zu nutzen. Als Folge dieser Plastizität könnte es im Laufe der Zeit zu einer Verlagerung von Haushaltsgenen auf die Megaplasmide gekommen sein, mit dem Ergebnis, daß heute essentielle Stoffwechselfgene auf den Megaplasmiden kodiert sind. Genduplikationen und die Aufnahme fremden genetischen Materials über lateralen Gentransfer hätten danach die Megaplasmide auf die heutige Größe anwachsen lassen. In diesem Sinne stellen die Megaplasmide der *Burkholderiaceae* ein genetisch variables Reservoir nützlicher Eigenschaften dar, aus dem sich der Organismus unter widrigen oder veränderten Lebensumständen bedient.

In der *Ralstonia*-Gattung könnte es in einem Vorläuferorganismus von *R. eutropha* H16 und *R. metallidurans* CH34 zur Aufnahme eines weiteren Plasmides gekommen sein, aus dem sich pHG1 und pMOL28 entwickelt haben (Abb. 4.2). In *R. metallidurans* CH34, einem Stamm, der ebenso wie *R. eutropha* H16 lithoautotroph wachsen kann, sind die dazu nötigen Hydrogenase-Gene auf dem Chromosom kodiert (Mergeay *et al.*, 1985). Dennoch könnten die Grundlagen für die Fähigkeit zu lithoautotrophen Wachstum auf einem Vorläuferplasmid von pHG1 und pMOL28 angelegt gewesen sein. Daß in *R. eutropha* H16 die genetischen Grundlagen des fakultativ lithoautotrophen und des fakultativ anaeroben Stoffwechsels auf pHG1 und Chr.2 verteilt vorliegen, spricht möglicherweise dafür, daß eine langfristige Verlagerung genetischer Ressourcen von den kleineren auf die größeren Megaplasmide oder auf das Chromosom stattfindet. Diese Verlagerung könnte nach dem Prinzip erfolgen, vorteilhafte kodierende Eigenschaften von neu erworbenen auf stabilere Replikons zu transferieren.

In den *Bordetella*-Spezies könnte eine Anpassung an spezielle Wachstumsbedingungen durch direkte Integration fremder DNA in ein Ursprungschromosom stattgefunden haben. Eine gegenüber den Chromosomen der *Ralstonia*- und *Burkholderia*-Spezies erhöhte Genomgröße und die hohe Anzahl bidirektionaler BLAST-Treffer, die in der gleichen Größenordnung liegt wie die Summe der Treffer von Chromosom und Megaplasmid der anderen Genome, deuten darauf hin, daß die Entwicklung dieser Genome nicht über ein zweites Replikon stattgefunden hat oder daß ein zweites Replikon nur in einem Zwischenstadium existierte. Potentielle kryptische *repA*-Gene, die auf in das Chromosom inserierte Plasmide hinweisen würden, konnten in den Genomen der *Bordetella*-Spezies nicht gefunden werden.

4.1.9 Bemerkungen zur taxonom. Einteilung der *Burkholderiaceae*

Innerhalb der letzten Jahre hat eine taxonomische Neuordnung des *Ralstonia*-Genus stattgefunden, in deren Folge *R. eutropha* und *R. metallidurans* der Gattung *Cupriavidus* zugeschrieben wurden (Vandamme & Coenye, 2004; Vaneechoutte *et al.*, 2004). Den Veröffentlichungen nach handelt es sich bei *R. eutropha* um ein Synonym des schon früher beschriebenen *Cupriavidus necator* (Makkar & Casida, 1987). Demnach ordnen sich die Genome der *Ralstonia*-Spezies aus Tab. 4.1 jetzt zwei verschiedenen Gattungen zu: *Cupriavidus* (*C. necator* N1, *C. necator* H16, *C. necator* JMP134, *C. metallidurans* CH34) und *Ralstonia* (*R. solanacearum* GMI1000).

Da bisher keine Informationen über die Genomorganisation von *C. necator* N1 im Vergleich zu *R. eutropha* H16 vorliegen, wurden beide Genome durch Pulsfeld-Gelelektrophorese parallel aufgetrennt und miteinander verglichen (Abb. 4.4; S. Voll-land, unveröffentlicht).

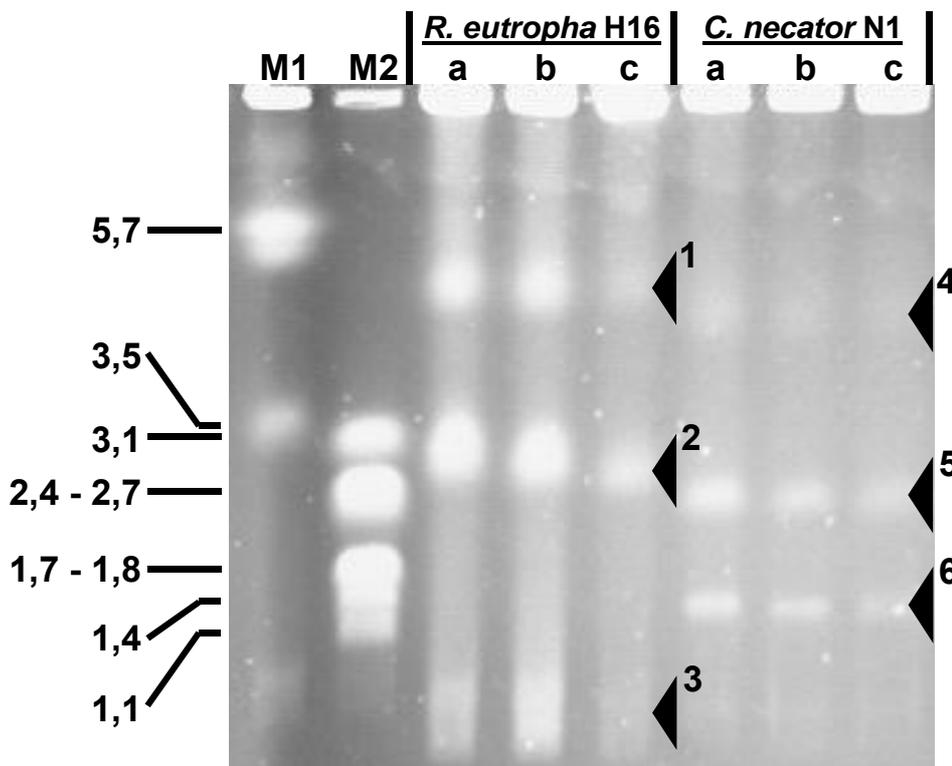


Abb. 4.4: Auftrennung genomischer DNA der Stämme *R. eutropha* H16 und *C. necator* N1 durch Pulsfeld-Gelelektrophorese. Es wurden jeweils drei gleiche Ansätze aufgetragen (*R. eutropha* H16 und *C. necator* N1: a,b,c). Abkürzungen: M1 und M2, Längenmarker. Größenangaben in Mbp. Zuordnung der DNA-Banden: 1, Chr.1 aus *R. eutropha* H16; 2, Chr.2 aus *R. eutropha* H16; 3, pHG1 aus *R. eutropha* H16; 4, mutmaßliches Chromosom aus *C. necator* N1; 5, mutmaßliches Megaplasmid 1 aus *C. necator* N1; 6, mutmaßliches Megaplasmid 2 aus *C. necator* N1.

Innerhalb der Familie der *Burkholderiaceae* verteilen sich die kodierende Eigenschaften des Genoms auf mehrere Replikons, die unterschiedlichen Ursprungs sein können und verschiedene Verwandtschaftsbeziehungen zeigen. Dadurch wird die taxonomische Klassifizierung dieser Organismen erheblich erschwert. An der jetzigen Einteilung erscheint zumindest die Trennung der Stämme *R. eutropha* H16 und *R. solanacearum* GMI1000 problematisch, weil sie die hohen Übereinstimmungen zwischen den Chromosomen ebenso wie den gemeinsamen Ursprung der potentiell gattungsspezifischen Megaplasmide vernachlässigt. Die Zuordnung von *R. eutropha* zur Spezies *Cupriavidus necator* sollte aufgrund der deutlichen Unterschiede in der Größe der Replikons, aus denen sich die Genome von *R. eutropha* H16 und *C. necator* N1 zusammensetzen, noch einmal überprüft werden.

4.1.10 Ausblick

Mit dem Verständnis der Genomorganisation der *Burkholderiaceae*, aus der sich die außerordentliche metabolische Vielseitigkeit dieser Familie speist, werden sich in Zukunft Ansätze für die Entwicklung von Hochleistungsstämmen ableiten lassen. Schließlich kann dabei auf ein umfangreiches genetisches Potential zurückgegriffen werden, das sich aus der Gesamtheit der verschiedenen Megaplasmide dieser Familie zusammensetzt. Möglicherweise ließen sich Eigenschaften unterschiedlicher Megaplasmide zu einer Art *Ralstonia*-Bibliothek verbinden, in der ein Basischromosom mit unterschiedlichen Megaplasmiden kombiniert würde, um aus Eigenschaften wie Autotrophie, Schwermetallresistenz oder Substratspezifität einen maßgeschneiderten Stamm zusammenzustellen. Dazu müßte zunächst die genaue Funktion der mutmaßlichen Megaplasmide und Chromosomen ebenso wie deren gesamtes kodierendes Potential erfaßt und verstanden werden. In *R. eutropha* H16 sollte dieses Verständnis in absehbarer Zeit durch eine Genomanalyse, bei der auf ausführliche experimentelle Daten zurückgegriffen werden kann, verfügbar sein. Zur Zeit wird an einer abschließenden Annotation und Charakterisierung von Chr.1 gearbeitet. Im Anschluß daran ist die Herstellung von DNA-Chips geplant, die eine Analyse der Transkription des gesamten Genoms von *R. eutropha* H16 unter verschiedenen Wachstumsbedingungen ermöglichen werden.

4.II Das *Methanosphaera stadtmanae*-Genomprojekt

4.II.1 Das *Methanosphaera stadtmanae*-Genom

Mit *Msp. stadtmanae* wurde das erste kommensale methanogene Archaeon vollständig sequenziert und annotiert. Von allen bisher sequenzierten Archaeen weist das Genom von *Msp. stadtmanae* den niedrigsten G+C-Gehalt (28 %) und die höchste Anzahl ribosomaler RNA-Operons (vier) auf. Gleichzeitig besitzt es die niedrigste Anzahl kodierender Sequenzen (1545 CDSs) von allen bisher sequenzierten Methanogenen (1687 CDSs bis 4540 CDSs). Die Genomgröße von *Msp. stadtmanae* (1.77 Mbp) liegt zwischen den Genomgrößen der *Methanosarcina*-Spezies (4.1 Mbp bis 5.75 Mbp) und der übrigen methanogenen Archaeen (1.66 Mbp bis 1.75 Mbp). Der Anteil der kodierenden Sequenz an der Gesamtsequenz (84 %) fällt höher aus als in den *Methanosarcina*-Spezies (74 % bis 75 %) und niedriger als in den übrigen methanogenen Archaeen (88 % bis 90 %).

4.II.2 Der methanogene Stoffwechsel

Im anaeroben Milieu des menschlichen Enddarms hat sich *Msp. stadtmanae* auf die Verwertung von Gärungsprodukten spezialisiert, die aus der anaeroben Fermentation von Nahrungsmittelresten, die vom Menschen nicht umgesetzt werden können, durch nicht-methanogene Bakterien stammen (Wolin & Miller, 1994). Dabei tritt *Msp. stadtmanae* in direkte Konkurrenz zu *Methanobrevibacter smithii*, dem häufigsten methanogenen Archaeon im menschlichen Enddarm, das für nahezu die gesamte Methanproduktion im Gastrointestinaltrakt verantwortlich ist (Miller & Wolin, 1982). Im Gegensatz zu *Mbr. smithii* hat sich *Msp. stadtmanae* auf einen methylotrophen Stoffwechsel spezialisiert, bei dem in analoger Weise zu den *Methanosarcina*-Spezies (Deppenmeier, 2004) Methanol mit molekularem Wasserstoff zu Methan reduziert wird (Miller & Wolin, 1985). Die CDSs der an diesen Reaktionen beteiligten Enzyme, einschließlich einer Untereinheit der ATP-Synthase, weisen eine Anpassung der *Codon*

Usage auf, die sie als hochexprimiert charakterisiert und die elementare Funktion dieser Enzyme im Stoffwechsel von *Msp. stadtmanae* belegt (siehe 3.II.2.3).

Die endogene Produktion von Methanol, das nahezu vollständig aus dem Abbau von Pectin stammt, liegt im menschlichen Gastrointestinaltrakt zwischen 0.3 und 0.6 Gramm pro Tag (Lindinger *et al.*, 1997). Pectin, ein Methyl-Ester polymerisierter Zuckersäuren, stellt einen Hauptbestandteil pflanzlichen Zellwände dar und wird im Darm durch pectinolytische Bakterien, insbesondere der Gattung *Bacteroides*, unter Abspaltung von Methanol umgesetzt (Salyers *et al.*, 1977). *In vivo*-Analysen weisen auf stark erhöhte Methanolkonzentrationen im menschlichen Körper insbesondere nach dem Verzehr von Früchten hin (Lindinger *et al.*, 1997). Beispielsweise entspricht der Anstieg der Methanolkonzentration im Atem nach Verzehr von einem Kilogramm Äpfel dem Genuß von 0.3 Litern 40 %igem Branntwein.

Zu einer methanogenen Reduktion von CO₂ ist *Msp. stadtmanae* im Gegensatz zu *Mbr. smithii* und den meisten anderen bisher sequenzierten methanogenen Archaeen nicht in der Lage. Dafür ist wahrscheinlich ein Mangel an Molybdopterin verantwortlich, das weder synthetisiert, noch aufgenommen werden kann (siehe 3.II.5.1). Molybdopterin stellt jedoch einen essentiellen Cofaktor für die Formyl-Methanofuran-Dehydrogenase dar, das Eingangsenzym der Methanogenese aus H₂ und CO₂ (Vorholt & Thauer, 1997). Zu allen weiteren Enzymen des CO₂-Reduktionsweges, einschließlich der Formyl-MFR-Dehydrogenase, konnten in *Msp. stadtmanae* analoge kodierende Sequenzen nachgewiesen werden (siehe 3.II.4). Überraschenderweise lassen sich anhand der Enzyme des methanogenen Stoffwechsels keine eindeutigen Verwandtschaftsverhältnisse von *Msp. stadtmanae* zu anderen methanogenen Archaeen ausmachen. Während die Mehrheit der CDSs die stärksten Übereinstimmungen in Sequenz und Organisation zu *Mth. thermotrophicus* zeigt, weichen einige Cluster von diesem Muster ab und entsprechen in der Reihenfolge der Gene dem *Methanosarcina*-Typ oder weichen von allen bisher beschriebenen Formen ab. Möglicherweise gehen diese Unterschiede in der Organisation der Gene auf eine dynamische Plastizität des Genoms von *Msp. stadtmanae* zurück, die sich in regelmäßigen intragenomischen Rekombinationsereignissen äußert (siehe 4.II.4).

4.II.3 Kommensalismus

Bis zu 10^{12} kommensale Mikroorganismen leben in jedem Gramm Darminhalt des Menschen (Mackie *et al.*, 1999). Wie Experimente mit Mäusen belegen, die unter keimfreien Bedingungen gehalten wurden, üben kommensale Mikroorganismen eine positive Wirkung auf die Ausbildung der Darmschleimhaut und des Immunsystems aus (Crabbe *et al.*, 1970; Benveniste *et al.*, 1971). Zu einer Störung des empfindlichen Gleichgewichts zwischen Bakterium und Wirt kann es jedoch kommen, wenn die kommensale Darmflora als pathogen wahrgenommen und durch das Immunsystem bekämpft wird – chronisch entzündliche Darmerkrankungen wie *Morbus Crohn* und *Colitis Ulcerosa* (*inflammatory bowel diseases*) sind die Folge (Macdonald & Monteleone, 2005). Da sich die Grenzen zwischen pathogenen und kommensalen Organismen in dieser Weise verschieben können, ist Kommensalismus auch als dynamisches Wettrüsten als Folge einer Koevolution bezeichnet worden, die Kommensale und Wirt davor bewahrt einander Schaden zuzufügen (Steinert *et al.*, 2000; Hooper & Gordon, 2001). Kommensale ebenso wie pathogene Organismen weisen allgemeine Oberflächenstrukturen auf (*pathogen associated molecular patterns*, PAMPs), die von Rezeptoren des menschlichen Immunsystems (*Toll-like receptors*, TLRs) erkannt werden (Macpherson & Uhr, 2004). Dementsprechend werden den Prozessen, die zur Ausbildung eines pathogenen oder eines symbiontischen Verhältnisses führen, ähnliche molekulare Mechanismen zugesprochen (Hentschel *et al.*, 2000). Ein gemeinsames Merkmal dieser Mechanismen ist die Erzeugung phänotypischer Diversität, durch die sich der kommensale oder pathogene Organismus maskiert und damit dem Immunsystem des Wirtes dauerhaft entzieht (Deutsch *et al.*, 1997). Durch genetische Reorganisationen, die zusammenfassend als Phasenvariation (*phase variation*) bezeichnet werden, wird innerhalb bestimmter Loci, die für antigenisch relevante Oberflächenstrukturen kodieren, eine lokal begrenzte Hypermutabilität erzeugt. Zu den genetischen Grundlagen dieser Hypermutabilität gehören Insertionen und Deletionen einzelner Basen oder kurzer Sequenzabschnitte innerhalb kodierender Sequenzen oder Promotorregionen.

Über eine Identifikation derjenigen Proteine, die keine Entsprechung in den Genomen anderer Archaeen aufweisen, konnten mehrere Gruppen von Proteinen identifiziert werden, die möglicherweise an der Erzeugung phänotypischer Diversität beteiligt sind oder auf andere Weise für die Etablierung des kommensalen Verhältnisses von

Msp. stadtmanae im menschlichen Enddarm benötigt werden. Spezifische Enzyme der Zellwandsynthese deuten darauf hin, daß *Msp. stadtmanae* Oberflächenstrukturen ausbildet, die Ähnlichkeiten zur O-Antigen-Komponente der Lipopolysaccharide Gram-negativer Bakterien aufweisen (siehe 4.II.3.1). Bei diesen Strukturen könnte es sich um Oligosaccharidketten handeln, die über N- und O-glykosidische Bindungen mit Asparagin-, Threonin- und Serin-Resten membranassoziierter Proteine aus der Gruppe der überlangen Proteine verknüpft sind (siehe 4.II.3.2). Intragenomische Rekombinationseignisse könnten ein Reservoir variierender überlanger ORFs erzeugt haben, deren variable Expression für die Erzeugung antigenischer Variabilität verantwortlich ist. Die Regulation dieser paralogenen Gene könnte über reversible DNA-Methylierungen durch Restriktions-Modifikations-Systeme stattfinden.

4.II.3.1 Zellwandsynthese

Msp. stadtmanae gehört zur Familie der *Methanobacteriaceae* (Miller & Wolin, 1985), die sich durch Zellwände mit hohen Anteilen von Pseudomurein auszeichnen (König *et al.*, 1982). Eine Glycoproteinschicht vom Typ der bakteriellen *S-layer* – vergleichbar mit den Lipopolysacchariden Gram-negativer Bakterien (Whitfield, 1995) – kann in *Msp. stadtmanae* nicht nachgewiesen werden (H. König, persönliche Mitteilung).

Die meisten Enzyme der Zellwandsynthese von *Msp. stadtmanae* sind in einem Bereich des Genoms kodiert, der an den Replikationsursprung grenzt (Abb. 3.25). Darunter befinden sich einige Enzyme, die keine signifikante Ähnlichkeit zu bekannten Proteinen anderer Archaeen aufweisen, dafür aber zu Enzymen der Zellwandsynthese Gram-negativer, vorwiegend pathogener Bakterien (siehe 3.II.5.2.1). Dort werden sie mit der Synthese und Assemblierung von Zuckerresten im O-Antigen, der äußeren, strukturell hochvariablen Komponente der Lipopolysaccharide (LPS) in Verbindung gebracht. Die chemische Zusammensetzung der O-Antigene in Gram-negativen Bakterien ist stammspezifisch, kann jedoch auch innerhalb einer einzigen Kultur eine große Heterogenität erreichen (Lerouge & Vanderleyden, 2002). In *Salmonella*-Spezies sind die Enzyme der O-Antigen-Synthese in *rfb*-Clustern kodiert, deren genetische Zusammensetzung für die hohe strukturelle Variabilität dieser Strukturen verantwortlich ist (Jiang *et al.*, 1991). LPS und darin besonders der O-Antigen-Komponente werden in pathogenen Organismen unterschiedliche Funktionen in wenigstens zwei Stadien des In-

fektionsprozesses zugeschrieben: Haftung an spezifische Gewebetypen während der Besiedelung und dauerhafte Überwindung von Abwehrmechanismen des Wirtes (Lerouge & Vanderleyden, 2002). Da sich *Msp. stadtmanae* durch die Existenz von *rfb*-analogen Genen von allen anderen bisher sequenzierten Archaeen unterscheidet, könnte diesen Genen eine ähnlich Funktion bei der Etablierung eines kommensalen Verhältnisses im menschlichen Darm zufallen.

In diesem Zusammenhang könnte eine in *Msp. stadtmanae* identifizierte CDS eine Rolle spielen, die in Gram-negativen Bakterien für ein Enzym kodiert, das die Übertragung von Phosphoryl-Cholin (ChoP) auf die LPS katalysiert (siehe 3.II.5.2.1). In vielen vorwiegend pathogenen, Gram-positiven und Gram-negativen Bakterien kann ChoP in der Zellwand nachgewiesen werden (Mosser & Tomasz, 1970; Gillespie *et al.*, 1996; Weiser *et al.*, 1998a). ChoP-Resten werden in der Pathogenese Funktionen für die Haftung und Invasion von Endothel- und Epithel-Zellen durch Wechselwirkungen mit dem Rezeptor des *platelet activating factors* zugeschrieben (Cundell *et al.*, 1995; Ring *et al.*, 1998). Da ChoP-Epitope allerdings auch Sensitivität gegenüber der serumvermittelten Zelltötung durch das menschliche Immunsystem vermitteln können (Weiser *et al.*, 1998b), werden Mechanismen der Phasenvariation diskutiert, mit deren Hilfe die Übertragung von ChoP in unterschiedlichen Stadien der Besiedelung reversibel deaktiviert werden kann (Serino & Virji, 2002). Innerhalb der Gattung *Neisseria* wurden verschiedene pathogene und kommensale Stämme charakterisiert, die sich in der Lokalisation von ChoP-Epitopen an unterschiedlichen Strukturen Zellwand unterscheiden (Serino & Virji, 2000). Während ChoP in kommensalen Stämmen durch die Cholin-Phosphotransferase auf LPS übertragen wird und sich ausschließlich dort nachweisen lässt, wird es in pathogenen *Neisseria*-Stämmen durch andere enzymatische Systeme übertragen und lässt sich in diesem Fall nur an exponierten Pilusstrukturen nachweisen. Obwohl in *Msp. stadtmanae* mit Ausnahme der Cholin-Phosphotransferase keine weiteren der in *H. influenzae* für Aufnahme, Synthese und Übertragung von ChoP charakterisierten Enzyme nachgewiesen wurden (Weiser *et al.*, 1997), könnte ChoP-Resten eine Funktion bei der Adhesion von *Msp. stadtmanae* an spezifische Darm-Epithelzellen im menschlichen Enddarm zukommen.

4.II.3.2 Überlange ORFs

Msp. stadtmanae wendet ca. 10 % seiner Genomsequenz für die Kodierung einer Gruppe von Proteinen auf, deren Funktion für den Organismus unbekannt ist. Die überlangen ORFs sind ohne erkennbare Ordnung oder konservierte Umgebung über das gesamte Genom verteilt und kodieren für Proteine von 800 bis 3400 Aminosäuren Länge, die schwache aber eindeutige Ähnlichkeit zu Proteinen unbekannter Funktion aus vorwiegend Gram-negativen Pathogenen aufweisen (siehe 3.II.5.2.2). Im Vergleich zum restlichen Genom zeichnen sich die überlangen ORFs durch einen unterdurchschnittlichen G+C-Gehalt, die überlangen Proteine durch eine veränderte Aminosäurezusammensetzung und eine aus kurzen, sich wiederholenden Aminosäuremustern zusammengesetzte Struktur aus (Abb. 3.31). Zu den strukturellen Merkmalen der überlangen Proteine gehören eine kurze Leader-ähnliche Sequenz, ein potentieller N-terminaler Membrananker, eine konservierte repetitive Sequenz aus etwa 50 Aminosäuren am C-Terminus und ein im Vergleich zum Rest des Genoms stark erhöhter Anteil der Aminosäuren Asparagin, Threonin und Serin. Weder bereits charakterisierte Proteindomänen, noch bekannte Leader-Sequenzen oder andere Aminosäuremuster, die auf eine posttranslationale Modifikation hinweisen würden, können innerhalb der Proteinsequenzen nachgewiesen werden. Zu einer möglichen Funktion des relativ konservierten C-terminalen Musters gibt es keine Informationen.

Wenn *Msp. stadtmanae* für die Etablierung eines kommensalen Verhältnisses im menschlichen Enddarm tatsächlich auf eine antigenische Variabilität von Oberflächenstrukturen angewiesen ist, um sich der Immunreaktion des Wirtes zu entziehen, so spricht einiges dafür, daß dazu die Gruppe der "überlangen ORFs" die genetischen Grundlagen liefert:

- Mit insgesamt 37 CDSs, die mehr als 10 % der kodierenden Sequenz von *Msp. stadtmanae* umfassen, bilden die überlangen ORFs einen umfangreichen Fundus ähnlicher, aber nicht identischer Proteine. Durch Expression verschiedener überlanger ORFs können Proteine von unterschiedlicher Länge, unterschiedlicher Anzahl konservierter Sequenzmuster und leicht variierender Aminosäurezusammensetzung generiert und so möglicherweise eine hohe phänotypische Diversität erzeugt werden.

- Der Regulation der Expression alternativer Loci, wie sie die überlangen ORFs darstellen, kann durch verschiedene Mechanismen bewerkstelligt werden, die unter dem Begriff Phasenvariation zusammengefaßt werden. Daran können reversible Methylierungen von Promotorbereichen durch ein Typ III Restriktions-Modifikations-System beteiligt sein (Srikhanta *et al.*, 2005), das außer in *Msp. stadtmanae* in keinem anderen Archaeon nachgewiesen wurde (siehe 3.II.5.2.3). In *Neisseria gonorrhoeae* werden antigenische Variationen von Pilusstrukturen durch intragenomische Rekombinationen erzeugt, in denen homologe, aber nicht identische Sequenzen durch nicht-reziproke Gen-Konversion ausgetauscht werden (Hagblom *et al.*, 1985; Koomey *et al.*, 1987).
- Die potentielle N-terminale Leader-Sequenz und anschließende Transmembranregion zeigen Übereinstimmungen mit Charakteristika der Sec-vermittelten Proteintranslokation über die Cytoplasmamembran (Pohlschröder *et al.*, 2004). Demnach besitzen Signalsequenzen der Klasse I eine N-terminale, positiv geladene, ein bis fünf Aminosäuren lange, nicht konservierte Signalsequenz, gefolgt von einer hydrophoben Domäne aus mindestens sechs Aminosäuren, gefolgt von einer ungeladenen polaren Domäne aus drei bis sieben Aminosäuren (von Heijne, 1990). Die positiven Ladungen der N-terminalen Leader-Sequenz interagieren wahrscheinlich mit negativ geladenen Phospholipiden an der cytoplasmatischen Seite der Membran und vermitteln auf diese Weise die Orientierung des Peptids während der Translokation (Lee & Bernstein, 2001). Die polare ungeladene dritte Domäne, zu der keine Entsprechung in den überlangen Proteinen aus *Msp. stadtmanae* existiert, wird von einer Signalpeptidase des Typs I erkannt und geschnitten (Albers & Driessen, 2002). Möglicherweise werden die überlangen Proteine über einen Sec-ähnlichen Mechanismus über die Cytoplasmamembran transportiert, bleiben jedoch über einen potentiellen N-terminalen Transmembran-Anker, der nach der Translokation nicht abgeschnitten wird, an die Membran assoziiert.
- Die Domänenstruktur der überlangen Proteine, in der kurze Sequenzmotive im Abstand von ca. 100 Aminosäuren wiederholt werden, ermöglicht einerseits die Generierung zusätzlicher Variabilität durch Veränderungen in der Anzahl von Wiederholungen und kann andererseits Zielsequenzen bereitstellen für posttranslationale Prozessierungen.

- Mit Asparagin, Threonin und Serin liegen in den überlangen Proteinen gerade die drei Aminosäuren stark konzentriert vor, deren Substituierung durch N- und O-verknüpfte Glycosylierung insbesondere in pathogenen Bakterien gut dokumentiert ist (Benz & Schmidt, 2002; Szymanski & Wren, 2005). Den Oligosaccharidresten der Glycoproteine werden Funktionen für die stabile Haftung an spezifische Gewebetypen, für die Invasion von Wirtszellen und als Antigene der Immunantwort zugeschrieben (Dobos *et al.*, 1996; Benz & Schmidt, 2001). An der Synthese und Assemblierung der Oligosaccharide und an der Verknüpfung mit Asparagin (N-Verknüpfung) bzw. mit Serin/Threonin (O-Verknüpfung) sind Enzyme beteiligt, die ursprünglich der LPS-Biosynthese zugeschrieben wurden (Szymanski *et al.*, 1999). Möglicherweise üben die in *Msp. stadtmanae* identifizierten spezifischen Enzyme der Zellwandsynthese eine ähnliche Funktion aus (siehe 4.II.3.1).

4.II.4 Spezifische Anpassungen des *Msp. stadtmanae*-Genoms

Im Vergleich zu anderen Archaeen und darunter besonders den methanogenen, zeigt das Genom von *Msp. stadtmanae* einige spezifische Anpassungen, die in enger Beziehung zu seinem besonderen Habitat zu sehen sind. Einerseits gewährleistet das Wachstum im menschlichen Enddarm eine Nährstoffversorgung, die *Msp. stadtmanae* eine Reduktion des Stoffwechsels und damit der genetischen Ressourcen erlaubt, die sich in der im Vergleich zu allen bisher sequenzierten methanogenen Archaeen niedrigsten Anzahl CDSs ausdrückt. Andererseits sind mit dem Wachstum in dieser potentiell feindlichen Umgebung Anpassungen verbunden, die sich aus der Auseinandersetzung mit dem menschlichen Immunsystem ergeben.

Die Reduktion der genetischen Ressourcen zeigt sich darin, daß *Msp. stadtmanae* im Gegensatz zu anderen methanogenen Archaeen die Fähigkeit zu autotrophem Wachstum, zur Methanogenese aus CO₂ und H₂, ebenso wie zur Synthese einiger Cofaktoren und Aminosäuren verloren hat. *Msp. stadtmanae* fehlt der CO-Dehydrogenase/Acetyl-CoA-Synthase-Komplex; ein Molybdopterin-Cofaktor kann nicht synthetisiert werden. Aus dem Medium müssen mindestens Ammonium, Acetat, Kohlendioxid, Thiamin, Biotin und Isoleucin aufgenommen werden (Miller & Wolin, 1985). Für die Auseinandersetzung mit dem menschlichen Immunsystem wendet *Msp. stadtmanae* wahrscheinlich deutlich mehr als ein Zehntel der gesamten Genomsequenz auf, die für spezifische

Enzyme der Zellwandsynthese und die Gruppe der überlangen Proteine kodieren. Diese Synthese dürfte einen Großteil der Stoffwechselaktivität von *Msp. stadtmanae* binden und erhebliche Anforderungen an die Transkriptions- und Translations-Maschinerie und dazu an die Energieversorgung dieser Enzymapparate stellen. Dementsprechend könnte die im Vergleich zu anderen Archaeen hohe Anzahl ribosomaler RNA-Operons die Infrastruktur für die gesteigerte Proteinsynthese-Aktivität bereitstellen, wohingegen sich der erhöhte Energiebedarf in einer Anpassung der *Codon Usage* der an der Methanogenese beteiligten Enzyme und der ATP-Synthase ausdrückt, die insgesamt als hochexpressiert charakterisiert wurden. In diesem Zusammenhang ist es nicht überraschend, daß die von der DSMZ, Braunschweig bezogene Stammkultur von *Msp. stadtmanae* eine stabile genomische Variante enthielt, in der eines der ribosomalen RNA-Operons durch eine Deletion in der 23S rRNA zerstört war. Da in Kultur keine Auseinandersetzung mit dem menschlichen Immunsystem und damit auch kein Bedarf an gesteigerter Syntheseaktivität besteht, stellt eine verminderte Ribosomenkonzentration unter diesen Bedingungen keinen Selektionsnachteil dar.

Möglicherweise belegt die Deletion innerhalb des ribosomalen RNA-Operons ebenso wie die beobachteten Abweichungen in der Anordnung einzelner Gene innerhalb konservierter Cluster eine gesteigerte Plastizität des *Msp. stadtmanae*-Genoms im Vergleich zu anderen methanogenen Archaeen. Diese dynamische Struktur könnte die Aufnahme fremder DNA durch lateralen Gentransfer begünstigt haben und damit die Grundvoraussetzung für eine Anpassung von *Msp. stadtmanae* an das besondere Habitat darstellen. Einen Ursprung dieser fremden Gene, die für die Ausbildung besonderer Oberflächenstrukturen verantwortlich sein könnten, stellen möglicherweise Gram-negative, pathogene Bakterien dar. Weiterhin könnte sich eine genomische Plastizität, die zur Duplikation von Genen und zur verstärkten Mutabilität durch Insertionen und Deletionen einzelner Basen oder kurzer Sequenzabschnitte führt, bei der Ausbildung variabler Oberflächenstrukturen als nützlich erwiesen haben. Bei der umfangreiche Gruppe der überlangen ORFs, die untereinander starke Homologien zeigen und dennoch variable Strukturen aufweisen, könnte es sich um das Ergebnis dieser genomischen Flexibilität handeln. Gleichzeitig würde die dynamische Struktur des *Msp. stadtmanae*-Genoms die gegenüber anderen methanogenen Archaeen leicht erhöhte Genomgröße in Verbindung mit einer geringeren Dichte der kodierenden Sequenzen erklären. Durch die Aufnahme fremder DNA würde die Größe des Genoms zunehmen, aufgrund vermehrter

Mutationsereignisse aber gleichzeitig ein Teil der kodierenden offenen Leserahmen verlorengehen.

Möglicherweise deutet schließlich der offensichtliche Widerspruch zwischen einem hochspezialisierten kleinen Genom auf der einen Seite und der Verschwendung kodierender Kapazitäten für überflüssige Enzyme des CO₂-Reduktions-Weges auf der anderen Seite darauf hin, daß es sich bei *Msp. stadtmanae* um einen jungen Organismus handelt, der die Fähigkeit, CO₂ zu Methan zu reduzieren, erst vor evolutionsgeschichtlich kurzer Zeit verloren hat. Veränderte Ernährungsgewohnheiten des Menschen – zum Beispiel durch vermehrt pflanzliche Nahrung – könnten zu einer erhöhten Produktion von Methanol im Enddarm und zur Spezialisierung von *Msp. stadtmanae* auf eine eingeschränkte Methanogenese aus Methanol und H₂ geführt haben.

4.II.5 Ausblick

Obwohl infektiöse Erkrankungen des Gastrointestinaltraktes in den westlichen Industriestaaten weitgehend kontrolliert werden, nimmt die Anzahl allergischer und entzündlicher Störungen dramatisch zu (Macdonald & Monteleone, 2005). Chronisch entzündliche Darmerkrankungen wie *Morbus Crohn* werden mit einer Reizung des Immunsystems durch an sich harmlose kommensale Mikroorganismen in Verbindung gebracht (Podolsky, 2002). Bis heute konnte allerdings nicht befriedigend geklärt werden, durch welche Mechanismen das menschliche Immunsystem zwischen kommensalen und pathogenen Organismen unterscheidet und worin die Ursachen einer Störung des sensiblen kommensalen Gleichgewichtes bestehen (Macpherson & Uhr, 2004). Die Analyse des *Msp. stadtmanae*-Genom könnte in diesem Zusammenhang wertvolle Erkenntnisse liefern und unter Umständen therapeutische Ansätze für die Behandlung entzündlicher Darmerkrankungen liefern. Mit der Identifikation *Msp. stadtmanae*-spezifischer Gene, die möglicherweise an der Expression antigenisch variabler Oberflächenstrukturen beteiligt sind, konnten vielversprechende Kandidaten für weiterführende Untersuchungen gefunden werden, die die Charakterisierung von Deletionsmutanten in einem *Msp. stadtmanae*-Darm-Modell einschließen müßten.

5 Zusammenfassung

5.1 Das *Ralstonia eutropha*-Genomprojekt

1. Die Chromosomen 1 und 2 des Genoms von *R. eutropha* H16 wurden vollständig sequenziert und einer eingehenden Analyse unterzogen. Chr.1 besteht aus 4.053 Mbp mit ca. 3882 CDSs, Chr.2 aus 2.912 Mbp mit ca. 2654 CDSs und das bereits sequenzierte Megaplasmid pHG1 (Schwartz *et al.*, 2003) aus 0.452 Mbp mit 420 CDSs. Hinsichtlich des G+C-Gehaltes (66 %) und der Dichte kodierender Sequenzen (88 %) und genomischer Inseln (14 %) unterscheiden sich Chr.1 und Chr.2 kaum, weichen aber deutlich von pHG1 ab (G+C-Gehalt: 62 %, Dichte kodierender Sequenzen: 81 %, Dichte genomischer Inseln: 33 %).
2. In den kodierenden Eigenschaften läßt sich ein Schwerpunkt von Chr.1 für Komponenten des Basis-Stoffwechsels und der Informationsverwaltung feststellen, wohingegen Chr.2 und pHG1 für Enzyme kodieren, die dem Organismus eine Erweiterung des Substratspektrums und Anpassung an besondere Wachstumsbedingungen ermöglichen. In Bezug auf diese Eigenschaften ist Chr.2 weniger spezifisch als pHG1.
3. Auf Chr.2 sind Schlüsselkomponenten des fakultativ chemolithoautotrophen und anaeroben Stoffwechsels und der Umsetzung von Kohlenhydraten über den Entner-Doudoroff-Weg kodiert sowie verschiedene Abbauwege aromatischer Kohlenwasserstoffe, der gesamte Beißelungsapparat und eine Reihe potentieller Toxin-Cluster.
4. In Bezug auf Mechanismen der Replikationsinitiation weist Chr.2 Charakteristika eines Megaplasmid auf, die in Verbindung mit Unterschieden in den kodierenden Eigenschaften zwischen Chr.1 und Chr.2 für einen plasmidären Ursprung dieses Replikons sprechen.
5. Die Zusammensetzung des *R. eutropha*-Genoms aus einem Chromosom und mehreren Megaplasmiden findet ihre Entsprechung in anderen *Burkholderiaceae*-

Genomen und läßt auf ein gemeinsames Vorläufergenom schließen, das aus einem einzigen Chromosom bestand und sich durch Aufnahme eines Plasmides und Integration lateral transferierter DNA zu den Genomen der heutigen *Burkholderiaceae*-Spezies differenziert hat.

5.II Das *Methanosphaera stadtmanae*-Genomprojekt

1. Das vollständig sequenzierte Genom von *Msp. stadtmanae* besteht aus einem Chromosom von 1 767 405 Mbp Länge, das von allen bisher sequenzierten Archaeen den niedrigsten G+C-Gehalt (28 %) und die höchste Anzahl ribosomaler RNA-Operons (vier) aufweist. Im Vergleich zu anderen sequenzierten Methanogenen enthält das Genom von *Msp. stadtmanae* die niedrigste Anzahl kodierender Sequenzen (1545 CDSs).
2. Die CDSs für die Enzyme der Methanogenese aus Methanol und H₂ sowie für eine Untereinheit der ATP-Synthase zeigen eine Anpassung der *Codon Usage*, die auf eine effiziente Translation hinweist und diese CDSs als hochexprimiert charakterisiert.
3. Die Reduktion des Stoffwechsels von *Msp. stadtmanae* im Vergleich zu anderen Methanogenen kann auf fehlende CDSs für einen CO-Dehydrogenase/Acetyl-CoA-Synthase-Komplex und für die Molybdopterin-Cofaktor-Biosynthese zurückgeführt werden. Allerdings weist das Genom von *Msp. stadtmanae* CDSs für den gesamten Enzymapparat des CO₂-Reduktionsweges auf.
4. Für die Expression potentieller Gene, die bisher in keinem anderen Archaeon nachgewiesen worden waren und damit wahrscheinlich für das kommensale Wachstum im menschlichen Enddarm benötigt werden, wendet *Msp. stadtmanae* mehr als 10 % der kodierenden Sequenz und möglicherweise einen Großteil seiner Stoffwechselaktivität auf. Zu diesen Genen gehören CDSs, die auf Proteinebene Ähnlichkeiten zu Enzymen aus der O-Antigen-Synthese der Lipopolysaccharide Gram-negativer Pathogene aufweisen.

5. Eine umfangreiche und in sich homologe Gruppe *Msp. stadtmanae*-spezifischer CDSs von überdurchschnittlicher Länge und besonderer Aminosäurezusammensetzung kodiert möglicherweise für variable Oberflächenstrukturen, die durch Substituierung an Asparagin, Threonin- und Serin-Resten in ähnlicher Weise wie die Lipopolysaccharide Gram-negativer Pathogene glykosyliert werden könnten, um *Msp. stadtmanae* antigenische Variabilität und damit Schutz vor dem menschlichen Immunsystem zu gewährleisten.

6 Literaturverzeichnis

Albers, S. V. & Driessen, A. M. (2002). Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey. *Arch Microbiol* **177**, 209-216.

Alex, L. A., Reeve, J. N., Orme-Johnson, W. H. & Walsh, C. T. (1990). Cloning, sequence determination, and expression of the genes encoding the subunits of the nickel-containing 8-hydroxy-5-deazaflavin reducing hydrogenase from *Methanobacterium thermoautotrophicum* delta H. *Biochemistry* **29**, 7237-7244.

Altarriba, M., Merino, S., Gavin, R., Canals, R., Rabaan, A., Shaw, J. G. & Tomas, J. M. (2003). A polar flagella operon (flg) of *Aeromonas hydrophila* contains genes required for lateral flagella expression. *Microb Pathog* **34**, 249-259.

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

Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S. & Locht, C. (2003). Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *J Bacteriol* **185**, 1470-1474.

Aspelin, A. L. (1997). Pesticide industry sales and usage: 1994 and 1995 market estimates. Washington, D.C.: U.S. EPA. Office of Prevention, Pesticides and Toxic Substances. Office of Pesticide Programs. Biological and Economic Analysis Division.

Badger, J. H. & Olsen, G. J. (1999). CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16**, 512-524.

Baliga, N. S., Bonneau, R., Facciotti, M. T., Pan, M., Glusman, G., Deutsch, E. W., Shannon, P., Chiu, Y., Weng, R. S., Gan, R. R., Hung, P., Date, S. V., Marcotte, E., Hood, L. & Ng, W. V. (2004). Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* **14**, 2221-2234.

Bartha, R. (1962). [Physiological studies on the chemolithotropic metabolism of recently isolated *Hydrogenomonas* strains.]. *Arch Mikrobiol* **41**, 313-350.

- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004).** The Pfam protein families database. *Nucleic Acids Res* **32**, D138-141.
- Benveniste, J., Lespinats, G., Adam, C. & Salomon, J. C. (1971).** Immunoglobulins in intact, immunized, and contaminated axenic mice: study of serum IgA. *J Immunol* **107**, 1647-1655.
- Benz, I. & Schmidt, M. A. (2001).** Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol Microbiol* **40**, 1403-1413.
- Benz, I. & Schmidt, M. A. (2002).** Never say never again: protein glycosylation in pathogenic bacteria. *Mol Microbiol* **45**, 267-276.
- Bertram, P. A., Schmitz, R. A., Linder, D. & Thauer, R. K. (1994).** Tungstate can substitute for molybdate in sustaining growth of *Methanobacterium thermoautotrophicum*. Identification and characterization of a tungsten isoenzyme of formylmethanofuran dehydrogenase. *Arch Microbiol* **161**, 220-228.
- Bickle, T. A. & Kruger, D. H. (1993).** Biology of DNA restriction. *Microbiol Rev* **57**, 434-450.
- Blackkolb, F. & Schlegel, H. G. (1968).** [Catabolic repression and enzyme repression by molecular hydrogen in *Hydrogenomonas*]. *Arch Mikrobiol* **62**, 129-143.
- 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., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997).** The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1474.
- Bowen, D., Rocheleau, T. A., Blackburn, M., Andreev, O., Golubeva, E., Bhartia, R. & ffrench-Constant, R. H. (1998).** Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* **280**, 2129-2132.
- Bowien, B. & Kusian, B. (2002).** Genetics and control of CO₂ assimilation in the chemoautotroph *Ralstonia eutropha*. *Arch Microbiol* **178**, 85-93.
- Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M. J., Haft, D. H., Nelson, W. C., Davidsen, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Utterback, T., Van Aken, S. E., Feldblyum, T. V., D'Ascenzo, M., Deng, W. L., Ramos, A. R., Alfano, J. R., Cartinhour, S., Chatterjee, A. K., Delaney, T. P., Lazarowitz, S. G., Martin, G. B., Schneider, D. J., Tang, X., Bender, C. L., White, O., Fraser, C. M.**

- & Collmer, A. (2003).** The complete genome sequence of the Arabidopsis and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc Natl Acad Sci U S A* **100**, 10181-10186.
- 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., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Gload, A., Scott, J. L., Geoghagen, N. S. & Venter, J. C. (1996).** Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058-1073.
- Burgdorf, T., Bommer, D. & Bowien, B. (2001).** Involvement of an unusual mol operon in molybdopterin cofactor biosynthesis in *Ralstonia eutropha*. *J Mol Microbiol Biotechnol* **3**, 619-629.
- Burgin, A. B., Parodos, K., Lane, D. J. & Pace, N. R. (1990).** The excision of intervening sequences from *Salmonella* 23S ribosomal RNA. *Cell* **60**, 405-414.
- Burian, J., Ausio, J., Phipps, B., Moore, S., Dougan, D. & Kay, W. (2003).** Hexamerization of RepA from the *Escherichia coli* plasmid pKL1. *Biochemistry* **42**, 10282-10287.
- Carepo, M. S., Azevedo, J. S., Porto, J. I., Bentes-Sousa, A. R., Batista Jda, S., Silva, A. L. & Schneider, M. P. (2004).** Identification of *Chromobacterium violaceum* genes with potential biotechnological application in environmental detoxification. *Genet Mol Res* **3**, 181-194.
- Chang, G. W. & Chang, J. T. (1975).** Evidence for the B12-dependent enzyme ethanolamine deaminase in *Salmonella*. *Nature* **254**, 150-151.
- Chang, Y. F., Young, R. & Struck, D. K. (1991).** The *Actinobacillus pleuropneumoniae* hemolysin determinant: unlinked appCA and appBD loci flanked by pseudogenes. *J Bacteriol* **173**, 5151-5158.
- Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L. & Walsh, C. T. (1990).** Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psiD) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *J Biol Chem* **265**, 4461-4471.
- Coenye, T. & Vandamme, P. (2003).** Diversity and significance of Burkholderia species occupying diverse ecological niches. *Environ Microbiol* **5**, 719-729.
- Coenye, T., Falsen, E., Vancanneyt, M., Hoste, B., Govan, J. R., Kersters, K. & Vandamme, P. (1999).** Classification of *Alcaligenes faecalis*-like isolates from the environment and human clinical samples as *Ralstonia gilardii* sp. nov. *Int J Syst Bacteriol* **49 Pt 2**, 405-413.

Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol Microbiol* **47**, 1495-1512.

Conrad, R. (1996). Soil microorganisms as controllers of atmospheric trace gases (H₂, CO, CH₄, OCS, N₂O, and NO). *Microbiol Rev* **60**, 609-640.

Crabbe, P. A., Nash, D. R., Bazin, H., Eyssen, H. & Heremans, J. F. (1970). Immunohistochemical observations on lymphoid tissues from conventional and germ-free mice. *Lab Invest* **22**, 448-457.

Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkila, I. & Tuomanen, E. I. (1995). Streptococcus pneumoniae anchor to activated human cells by the receptor for platelet-activating factor. *Nature* **377**, 435-438.

Davis, D. H., Doudoroff, M., Stanier, R. Y. & Mandel, M. (1969). Proposal to reject the genus *Hydrogenomonas*: Taxonomic implications. *Int J Syst Bacteriol* **19**, 375-390.

Davis, D. H., Stanier, R. Y., Doudoroff, M. & Mandel, M. (1970). Taxonomic studies on some gram negative polarly flagellated "hydrogen bacteria" and related species. *Arch Mikrobiol* **70**, 1-13.

De Baere, T., Steyaert, S., Wauters, G., Des Vos, P., Goris, J., Coenye, T., Suyama, T., Verschraegen, G. & Vanechoutte, M. (2001). Classification of *Ralstonia pickettii* biovar 3/'thomasi' strains (Pickett 1994) and of new isolates related to nosocomial recurrent meningitis as *Ralstonia mannitolytica* sp. nov. *Int J Syst Evol Microbiol* **51**, 547-558.

De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W. & Moxon, E. R. (2000). The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol* **35**, 211-222.

De Felice, M., Esposito, L., Pucci, B., De Falco, M., Rossi, M. & Pisani, F. M. (2004). A CDC6-like factor from the archaea *Sulfolobus solfataricus* promotes binding of the mini-chromosome maintenance complex to DNA. *J Biol Chem* **279**, 43008-43012.

de Poorter, L. M., Geerts, W. G., Theuvenet, A. P. & Keltjens, J. T. (2003). Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in *Methanothermobacter thermautotrophicus*. *Eur J Biochem* **270**, 66-75.

Deitsch, K. W., Moxon, E. R. & Wellems, T. E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* **61**, 281-293.

- del Solar, G. & Espinosa, M. (2000).** Plasmid copy number control: an ever-growing story. *Mol Microbiol* **37**, 492-500.
- Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. (1999).** Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636-4641.
- Deppenmeier, U. (2002a).** Redox-driven proton translocation in methanogenic Archaea. *Cell Mol Life Sci* **59**, 1513-1533.
- Deppenmeier, U. (2002b).** The unique biochemistry of methanogenesis. *Prog Nucleic Acid Res Mol Biol* **71**, 223-283.
- Deppenmeier, U. (2004).** The membrane-bound electron transport system of Methanosarcina species. *J Bioenerg Biomembr* **36**, 55-64.
- Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., Bruggemann, H., Lienard, T., Christmann, A., Bomeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H. P., Gunsalus, R. P., Fritz, H. J. & Gottschalk, G. (2002).** The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea. *J Mol Microbiol Biotechnol* **4**, 453-461.
- Ding, Y. H., Zhang, S. P., Tomb, J. F. & Ferry, J. G. (2002).** Genomic and proteomic analyses reveal multiple homologs of genes encoding enzymes of the methanol:coenzyme M methyltransferase system that are differentially expressed in methanol- and acetate-grown Methanosarcina thermophila. *FEMS Microbiol Lett* **215**, 127-132.
- Dobos, K. M., Khoo, K. H., Swiderek, K. M., Brennan, P. J. & Belisle, J. T. (1996).** Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of Mycobacterium tuberculosis. *J Bacteriol* **178**, 2498-2506.
- Doolittle, W. F. (1999).** Phylogenetic classification and the universal tree. *Science* **284**, 2124-2129.
- Eisen, J. A., Heidelberg, J. F., White, O. & Salzberg, S. L. (2000).** Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol* **1**, RESEARCH0011.
- Ellermann, J., Hedderich, R., Bocher, R. & Thauer, R. K. (1988).** The final step in methane formation. Investigations with highly purified methyl-CoM reductase (component C) from Methanobacterium thermoautotrophicum (strain Marburg). *Eur J Biochem* **172**, 669-677.

- Epstein, W. (2003).** The roles and regulation of potassium in bacteria. *Prog Nucleic Acid Res Mol Biol* **75**, 293-320.
- Evguenieva-Hackenberg, E. & Selenska-Pobell, S. (1995).** Variability of the 5'-end of the large subunit rDNA and presence of a new short class of rRNA in Rhizobiaceae. *Lett Appl Microbiol* **21**, 402-405.
- Ewing, B. & Green, P. (1998).** Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194.
- Faust, L. R., Connor, J. A., Roof, D. M., Hoch, J. A. & Babior, B. M. (1990).** Cloning, sequencing, and expression of the genes encoding the adenosylcobalamin-dependent ethanolamine ammonia-lyase of *Salmonella typhimurium*. *J Biol Chem* **265**, 12462-12466.
- Felmlee, T., Pellett, S. & Welch, R. A. (1985).** Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. *J Bacteriol* **163**, 94-105.
- Felsenstein, J. (1989).** Phylip -- Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166.
- Ferry, J. G. (1995).** CO dehydrogenase. *Annu Rev Microbiol* **49**, 305-333.
- Ferry, J. G. (1997).** Enzymology of the fermentation of acetate to methane by *Methanosarcina thermophila*. *Biofactors* **6**, 25-35.
- Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B. & Puhler, A. (2001).** The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci U S A* **98**, 9889-9894.
- Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I. & Miller, J. H. (2002).** Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci U S A* **99**, 984-989.
- Florin, T. H., Zhu, G., Kirk, K. M. & Martin, N. G. (2000).** Shared and unique environmental factors determine the ecology of methanogens in humans and rats. *Am J Gastroenterol* **95**, 2872-2879.
- Fraser, G. M., Bennett, J. C. & Hughes, C. (1999).** Substrate-specific binding of hook-associated proteins by FlgN and FliT, putative chaperones for flagellum assembly. *Mol Microbiol* **32**, 569-580.

- Fraser, J. A., Davis, M. A. & Hynes, M. J. (2001).** The formamidase gene of *Aspergillus nidulans*: regulation by nitrogen metabolite repression and transcriptional interference by an overlapping upstream gene. *Genetics* **157**, 119-131.
- Friedrich, C. G. & Mitrenga, G. (1981).** Utilization of aliphatic amides and formation of two different amidases by *Alcaligenes eutrophus*. *J Gen Microbiol* **125**, 367-374.
- Friedrich, C. G., Bowien, B. & Friedrich, B. (1979).** Formate and oxalate metabolism in *Alcaligenes eutrophus*. *J Gen Microbiol* **115**, 185-192.
- Frund, C., Priefert, H., Steinbüchel, A. & Schlegel, H. G. (1989).** Biochemical and genetic analyses of acetoin catabolism in *Alcaligenes eutrophus*. *J Bacteriol* **171**, 6539-6548.
- Fuller, R. S., Funnell, B. E. & Kornberg, A. (1984).** The dnaA protein complex with the *E. coli* chromosomal replication origin (oriC) and other DNA sites. *Cell* **38**, 889-900.
- Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G. & Liebl, W. (2004).** Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc Natl Acad Sci U S A* **101**, 9091-9096.
- Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway de Macario, E., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W. & Birren, B. (2002).** The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res* **12**, 532-542.
- Gassel, M., Mollenkamp, T., Puppe, W. & Altendorf, K. (1999).** The KdpF subunit is part of the K(+)-translocating Kdp complex of *Escherichia coli* and is responsible for stabilization of the complex in vitro. *J Biol Chem* **274**, 37901-37907.
- Gencic, S. & Grahame, D. A. (2003).** Nickel in subunit beta of the acetyl-CoA decarbonylase/synthase multienzyme complex in methanogens. Catalytic properties and evidence for a binuclear Ni-Ni site. *J Biol Chem* **278**, 6101-6110.
- Gerdes, K., Moller-Jensen, J. & Bugge Jensen, R. (2000).** Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol* **37**, 455-466.

Gillespie, S. H., Ainscough, S., Dickens, A. & Lewin, J. (1996). Phosphorylcholine-containing antigens in bacteria from the mouth and respiratory tract. *J Med Microbiol* **44**, 35-40.

Glaser, P., Sakamoto, H., Bellalou, J., Ullmann, A. & Danchin, A. (1988). Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. *Embo J* **7**, 3997-4004.

Goosen, N., Huinen, R. G. & van de Putte, P. (1992). A 24-amino-acid polypeptide is essential for the biosynthesis of the coenzyme pyrrolo-quinoline-quinone. *J Bacteriol* **174**, 1426-1427.

Goosen, N., Horsman, H. P., Huinen, R. G. & van de Putte, P. (1989). *Acinetobacter calcoaceticus* genes involved in biosynthesis of the coenzyme pyrrolo-quinoline-quinone: nucleotide sequence and expression in *Escherichia coli* K-12. *J Bacteriol* **171**, 447-455.

Gorenflo, V., Schmack, G., Vogel, R. & Steinbüchel, A. (2001). Development of a process for the biotechnological large-scale production of 4-hydroxyvalerate-containing polyesters and characterization of their physical and mechanical properties. *Biomacromolecules* **2**, 45-57.

Gottschalk, G. (1964). [Utilization of Fructose by *Hydrogenomonas* H 16. II. Cryptic Behavior Towards Glucose.]. *Arch Mikrobiol* **49**, 96-102.

Gottschalk, G. & Thauer, R. K. (2001). The Na⁽⁺⁾-translocating methyltransferase complex from methanogenic archaea. *Biochim Biophys Acta* **1505**, 28-36.

Gottschalk, G., Eberhardt, U. & Schlegel, H. G. (1964). [Utilization of Fructose by *Hydrogenomonas* H 16. (I)]. *Arch Mikrobiol* **48**, 95-108.

Grahame, D. A. (1991). Catalysis of acetyl-CoA cleavage and tetrahydrosarcinapterin methylation by a carbon monoxide dehydrogenase-corrinoid enzyme complex. *J Biol Chem* **266**, 22227-22233.

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

Grunden, A. M. & Shanmugam, K. T. (1997). Molybdate transport and regulation in bacteria. *Arch Microbiol* **168**, 345-354.

Guenzi, E., Galli, G., Grgurina, I., Gross, D. C. & Grandi, G. (1998). Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J Biol Chem* **273**, 32857-32863.

- Guilloton, M. B., Lamblin, A. F., Kozliak, E. I., Gerami-Nejad, M., Tu, C., Silverman, D., Anderson, P. M. & Fuchs, J. A. (1993).** A physiological role for cyanate-induced carbonic anhydrase in *Escherichia coli*. *J Bacteriol* **175**, 1443-1451.
- Guo, F. B., Ou, H. Y. & Zhang, C. T. (2003).** ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res* **31**, 1780-1789.
- Hagblom, P., Segal, E., Billyard, E. & So, M. (1985).** Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* **315**, 156-158.
- Haren, L., Ton-Hoang, B. & Chandler, M. (1999).** Integrating DNA: transposases and retroviral integrases. *Annu Rev Microbiol* **53**, 245-281.
- Harkness, D. R. (1966).** Bacterial growth on aminoalkylphosphonic acids. *J Bacteriol* **92**, 623-627.
- Harms, U. & Thauer, R. K. (1996).** Methylcobalamin: coenzyme M methyltransferase isoenzymes MtaA and MtbA from *Methanosarcina barkeri*. Cloning, sequencing and differential transcription of the encoding genes, and functional overexpression of the mtaA gene in *Escherichia coli*. *Eur J Biochem* **235**, 653-659.
- Harms, U., Weiss, D. S., Gartner, P., Linder, D. & Thauer, R. K. (1995).** The energy conserving N5-methyltetrahydromethanopterin:coenzyme M methyltransferase complex from *Methanobacterium thermoautotrophicum* is composed of eight different subunits. *Eur J Biochem* **228**, 640-648.
- Hedderich, R., Koch, J., Linder, D. & Thauer, R. K. (1994).** The heterodisulfide reductase from *Methanobacterium thermoautotrophicum* contains sequence motifs characteristic of pyridine-nucleotide-dependent thioredoxin reductases. *Eur J Biochem* **225**, 253-261.
- Hendrickson, E. L., Kaul, R., Zhou, Y., Bovee, D., Chapman, P., Chung, J., Conway de Macario, E., Dodsworth, J. A., Gillett, W., Graham, D. E., Hackett, M., Haydock, A. K., Kang, A., Land, M. L., Levy, R., Lie, T. J., Major, T. A., Moore, B. C., Porat, I., Palmeiri, A., Rouse, G., Saenphimmachak, C., Soll, D., Van Dien, S., Wang, T., Whitman, W. B., Xia, Q., Zhang, Y., Larimer, F. W., Olson, M. V. & Leigh, J. A. (2004).** Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus marisaludis*. *J Bacteriol* **186**, 6956-6969.
- Hentschel, U., Steinert, M. & Hacker, J. (2000).** Common molecular mechanisms of symbiosis and pathogenesis. *Trends Microbiol* **8**, 226-231.

Hochheimer, A., Hedderich, R. & Thauer, R. K. (1998). The formylmethanofuran dehydrogenase isoenzymes in *Methanobacterium wolfei* and *Methanobacterium thermoautotrophicum*: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme. *Arch Microbiol* **170**, 389-393.

Hochheimer, A., Linder, D., Thauer, R. K. & Hedderich, R. (1996). The molybdenum formylmethanofuran dehydrogenase operon and the tungsten formylmethanofuran dehydrogenase operon from *Methanobacterium thermoautotrophicum*. Structures and transcriptional regulation. *Eur J Biochem* **242**, 156-162.

Hofer, B., Backhaus, S. & Timmis, K. N. (1994). The biphenyl/polychlorinated biphenyl-degradation locus (bph) of *Pseudomonas* sp. LB400 encodes four additional metabolic enzymes. *Gene* **144**, 9-16.

Hofmann, K. & Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol Chem Hoppe-Seyler* **374**, 166.

Hogrefe, C., Romermann, D. & Friedrich, B. (1984). *Alcaligenes eutrophus* hydrogenase genes (Hox). *J Bacteriol* **158**, 43-48.

Holden, M. T., Titball, R. W., Peacock, S. J., Cerdeno-Tarraga, A. M., Atkins, T., Crossman, L. C., Pitt, T., Churcher, C., Mungall, K., Bentley, S. D., Sebaihia, M., Thomson, N. R., Bason, N., Beacham, I. R., Brooks, K., Brown, K. A., Brown, N. F., Challis, G. L., Cherevach, I., Chillingworth, T., Cronin, A., Crossett, B., Davis, P., DeShazer, D., Feltwell, T., Fraser, A., Hance, Z., Hauser, H., Holroyd, S., Jagels, K., Keith, K. E., Maddison, M., Moule, S., Price, C., Quail, M. A., Rabinowitsch, E., Rutherford, K., Sanders, M., Simmonds, M., Songvilai, S., Stevens, K., Tumapa, S., Vesaratchavest, M., Whitehead, S., Yeats, C., Barrell, B. G., Oyston, P. C. & Parkhill, J. (2004). Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**, 14240-14245.

Hooper, L. V. & Gordon, J. I. (2001). Commensal host-bacterial relationships in the gut. *Science* **292**, 1115-1118.

Horwitz, J. P., Chua, J., Curby, R. J., Tomson, A. J., Darooge, M. A., Fisher, B. E., Mauricio, J. & Klundt, I. (1964). Substrates for Cytochemical Demonstration of Enzyme Activity. I. Some Substituted 3-Indolyl-Beta-D-Glycopyranosides. *J Med Chem* **53**, 574-575.

Hynes, M. J. (1975). Amide utilization in *Aspergillus nidulans*: evidence for a third amidase enzyme. *J Gen Microbiol* **91**, 99-109.

- Ide, T., Bäumer, S. & Deppenmeier, U. (1999).** Energy conservation by the H₂:heterodisulfide oxidoreductase from *Methanosarcina mazei* Go1: identification of two proton-translocating segments. *J Bacteriol* **181**, 4076-4080.
- Jiang, X. M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991).** Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol Microbiol* **5**, 695-713.
- Johnson, B. F. & Stanier, R. Y. (1971).** Dissimilation of aromatic compounds by *Alcaligenes eutrophus*. *J Bacteriol* **107**, 468-475.
- Johnson, M. E. & Rajagopalan, K. V. (1987).** Involvement of chlA, E, M, and N loci in *Escherichia coli* molybdopterin biosynthesis. *J Bacteriol* **169**, 117-125.
- Johnson, M. K., Rees, D. C. & Adams, M. W. (1996).** Tungstoenzymes. *Chem Rev* **96**, 2817-2840.
- Jones, P. W. & Turner, J. M. (1984).** Interrelationships between the enzymes of ethanolamine metabolism in *Escherichia coli*. *J Gen Microbiol* **130**, 299-308.
- Kappler, U., Friedrich, C. G., Truper, H. G. & Dahl, C. (2001).** Evidence for two pathways of thiosulfate oxidation in *Starkeya novella* (formerly *Thiobacillus novellus*). *Arch Microbiol* **175**, 102-111.
- Karlin, S. & Mrazek, J. (2000).** Predicted highly expressed genes of diverse prokaryotic genomes. *J Bacteriol* **182**, 5238-5250.
- Karlinsey, J. E., Lonner, J., Brown, K. L. & Hughes, K. T. (2000).** Translation/secretion coupling by type III secretion systems. *Cell* **102**, 487-497.
- Kasahara, T. & Kato, T. (2003).** Nutritional biochemistry: A new redox-cofactor vitamin for mammals. *Nature* **422**, 832.
- Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. & Kikuchi, H. (1998).** Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res* **5**, 55-76.
- Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kus-**

hida, N., Oguchi, A., Kikuchi, H. & et al. (1999). Complete genome sequence of an aerobic hyperthermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res* **6**, 83-101, 145-152.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. & Kikuchi, H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res* **8**, 123-140.

Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K. & Suzuki, M. (2000). Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc Natl Acad Sci U S A* **97**, 14257-14262.

Kelly, R. F. & Whitfield, C. (1996). Clonally diverse rfb gene clusters are involved in expression of a family of related D-galactan O antigens in *Klebsiella* species. *J Bacteriol* **178**, 5205-5214.

Keltjens, J. T. & Vogels, G. D. (1993). Conversion of methanol and methylamines to methane and carbon dioxide. In *Methanogenesis*, pp. pp. 253-303. Edited by J. G. Ferry. New York: Chapman & Hall.

Killgore, J., Smidt, C., Duich, L., Romero-Chapman, N., Tinker, D., Reiser, K., Melko, M., Hyde, D. & Rucker, R. B. (1989). Nutritional importance of pyrroloquinoline quinone. *Science* **245**, 850-852.

Kim, S. I., Leem, S. H., Choi, J. S. & Ha, K. S. (1998). Organization and transcriptional characterization of the *cat1* gene cluster in *Acinetobacter lwoffii* K24. *Biochem Biophys Res Commun* **243**, 289-294.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olsen, G. J., Fraser, C. M., Smith, H. O., Woese, C. R. & Venter, J. C. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364-370.

König, C., Sammler, I., Wilde, E. & Schlegel, H. G. (1969). [Constitutive glucose-6-phosphate dehydrogenase in mutants utilizing glucose, which are derived from cryptic wildtype strains]. *Arch Mikrobiol* **67**, 51-57.

- König, H., Kralik, R. & Kandler, O. (1982).** Structure and modifications of pseudomurein in Methanobacteriales. *Zbl Bakt Hyg* **1**, 179-191.
- Koomey, M., Gotschlich, E. C., Robbins, K., Bergstrom, S. & Swanson, J. (1987).** Effects of recA mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae*. *Genetics* **117**, 391-398.
- Krieg, N. R. & Holt, J. G. (1984).** *Bergey's Manual of Systematic Bacteriology*. Baltimore: Williams & Wilkins.
- Kusian, B. & Bowien, B. (1997).** Organization and regulation of cbb CO₂ assimilation genes in autotrophic bacteria. *FEMS Microbiol Rev* **21**, 135-155.
- Kutsukake, K., Iyoda, S., Ohnishi, K. & Iino, T. (1994).** Genetic and molecular analyses of the interaction between the flagellum-specific sigma and anti-sigma factors in *Salmonella typhimurium*. *Embo J* **13**, 4568-4576.
- Lambert, A., Osteras, M., Mandon, K., Poggi, M. C. & Le Rudulier, D. (2001).** Fructose uptake in *Sinorhizobium meliloti* is mediated by a high-affinity ATP-binding cassette transport system. *J Bacteriol* **183**, 4709-4717.
- Laoide, B. M. & Ullmann, A. (1990).** Virulence dependent and independent regulation of the *Bordetella pertussis* cya operon. *Embo J* **9**, 999-1005.
- Lawrence, J. G. & Ochman, H. (1998).** Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* **95**, 9413-9417.
- Lee, H. C. & Bernstein, H. D. (2001).** The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A* **98**, 3471-3476.
- Lehmacher, A. (1994).** Cloning, sequencing and transcript analysis of the gene encoding formylmethanofuran: tetrahydromethanopterin formyltransferase from the hyperthermophilic *Methanothermobacter fervidus*. *Mol Gen Genet* **242**, 73-80.
- Lengeler, J. W., Drews, G. & Schlegel, H. G. (1999).** *Biology of the Prokaryotes*. Stuttgart: Georg Thieme Verlag.

- Lenz, O., Bernhard, M., Buhrke, T., Schwartz, E. & Friedrich, B. (2002).** The hydrogen-sensing apparatus in *Ralstonia eutropha*. *J Mol Microbiol Biotechnol* **4**, 255-262.
- Lerouge, I. & Vanderleyden, J. (2002).** O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol Rev* **26**, 17-47.
- Levitt, M. D. & Ingelfinger, F. J. (1968).** Hydrogen and methane production in man. *Ann N Y Acad Sci* **150**, 75-81.
- Lin, W., Fullner, K. J., Clayton, R., Sexton, J. A., Rogers, M. B., Calia, K. E., Calderwood, S. B., Fraser, C. & Mekalanos, J. J. (1999).** Identification of a vibrio cholerae RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc Natl Acad Sci U S A* **96**, 1071-1076.
- Lindinger, W., Taucher, J., Jordan, A., Hansel, A. & Vogel, W. (1997).** Endogenous production of methanol after the consumption of fruit. *Alcohol Clin Exp Res* **21**, 939-943.
- Liu, D., Burton, S., Glancy, T., Li, Z. S., Hampton, R., Meade, T. & Merlo, D. J. (2003).** Insect resistance conferred by 283-kDa *Photobacterium luminescens* protein TcdA in *Arabidopsis thaliana*. *Nat Biotechnol* **21**, 1222-1228.
- Lobry, J. R. (1996).** Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* **13**, 660-665.
- Lopez, P., Philippe, H., Myllykallio, H. & Forterre, P. (1999).** Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol* **32**, 883-886.
- Lowe, T. M. & Eddy, S. R. (1997).** tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964.
- Lundgren, M., Andersson, A., Chen, L., Nilsson, P. & Bernander, R. (2004).** Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**, 7046-7051.
- Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P. O., Crosa, J. H. & O'Connell, M. (2001).** Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *J Bacteriol* **183**, 2576-2585.
- Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M. & Weiser, J. N. (2000).** The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. *Mol Microbiol* **35**, 234-245.

- Macdonald, T. T. & Monteleone, G. (2005).** Immunity, inflammation, and allergy in the gut. *Science* **307**, 1920-1925.
- Mackie, R. I., Sghir, A. & Gaskins, H. R. (1999).** Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* **69**, 1035S-1045S.
- Macpherson, A. J. & Uhr, T. (2004).** Compartmentalization of the mucosal immune responses to commensal intestinal bacteria. *Ann N Y Acad Sci* **1029**, 36-43.
- Mahenthiralingam, E., Urban, T. A. & Goldberg, J. B. (2005).** The multifarious, multireplicon Burkholderia cepacia complex. *Nat Rev Microbiol* **3**, 144-156.
- Makkar, N. S. & Casida, L. E. (1987).** *Cupriavidus necator* gen. nov., sp. nov.: a nonobligate bacterial predator of bacteria in soil. *Int J Syst Bacteriol* **37**, 323-326.
- Marchesi, J. R., Weightman, A. J., Cragg, B. A., Parkes, R. J. & Fry, J. C. (2001).** Methanogen and bacterial diversity and distribution in deep gas hydrate sediments from the Cascadia Margin as revealed by 16S rRNA molecular analysis. *FEMS Microbiol Ecol* **34**, 221-228.
- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. & Bryant, S. H. (2005).** CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33 Database Issue**, D192-196.
- Marzluf, G. A. (1997).** Genetic regulation of nitrogen metabolism in the fungi. *Microbiol Mol Biol Rev* **61**, 17-32.
- Mergeay, M., Nies, D., Schlegel, H. G., Gerits, J., Charles, P. & Van Gijsegem, F. (1985).** *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J Bacteriol* **162**, 328-334.
- Merkl, R. (2003).** A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. *J Mol Evol* **57**, 453-466.
- Merkl, R. (2004).** SIGI: score-based identification of genomic islands. *BMC Bioinformatics* **5**, 22.
- Messer, W. (2002).** The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* **26**, 355-374.

- Meuer, J., Kuettner, H. C., Zhang, J. K., Hedderich, R. & Metcalf, W. W. (2002).** Genetic analysis of the archaeon *Methanosarcina barkeri* Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc Natl Acad Sci U S A* **99**, 5632-5637.
- Meulenberg, J. J., Sellink, E., Riegman, N. H. & Postma, P. W. (1992).** Nucleotide sequence and structure of the *Klebsiella pneumoniae* pqq operon. *Mol Gen Genet* **232**, 284-294.
- Miller, T. L. & Wolin, M. J. (1982).** Enumeration of *Methanobrevibacter smithii* in human feces. *Arch Microbiol* **131**, 14-18.
- Miller, T. L. & Wolin, M. J. (1983).** Oxidation of hydrogen and reduction of methanol to methane is the sole energy source for a methanogen isolated from human feces. *J Bacteriol* **153**, 1051-1055.
- Miller, T. L. & Wolin, M. J. (1985).** *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch Microbiol* **141**, 116-122.
- Miller, W. L., Pabbaraju, K. & Sanderson, K. E. (2000).** Fragmentation of 23S rRNA in strains of *Proteus* and *Providencia* results from intervening sequences in the *rrn* (rRNA) genes. *J Bacteriol* **182**, 1109-1117.
- Mitchell, A. E., Jones, A. D., Mercer, R. S. & Rucker, R. B. (1999).** Characterization of pyrroloquinoline quinone amino acid derivatives by electrospray ionization mass spectrometry and detection in human milk. *Anal Biochem* **269**, 317-325.
- Morgenstern, B. (1999).** DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**, 211-218.
- Morris, J., Donnelly, D. F., O'Neill, E., McConnell, F. & O'Gara, F. (1994).** Nucleotide sequence analysis and potential environmental distribution of a ferric pseudobactin receptor gene of *Pseudomonas* sp. strain M114. *Mol Gen Genet* **242**, 9-16.
- Mosser, J. L. & Tomasz, A. (1970).** Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. *J Biol Chem* **245**, 287-298.
- Mukhopadhyay, B., Purwantini, E., Pihl, T. D., Reeve, J. N. & Daniels, L. (1995).** Cloning, sequencing, and transcriptional analysis of the coenzyme F420-dependent methylene-5,6,7,8-tetrahydromethanopterin dehydrogenase gene from *Methanobacterium thermoautotrophicum* strain Marburg and functional expression in *Escherichia coli*. *J Biol Chem* **270**, 2827-2832.

- Nambu, T. & Kutsukake, K. (2000).** The Salmonella FlgA protein, a putative periplasmic chaperone essential for flagellar P ring formation. *Microbiology* **146** (Pt 5), 1171-1178.
- Nandadasa, H. G., Andreesen, M. & Schlegel, H. G. (1974).** The utilization of 2-ketogluconate by Hydrogenomonas eutropha H 16. *Arch Microbiol* **99**, 15-23.
- Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T. A., Welti, R., Goo, Y. A., Leithauser, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W., Maddocks, D. G., Jablonski, P. E., Krebs, M. P., Angevine, C. M., Dale, H., Isenbarger, T. A., Peck, R. F., Pohlschroder, M., Spudich, J. L., Jung, K. W., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P., Omer, A. D., Ebhardt, H., Lowe, T. M., Liang, P., Riley, M., Hood, L. & DasSarma, S. (2000).** Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci U S A* **97**, 12176-12181.
- Nierman, W. C., DeShazer, D., Kim, H. S., Tettelin, H., Nelson, K. E., Feldblyum, T., Ulrich, R. L., Ronning, C. M., Brinkac, L. M., Daugherty, S. C., Davidsen, T. D., Deboy, R. T., Dimitrov, G., Dodson, R. J., Durkin, A. S., Gwinn, M. L., Haft, D. H., Khouri, H., Kolonay, J. F., Madupu, R., Mohammoud, Y., Nelson, W. C., Radune, D., Romero, C. M., Sarria, S., Selengut, J., Shamblin, C., Sullivan, S. A., White, O., Yu, Y., Zafar, N., Zhou, L. & Fraser, C. M. (2004).** Structural flexibility in the Burkholderia mallei genome. *Proc Natl Acad Sci U S A* **101**, 14246-14251.
- Nowack, B. (1998).** The behavior of phosphonates in wastewater treatment plants of Switzerland. *Water Res* **32**, 1271-1279.
- Nowack, B. (2003).** Environmental chemistry of phosphonates. *Water Res* **37**, 2533-2546.
- Oh, J. I. & Bowien, B. (1998).** Structural analysis of the fds operon encoding the NAD⁺-linked formate dehydrogenase of Ralstonia eutropha. *J Biol Chem* **273**, 26349-26360.
- Okamura, H., Denawa, M., Ohniwa, R. & Takeyasu, K. (2003).** P-type ATPase superfamily: evidence for critical roles for kingdom evolution. *Ann N Y Acad Sci* **986**, 219-223.
- Olsen, G. J. & Woese, C. R. (1997).** Archaeal genomics: an overview. *Cell* **89**, 991-994.
- Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., Bhattacharyya, A., Burd, H., Gardner, W., Hanke, P., Kapatral, V., Mikhailova, N., Vasieva, O., Osterman, A., Vonstein, V., Fonstein, M., Ivanova, N. & Kyrpides, N. (2003).** The ERGO genome analysis and discovery system. *Nucleic Acids Res* **31**, 164-171.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Holden, M. T., Churcher, C. M., Bentley, S. D., Mungall, K. L., Cerdeno-Tarraga, A. M., Temple, L., James, K., Harris, B., Quail, M. A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B. G. & Maskell, D. J. (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**, 32-40.

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

Peled, Y., Gilat, T., Liberman, E. & Bujanover, Y. (1985). The development of methane production in childhood and adolescence. *J Pediatr Gastroenterol Nutr* **4**, 575-579.

Pfützner, J. & Schlegel, H. G. (1973). [Denitrification in *Hydrogenomonas eutropha* strain H 16]. *Arch Mikrobiol* **90**, 199-211.

Pitt, P., de Bruijn, K. M., Beeching, M. F., Goldberg, E. & Blendis, L. M. (1980). Studies on breath methane: the effect of ethnic origins and lactulose. *Gut* **21**, 951-954.

Pitterle, D. M. & Rajagopalan, K. V. (1989). Two proteins encoded at the *chlA* locus constitute the converting factor of *Escherichia coli* *chlA1*. *J Bacteriol* **171**, 3373-3378.

Podolsky, D. K. (2002). The current future understanding of inflammatory bowel disease. *Best Pract Res Clin Gastroenterol* **16**, 933-943.

Pohlschröder, M., Dilks, K., Hand, N. J. & Wesley Rose, R. (2004). Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiol Rev* **28**, 3-24.

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

Ralph, D. & McClelland, M. (1993). Intervening sequence with conserved open reading frame in eubacterial 23S rRNA genes. *Proc Natl Acad Sci U S A* **90**, 6864-6868.

Reay, D. S. (2003). Sinking methane. *Biologist (London)* **50**, 15-19.

- Reeve, J. N., Beckler, G. S., Cram, D. S., Hamilton, P. T., Brown, J. W., Krzycki, J. A., Kolodziej, A. F., Alex, L., Orme-Johnson, W. H. & Walsh, C. T. (1989).** A hydrogenase-linked gene in Methanobacterium thermoautotrophicum strain delta H encodes a polyferredoxin. *Proc Natl Acad Sci U S A* **86**, 3031-3035.
- Reeves, P. (1993).** Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. *Trends Genet* **9**, 17-22.
- Rhoads, D. B., Waters, F. B. & Epstein, W. (1976).** Cation transport in Escherichia coli. VIII. Potassium transport mutants. *J Gen Physiol* **67**, 325-341.
- Ring, A., Weiser, J. N. & Tuomanen, E. I. (1998).** Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *J Clin Invest* **102**, 347-360.
- Risberg, A., Schweda, E. K. & Jansson, P. E. (1997).** Structural studies of the cell-envelope oligosaccharide from the lipopolysaccharide of Haemophilus influenzae strain RM.118-28. *Eur J Biochem* **243**, 701-707.
- Rivers, S. L., McNairn, E., Blasco, F., Giordano, G. & Boxer, D. H. (1993).** Molecular genetic analysis of the moa operon of Escherichia coli K-12 required for molybdenum cofactor biosynthesis. *Mol Microbiol* **8**, 1071-1081.
- Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. & Dunn, D. M. (2001).** Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. *Methods Enzymol* **330**, 134-157.
- Rocha, E. P., Danchin, A. & Viari, A. (1999).** Universal replication biases in bacteria. *Mol Microbiol* **32**, 11-16.
- Rospert, S., Linder, D., Ellermann, J. & Thauer, R. K. (1990).** Two genetically distinct methyl-coenzyme M reductases in Methanobacterium thermoautotrophicum strain Marburg and delta H. *Eur J Biochem* **194**, 871-877.
- Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S. & Church, G. M. (1993).** Characterization of the cobalamin (vitamin B12) biosynthetic genes of Salmonella typhimurium. *J Bacteriol* **175**, 3303-3316.
- Rouviere, P. E. & Wolfe, R. S. (1987).** 2',3'-Dialdehyde of ATP: a specific, irreversible inhibitor of component A3 of the methylreductase system of Methanobacterium thermoautotrophicum. *J Bacteriol* **169**, 1737-1739.

- Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. (2001).** Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. *Nat Struct Biol* **8**, 42-46.
- Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000).** The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508-513.
- Ruppert, C., Wimmers, S., Lemker, T. & Muller, V. (1998).** The A1A0 ATPase from *Methanosarcina mazei*: cloning of the 5' end of the *aha* operon encoding the membrane domain and expression of the proteolipid in a membrane-bound form in *Escherichia coli*. *J Bacteriol* **180**, 3448-3452.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000).** Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brotier, P., Camus, J. C., Cattolico, L., Chandler, M., Choisine, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J. & Boucher, C. A. (2002).** Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**, 497-502.
- Salisbury, S. A., Forrest, H. S., Cruse, W. B. & Kennard, O. (1979).** A novel coenzyme from bacterial primary alcohol dehydrogenases. *Nature* **280**, 843-844.
- Salyers, A. A., Vercellotti, J. R., West, S. E. & Wilkins, T. D. (1977).** Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl Environ Microbiol* **33**, 319-322.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular cloning: a laboratory manual*: Cold Spring Harbour Laboratory.
- Samuel, G. & Reeves, P. (2003).** Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503-2519.
- Sander, C. & Schneider, R. (1991).** Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* **9**, 56-68.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977).** DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467.

- Sauer, K., Harms, U. & Thauer, R. K. (1997).** Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*. Purification, properties and encoding genes of the corrinoid protein MT1. *Eur J Biochem* **243**, 670-677.
- Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R. & Moxon, E. R. (2000).** Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol Microbiol* **37**, 207-215.
- Schiex, T., Gouzy, J., Moisan, A. & de Oliveira, Y. (2003).** FrameD: A flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* **31**, 3738-3741.
- Schwartz, E. & Friedrich, B. (2001).** A physical map of the megaplasmid pHG1, one of three genomic replicons in *Ralstonia eutropha* H16. *FEMS Microbiol Lett* **201**, 213-219.
- Schwartz, E., Henne, A., Cramm, R., Eitinger, T., Friedrich, B. & Gottschalk, G. (2003).** Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H(2)-based lithoautotrophy and anaerobiosis. *J Mol Biol* **332**, 369-383.
- Schwarzer, D. & Marahiel, M. A. (2001).** Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* **88**, 93-101.
- Selenska-Pobell, S., Evguenieva-Hackenberg, E., Radeva, G. & Squartini, A. (1996).** Characterization of *Rhizobium 'hedysari'* by RFLP analysis of PCR amplified rDNA and by genomic PCR fingerprinting. *J Appl Bacteriol* **80**, 517-528.
- Serino, L. & Virji, M. (2000).** Phosphorylcholine decoration of lipopolysaccharide differentiates commensal *Neisseriae* from pathogenic strains: identification of licA-type genes in commensal *Neisseriae*. *Mol Microbiol* **35**, 1550-1559.
- Serino, L. & Virji, M. (2002).** Genetic and functional analysis of the phosphorylcholine moiety of commensal *Neisseria* lipopolysaccharide. *Mol Microbiol* **43**, 437-448.
- Setzke, E., Hedderich, R., Heiden, S. & Thauer, R. K. (1994).** H2: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties. *Eur J Biochem* **220**, 139-148.
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M.**

- E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W. & Van der Oost, J. (2001).** The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98**, 7835-7840.
- Shima, S., Warkentin, E., Thauer, R. K. & Ermler, U. (2002).** Structure and function of enzymes involved in the methanogenic pathway utilizing carbon dioxide and molecular hydrogen. *J. Biosci. Bioeng.* **93**, 519-530.
- Shingler, V., Powlowski, J. & Marklund, U. (1992).** Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *J Bacteriol* **174**, 711-724.
- Shuman, S. (1991).** Recombination mediated by vaccinia virus DNA topoisomerase I in *Escherichia coli* is sequence specific. *Proc Natl Acad Sci U S A* **88**, 10104-10108.
- Simpson, P. G. & Whitman, W. B. (1993).** Anabolic pathways in methanogens. In *Methanogenesis: Ecology, Physiology, Biochemistry, and Genetics.*, pp. 445-472. New York/London: Chapman & Hall.
- Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., Polushin, N. N., Shcherbinina, O. V., Shakhova, V. V., Belova, G. I., Aravind, L., Natale, D. A., Rogozin, I. B., Tatusov, R. L., Wolf, Y. I., Stetter, K. O., Malykh, A. G., Koonin, E. V. & Kozyavkin, S. A. (2002).** The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99**, 4644-4649.
- Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N. & et al. (1997).** Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179**, 7135-7155.
- Song, X. M., Forsgren, A. & Janson, H. (1999).** Fragmentation heterogeneity of 23S ribosomal RNA in *Haemophilus* species. *Gene* **230**, 287-293.
- Sorgenfrei, O., Muller, S., Pfeiffer, M., Sniezko, I. & Klein, A. (1997).** The [NiFe] hydrogenases of *Methanococcus voltae*: genes, enzymes and regulation. *Arch Microbiol* **167**, 189-195.
- Sparling, R., Blaut, M. & Gottschalk, G. (1993).** Bioenergetic studies of *Methanosphaera stadtmanae*, an obligate H₂-methanol utilizing methanogen. *Can J Microbiol* **39**, 742-748.

- Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M. & Jennings, M. P. (2005).** The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A* **102**, 5547-5551.
- Staden, R. (1996).** The Staden sequence analysis package. *Mol Biotechnol* **5**, 233-241.
- Starkey, R. L. (1935).** Isolation of some bacteria which oxidize thiosulfate. *Soil Science* **39**, 197-219.
- Stein, T., Vater, J., Kruff, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R. & Morris, H. R. (1996).** The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *J Biol Chem* **271**, 15428-15435.
- Steinbüchel, A. & Schlegel, H. G. (1991).** Physiology and molecular genetics of poly(beta-hydroxyalkanoic acid) synthesis in *Alcaligenes eutrophus*. *Mol Microbiol* **5**, 535-542.
- Steinbüchel, A., Frund, C., Jendrossek, D. & Schlegel, H. G. (1987).** Isolation of mutants of *Alcaligenes eutrophus* unable to derepress the fermentative alcohol dehydrogenase. *Arch. Microbiol.* **148**, 178-186.
- Steinert, M., Hentschel, U. & Hacker, J. (2000).** Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* **87**, 1-11.
- Stojanowic, A., Mander, G. J., Duin, E. C. & Hedderich, R. (2003).** Physiological role of the F420-non-reducing hydrogenase (Mvh) from *Methanothermobacter marburgensis*. *Arch Microbiol* **180**, 194-203.
- Stumpe, S., Schlösser, A., Schleyer, M. & Bakker, E. P. (1996).** K⁺ circulation across the prokaryotic cell membrane: K⁺ uptake systems. In *Handbook of Biological Physics.*, pp. 473-499. Edited by W. N. Konings, H. R. Kaback & J. S. Lolkema.
- Sung, Y. C. & Fuchs, J. A. (1988).** Characterization of the *cyn* operon in *Escherichia coli* K12. *J Biol Chem* **263**, 14769-14775.
- Szymanski, C. M. & Wren, B. W. (2005).** Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* **3**, 225-237.
- Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. (1999).** Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol* **32**, 1022-1030.

- Taghavi, S., Provoost, A., Mergeay, M. & van der Lelie, D. (1996).** Identification of a partition and replication region in the *Alcaligenes eutrophus* megaplasmid pMOL28. *Mol Gen Genet* **250**, 169-179.
- Tans-Kersten, J., Huang, H. & Allen, C. (2001).** *Ralstonia solanacearum* needs motility for invasive virulence on tomato. *J Bacteriol* **183**, 3597-3605.
- Tech, M. & Merkl, R. (2003).** YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol* **3**, 441-451.
- Tersteegen, A. & Hedderich, R. (1999).** Methanobacterium thermoautotrophicum encodes two multisubunit membrane-bound [NiFe] hydrogenases. Transcription of the operons and sequence analysis of the deduced proteins. *Eur J Biochem* **264**, 930-943.
- Tettelin, H., Radune, D., Kasif, S., Khouri, H. & Salzberg, S. L. (1999).** Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**, 500-507.
- van de Wijngaard, W. M., Creemers, J., Vogels, G. D. & van der Drift, C. (1991).** Methanogenic pathways in *Methanosphaera stadtmanae*. *FEMS Microbiol Lett* **64**, 207-211.
- Vandamme, P. & Coenye, T. (2004).** Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Microbiol* **54**, 2285-2289.
- Vaneechoutte, M., Kampfer, P., De Baere, T., Falsen, E. & Verschraegen, G. (2004).** *Wautersia* gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia eutropha* and related species, and proposal of *Ralstonia* [*Pseudomonas*] *syzygii* (Roberts et al. 1990) comb. nov. *Int J Syst Evol Microbiol* **54**, 317-327.
- Vaupel, M. & Thauer, R. K. (1995).** Coenzyme F420-dependent N₅,N₁₀-methylene-tetrahydro-methanopterin reductase (Mer) from *Methanobacterium thermoautotrophicum* strain Marburg. Cloning, sequencing, transcriptional analysis, and functional expression in *Escherichia coli* of the mer gene. *Eur J Biochem* **231**, 773-778.
- Vicente, M. & Canovas, J. L. (1973).** Glucolysis in *Pseudomonas putida*: physiological role of alternative routes from the analysis of defective mutants. *J Bacteriol* **116**, 908-914.
- von Heijne, G. (1990).** The signal peptide. *J Membr Biol* **115**, 195-201.
- Vorholt, J. A. & Thauer, R. K. (1997).** The active species of 'CO₂' utilized by formylmethanofuran dehydrogenase from methanogenic Archaea. *Eur J Biochem* **248**, 919-924.

- Walderhaug, M. O., Polarek, J. W., Voelkner, P., Daniel, J. M., Hesse, J. E., Altendorf, K. & Epstein, W. (1992).** KdpD and KdpE, proteins that control expression of the kdpABC operon, are members of the two-component sensor-effector class of regulators. *J Bacteriol* **174**, 2152-2159.
- Wandersman, C. & Delepelaire, P. (1990).** TolC, an Escherichia coli outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci U S A* **87**, 4776-4780.
- Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D. & French-Constant, R. H. (2001).** The tc genes of Photorhabdus: a growing family. *Trends Microbiol* **9**, 185-191.
- Weiser, J. N., Williams, A. & Moxon, E. R. (1990).** Phase-variable lipopolysaccharide structures enhance the invasive capacity of Haemophilus influenzae. *Infect Immun* **58**, 3455-3457.
- Weiser, J. N., Shchepetov, M. & Chong, S. T. (1997).** Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of Haemophilus influenzae. *Infect Immun* **65**, 943-950.
- Weiser, J. N., Goldberg, J. B., Pan, N., Wilson, L. & Virji, M. (1998a).** The phosphorylcholine epitope undergoes phase variation on a 43-kilodalton protein in Pseudomonas aeruginosa and on pili of Neisseria meningitidis and Neisseria gonorrhoeae. *Infect Immun* **66**, 4263-4267.
- Weiser, J. N., Pan, N., McGowan, K. L., Musher, D., Martin, A. & Richards, J. (1998b).** Phosphorylcholine on the lipopolysaccharide of Haemophilus influenzae contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *J Exp Med* **187**, 631-640.
- Welch, R. A. (2001).** RTX toxin structure and function: a story of numerous anomalies and few analogies in toxin biology. *Curr Top Microbiol Immunol* **257**, 85-111.
- Welch, R. A., Forestier, C., Lobo, A., Pellett, S., Thomas, W., Jr. & Rowe, G. (1992).** The synthesis and function of the Escherichia coli hemolysin and related RTX exotoxins. *FEMS Microbiol Immunol* **5**, 29-36.
- Welch, R. A., Bauer, M. E., Kent, A. D., Leeds, J. A., Moayeri, M., Regassa, L. B. & Swenson, D. L. (1995).** Battling against host phagocytes: the wherefore of the RTX family of toxins? *Infect Agents Dis* **4**, 254-272.
- White, A. K. & Metcalf, W. W. (2004).** Two C-P lyase operons in Pseudomonas stutzeri and their roles in the oxidation of phosphonates, phosphite, and hypophosphite. *J Bacteriol* **186**, 4730-4739.
- Whitfield, C. (1995).** Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* **3**, 178-185.

- Wiezer, A. & Merkl, R. (2003).** secureBLAST. *In Silico Biol* **3**, 405-409.
- Wilde, E. (1962).** Untersuchungen über Wachstum und Speicherstoffsynthese von *Hydrogenomonas*. *Arch Microbiol* **43**, 109-137.
- Windhövel, U. & Bowien, B. (1991).** Identification of cfxR, an activator gene of autotrophic CO₂ fixation in *Alcaligenes eutrophus*. *Mol Microbiol* **5**, 2695-2705.
- Wolin, M. J. & Miller, T. L. (1994).** Acetogenesis from CO₂ in the human colonic ecosystem. In *Acetogenesis*, pp. 365-385. Edited by H. I. Drake. New York: Chapman and Hall.
- Wong, D., Lin, Z., Juck, D., Terrick, K. A. & Sparling, R. (1994).** Electron transfer reactions for the reduction of NADP⁺ in *Methanospiraeta stadtmanae*. *FEMS Microbiol Lett* **120**, 285-290.
- Wood, W. T., Gettrust, J. F., Chapman, N. R., Spence, G. D. & Hyndman, R. D. (2002).** Decreased stability of methane hydrates in marine sediments owing to phase-boundary roughness. *Nature* **420**, 656-660.
- Wuebbens, M. M. & Rajagopalan, K. V. (1995).** Investigation of the early steps of molybdopterin biosynthesis in *Escherichia coli* through the use of in vivo labeling studies. *J Biol Chem* **270**, 1082-1087.
- Wyborn, N. R., Scherr, D. J. & Jones, C. W. (1994).** Purification, properties and heterologous expression of formamidase from *Methylophilus methylotrophus*. *Microbiology* **140**, 191-195.
- Yabuuchi, E., Kosako, Y., Yano, I., Hotta, H. & Nishiuchi, Y. (1995).** Transfer of two Burkholderia and an Alcaligenes species to Ralstonia gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov. *Microbiol Immunol* **39**, 897-904.

Danksagung

Ganz besonders danke ich Herrn Prof. Dr. G. Gottschalk für seine uneingeschränkte Unterstützung, die ich in jeder erdenklichen Hinsicht genossen habe. Ich danke ihm dafür, daß ich zwei interessante Projekte in einer Doktorarbeit behandeln konnte und daß er mir die Freiheit zu selbstständiger Arbeit ließ.

Ebenso danke ich Herrn Prof. Dr. B. Bowien für die freundliche Betreuung meiner Arbeit: für die offene Tür im Erdgeschoß, hinter der ich stets seine unerschöpfliche Kompetenz und ansteckende Begeisterung für *Ralstonia eutropha* und dazu eine gute Portion Humor vorgefunden habe.

Innerhalb meiner Arbeitsgruppe danke ich allen Kollegen vom Göttingen Genomics Laboratory für ihre gute Zusammenarbeit. Als weiteren Partnern danke ich aus der *Ralstonia*-Kooperation Frau Prof. Dr. B. Friedrich, Herrn Prof. Dr. A. Steinbüchel, Dr. Bernhard Kusian, Dr. Anne Pohlmann und Frank Reinecke. Innerhalb des *Methanospaera*-Projektes danke ich Prof. Dr. R. Thauer und insbesondere Henning Seedorf für ihre Unterstützung.

Meinen Kollegen vom Göttingen Genomics Laboratory und aus dem Institut für Mikrobiologie und Genetik danke ich außerdem für die gute und konstruktive Atmosphäre, die mich während meiner gesamten Arbeit umgeben hat. Ich danke Heiko Liesegang für sein offenes Ohr, Axel Strittmatter für seine Kritik, Birgit Veit für ihre Freundschaft, Christian Hildmann für den "Hildmann", Jarek Sobkowiak für "vielen", Anke Henne für ihre Energie, Frank Hoster für seine Ruhe, Arnim Wiezer für Linux, Rainer Merkl für SIGI, Silke Dencker für ihre Kraft, Iwona Decker für ihre Übersicht, Monika Gryzwa für ihre Amplifikationen, Marita Kalck für ihren Spott, Olaf Waase für Hilfe, Silke Steckel für ihre Sequenzen, Elzbieta Brzuszkiewicz für ihre Freundlichkeit, Frauke Meyer für Musik und Sönke Andres für Gelassenheit.

Abseits von meiner Arbeit haben mir viele gute Freunde in schwierigen Zeiten geholfen – man könnte sagen: mich ertragen! Dafür danke ich Maik Heistermann und Michael Jackenkroll und Lars Maschine und Thomas Stodulka und Wulf Winkelmüller und Holger Kramer und Dirk Katterbach und Janna Lau und Philipp Künstler und Milena Nin-kovic. Und Philip Scheibe und Ingmar Jürgens sowieso.

Ich danke meinen Eltern und Onkel Bob.

Lebenslauf

Wolfgang Florian Fricke	Kreuzbergring 2 37075 Göttingen
Geburtsdatum, -ort	28.10.1974, Lemgo
08/1981 - 07/1984	Grundschule Ost, Lemgo
08/1984 - 06/1994	Engelbert-Kämpfer-Gymnasium, Lemgo
10/1995 - 10/2000	Studium der Biologie Georg-August-Universität Göttingen
10/99	Diplomprüfung Hauptfach: Molekulare Genetik Nebenfächer: Biochemie, Botanik
11/99 - 10/2000	Diplomarbeit bei Prof. Dr. H.-W. Heldt, Albrecht-Haller-Institut für Pflanzenwissenschaften, Abteilung Biochemie der Pflanze Thema: „Untersuchungen zur Funktion von <i>ENOD40</i> - Genen aus Leguminosen und Aktinorhizapflanzen“
seit 02/2003	Promotion bei Prof. Dr. G. Gottschalk Institut für Mikrobiologie und Genetik, Göttingen Genomics Laboratory Thema: „Entschlüsselung der Genome von <i>Ralstonia eu-</i> <i>troph</i> a H16 und <i>Methanosphaera stadtmanae</i> und verglei- chende Untersuchungen zu Anpassungen der Genomor- ganisation“