



Universität Bielefeld

Fakultät für Biologie  
Lehrstuhl für Genetik

**Abschlussbericht für das Landesamt für Natur, Umwelt und  
Verbraucherschutz Nordrhein-Westfalen (LANUV NRW)  
zu dem Forschungsvorhaben**

## **Sequenzierung eines Kläranlagen-spezifischen Mobiloms unter Einsatz der 454-Technologie**

**Bielefeld, Mai 2007**

**Prof. Dr. A. Pühler**  
Fakultät für Biologie  
Lehrstuhl für Genetik



## **Abschlußbericht für das Forschungsvorhaben:**

### **Sequenzierung eines Kläranlagen-spezifischen Mobiloms unter Einsatz der 454-Technologie**

Projektlaufzeit: 01.08 – 31.12.2006

Berichtszeitraum: 01.08– 31.12.2006

Prof. Dr. Alfred Pühler

Lehrstuhl für Genetik

Fakultät für Biologie

Universität Bielefeld

Universitätsstrasse 25

33615 Bielefeld

Dr. Andreas Schlüter

Lehrstuhl für Genetik

Fakultät für Biologie

Universität Bielefeld

Universitätsstrasse 25

33615 Bielefeld

Dr. Rafael Szczepanowski

Lehrstuhl für Genetik

Fakultät für Biologie

Universität Bielefeld

Universitätsstrasse 25

33615 Bielefeld

## Inhaltsverzeichnis

<b>1. Einleitung</b> .....	<b>1</b>
1.1 Plasmide und ihre Bedeutung für Bakterien .....	1
1.2 In welchem Zusammenhang ist die Sequenzierung des Mobiloms zu sehen? .....	2
1.3 Sequenzierung des Kläranlagen-spezifischen Mobiloms .....	2
1.4 Die 454 Sequenzier-Technologie .....	3
1.5 Assemblierung, Annotation und Analyse.....	4
1.6 Verwertung der Sequenzdaten.....	4
<b>2. Aufgabenstellung</b> .....	<b>5</b>
<b>3. Präparation von Gesamt-Plasmid-DNA aus Belebtschlamm Bakterien einer kommunalen Abwasserkläranlage für die Sequenzanalyse</b> .....	<b>5</b>
<b>4. Ergebnisse der Hochdurchsatz Sequenzierung mittels 454-Technologie</b> .....	<b>7</b>
4.1 Kategorisierung der 346.427 Reads entsprechend der Gene Ontology .....	7
Terminologie .....	7
4.2 Abgleich der einzelnen Reads gegen eine Plasmid- und eine Resistenzdatenbank.....	8
4.3 Assemblierung der einzelnen Sequenzen und die Auswertung in SAMS.....	9
4.4 Translation der DNA-Sequenzen der 346.427 Reads und Abgleich gegen die Pfam-Datenbank.....	10
<b>5. Ausblick</b> .....	<b>12</b>
<b>6. Literatur</b> .....	<b>14</b>
<b>7. Anhang</b> .....	<b>17</b>
7.1 Tabellen.....	17
7.2 Abkürzungsverzeichnis und Glossar .....	29

## 1. Einleitung

### 1.1 Plasmide und ihre Bedeutung für Bakterien

Plasmide tragen maßgeblich zum sogenannten bakteriellen Genpool bei, dem Teil der genetischen Information, der zwischen verschiedenen bakteriellen Spezies horizontal transferiert werden kann. Plasmide kommen in einer Vielzahl unterschiedlicher Replikontypen und Größen und mit unterschiedlichem Informationsgehalt vor. Viele Plasmide sind selbst-transmissibel, das heißt sie kodieren Gene, deren Genprodukte den Plasmid-Transfer von einer Donorzelle in eine Rezipientenzelle bewerkstelligen; andere Plasmide sind durch selbst-transmissible Plasmide mobilisierbar. Darüber hinaus sind Plasmide in der Lage, neue genetische Informationen aufzunehmen und bestehende Informationen neu zu kombinieren und dadurch zur Evolution und Adaptation der Wirtsbakterien an veränderte Umweltbedingungen beizutragen (Davison, 1999). Zur genetischen Fracht eines Plasmids zählen Gene, die keine Plasmid-spezifischen Funktionen kodieren, wie z.B. Schwermetall- und Antibiotikaresistenz, Pathogenität oder Abbau von Schadstoffen (Dröge *et al.*, 2000; Heuer *et al.*, 2002; Smalla & Sobecky, 2002). Neben der Fracht eines Plasmids spielt auch das Rückgrat der Plasmide, das für deren Mobilität und stabile Erhaltung kodiert, eine entscheidende Rolle für die Anpassung der Bakterien. Als extrachromosomale DNA-Moleküle, die autonom replizieren, lassen sich Plasmide gut isolieren und von der chromosomalen DNA der Mikroorganismen separieren. Deshalb kann die Gesamtheit der bakteriellen Plasmide einer Population gesondert untersucht werden.

Plasmide spielen in Kläranlagenbakterien eine besondere Rolle bei der Anpassung der Wirtsorganismen an die besonderen Bedingungen in Abwässern und Klärschlamm. Der Eintrag von toxischen Verbindungen, Desinfektionsmitteln und Antibiotika mit dem Abwasser der Krankenhäuser, der privaten Haushalte, der Landwirtschaft und der Industrie (Feuerpfeil *et al.*, 1999; Guardabassi *et al.*, 1999; Hartmann *et al.*, 1998; Hartmann *et al.*, 1999; Kümmerer, 2001; Kümmerer, 2003; Kümmerer & Henninger, 2003; Kümmerer *et al.*, 2000; Ohlsen *et al.*, 2003) unterstützt die Selektion von resistenten Bakterien und fördert die Verbreitung der entsprechenden Determinanten, die meist Plasmid-kodiert sind (Seveno *et al.*, 2002). Aufgrund der hohen mikrobiellen Zelldichte und Diversität, des komplexen Nährstoffangebots und der Annahme, dass etwa die Hälfte der in Kläranlagen lebenden Bakterien ein oder mehrere Plasmide pro Zelle enthält, ist von einer großen Diversität des Plasmid-Pools auszugehen. Ansätze, die auf die Untersuchung der Gesamtheit aller Plasmide einer Population zielen, lassen folglich viele neue Erkenntnisse erwarten. Im Hinblick auf

Resistenzdeterminanten ist es sehr wahrscheinlich, dass Plasmid-kodierte Resistenzgene aufgefunden werden, die mit bisherigen Methoden nicht detektierbar waren. Außerdem sollte eine derartige Analyse einen Überblick über alle Plasmid-kodierten Resistenzdeterminanten in Kläranlagenbakterien geben.

### *1.2 In welchem Zusammenhang ist die Sequenzierung des Mobiloms zu sehen?*

In bisherigen Arbeiten des Lehrstuhls für Genetik der Universität Bielefeld wurden Plasmide, die aus Belebtschlamm Bakterien kommunaler Kläranlagen stammen, mit Methoden der Genomforschung auf ihren Gehalt an Antibiotikaresistenzgenen hin untersucht. In diesem Zusammenhang wurde in Zusammenarbeit mit dem Landesumweltamt Nordrhein-Westfalen ein Projekt betrieben, das die Entwicklung und Evaluierung eines Resistenzgen-Microarrays für die Detektion von Resistenzdeterminanten in Bakterien aus Kläranlagen zum Ziel hatte. Ergänzend hierzu wurde geplant, das Kläranlagen-spezifische Mobilom durch Sequenzierung näher zu charakterisieren. Das Mobilom definiert sich als die Gesamtheit aller mobilen, autonom-replizierenden genetischen Einheiten aus einer bakteriellen Population. Nukleinsäuresequenzdaten des Mobiloms einer Kläranlage sollen mit Hilfe einer neuartigen Sequenziermethode gewonnen werden, die in der Lage ist, 20 Millionen Basen in einem vier-Stunden-Durchlauf zu generieren. Der Ansatz soll einen Überblick über alle Plasmid-kodierten Resistenzdeterminanten liefern, die in Kläranlagenbakterien vorkommen, um festzustellen, wie vollständig die Erfassung der Resistenzdeterminanten in der Kläranlage mit den für den Chip ausgewählten Sonden ist.

### *1.3 Sequenzierung des Kläranlagen-spezifischen Mobiloms*

Ziel des Projektes war es, einen Überblick über die genetische Information zu bekommen, die auf dem Plasmid-Mobilom von Kläranlagenbakterien kodiert ist. Bisher wurden aus Kläranlagenbakterien isolierte Plasmide auf individueller Basis untersucht (Dröge *et al.*, 2000; Heuer *et al.*, 2004; Schlüter *et al.*, 2003; Schlüter *et al.*, 2005; Szczepanowski *et al.*, 2004; Szczepanowski *et al.*, 2005; Tauch *et al.*, 2003; Tennstedt *et al.*, 2003; Tennstedt *et al.*, 2005). Diese Art der Analyse beschränkt sich auf Plasmide, die in kultivierbaren Bakterien vorliegen und solche, die in bekannten, kultivierbaren Bakterien replikationsfähig sind. Es ist jedoch davon auszugehen, dass die Mehrzahl (ca. 85-99 %) der Kläranlagen-Bakterien nicht kultivierbar ist (Amann *et al.*, 1995). Damit sind auch Plasmide, die in nicht-kultivierbaren Bakterien vorkommen, nur schwer zugänglich. Folglich ist auch nur ein Bruchteil der Plasmid-kodierten Resistenzfracht aus Kläranlagenbakterien bekannt. An diesem Punkt setzt das hier beschriebene Projekt an. Als Ausgangsmaterial sollte aus der Bakterienpopulation

einer Kläranlage die extrachromosomale (Plasmid) DNA-Fraktion gewonnen werden. Methoden zur Isolierung von Gesamt-Plasmid-DNA wurden am Lehrstuhl für Genetik der Fakultät für Biologie der Universität Bielefeld bereits erfolgreich eingesetzt (Dröge *et al.*, 2000; Tennstedt *et al.*, 2003). Eine etwaige Verunreinigung der Plasmid-DNA durch chromosomale DNA wird mittels Cäsiumchlorid-Dichtegradienten-Zentrifugation oder durch eine Exonuklease-Behandlung der Plasmid-Präparation abgetrennt. Auf diese Weise können Plasmide in einem Größenbereich von 5 bis 200 kb isoliert werden. Präparierte Plasmid-DNA kann sofort, ohne vorherige Klonierung, einer Sequenzierung mit Hilfe der 454-Technologie zugeführt werden (Margulies *et al.*, 2005).

#### 1.4 Die 454 Sequenzier-Technologie

Die 454-Sequenzierung stellt eine sehr schnelle Methode der DNA-Sequenzierung dar, die ohne die zeit- und kostenintensive Klonierung von DNA-Fragmenten auskommt (im Gegensatz zur Sanger-Sequenzierung). Als erstes wird die DNA isoliert und in kleine Fragmente (ca. 300 – 500 bp) geschert. Anschließend werden die Enden der DNA-Fragmente zu glatten Enden aufpolymerisiert und kleine Adapter angehängt. Diese Adapter dienen einerseits als Primer-Sequenzen für die PCR-Amplifikation und Sequenzierung, andererseits trägt einer der Adapter eine 5'-Biotinmarkierung, die eine Immobilisierung des DNA-Fragments an Streptavidin-beschichtete Kügelchen ermöglicht. Mittels limitierender Verdünnung werden die Kügelchen mit einzelsträngigen DNA-Fragmenten vermischt (Überschuss an Kügelchen im Vergleich zu DNA-Fragmenten). Die Einzelstrang-DNA-Kügelchen werden anschließend in Mikroreaktoren (ein DNA-Kügelchen pro Reaktor) einer PCR unterzogen mit dem Ergebnis der klonalen Vermehrung des an das Kügelchen gebundenen DNA-Fragments.

Nach anschließender Denaturierung der DNA-Doppelstränge werden die Kügelchen, die Einzelstrang-DNA Moleküle tragen, jeweils einzeln in den Kammern eines Glasfaser-Objektträgers positioniert und mit Enzym-Kügelchen (kleinere Kügelchen, an welche Enzyme gebunden sind, die für die Erzeugung eines Chemilumineszenzsignals notwendig sind) überschichtet. Die Sequenzierung beruht auf der Synthese des komplementären Stranges. Ein Zyklus setzt sich aus folgenden Schritten zusammen:

- Zugabe eines bestimmten Nukleotid-Typs (z.B. C)
- Ist das Nukleotid komplementär, erfolgt Einbau des Nukleotids unter Freisetzung von Pyrophosphat.
- Verwendung des Pyrophosphats durch Enzyme der Enzym-Kügelchen zur Erzeugung

eines Chemilumineszenzsignals

- Detektion des Signals (die Signalstärke ist abhängig von der Zahl eingebauter Nukleotide)
- Entfernung überschüssiger Nukleotide
- Neuer Reaktionszyklus durch Zugabe eines anderen Nukleotid-Typs (z.B. A)

Die Objektträger beherbergen entweder 0,68 Mio. oder 1,2 Mio. Kammern, so dass bei einer 30 - 40 %igen Beladung mit DNA-Kügelchen und einer durchschnittlichen Leseweite von ca. 100 Basen zwischen 10 und 20 Megabasen an auswertbarer Sequenz in einem Lauf entstehen. Theoretisch kann nach dieser Methode ein Genom mit einer Größe von 2 Megabasen innerhalb von 4 Tagen mit einer 10-fachen Abdeckung sequenziert werden (mittels Sanger-Sequenzierung – mehrere Wochen).

### *1.5 Assemblierung, Annotation und Analyse*

Am Lehrstuhl für Genetik wurde eine Software Pipeline, die auf Standard-Programmen beruht (PHRED/PHPAP, CAP3 und Consed/Autofinish (Gordon *et al.*, 1998; Gordon *et al.*, 2001)), für die Durchführung, Kontrolle, Assemblierung und den Abschluss von Nucleinsäure-Sequenzierprojekten entwickelt (Kaiser *et al.*, 2003). Rohe Sequenzdaten werden mit Hilfe der Programme PHRED (Bestimmung der Basenabfolge) und BioMake (Vorhölter *et al.*, 2003) (Qualitätskontrolle) verarbeitet, um sie dann mittels der Programme CAP3 und PHRAP zu assemblieren. BioMake erlaubt eine erste Annotation der Sequenzen in einem sehr frühen Projekt-Stadium. Hierfür werden Contigs in das Annotationsprogramm GenDB (Vorhölter *et al.*, 2003) geladen. GenDB wird sowohl für manuelle als auch automatische Annotationsstrategien benutzt. Gene werden entsprechend den Standards, die in den Datenbanken GeneOntology (GO) und COG (Clusters of Orthologous Groups of Protein) definiert wurden, annotiert und funktionell kategorisiert.

### *1.6 Verwertung der Sequenzdaten*

Anfallende Sequenzdaten sollten in erster Instanz auf das Vorkommen von Resistenzgenen untersucht werden. In dieser Hinsicht dienen die Ergebnisse auch der Evaluierung von Daten, die in dem parallelen Projekt über Hybridisierung von DNA-Proben mit Resistenz-Microarrays gewonnen wurden. Andererseits sollen aber auch durch das Sequenzierprojekt neu entdeckte Resistenzdeterminanten für die Erweiterung der Resistenz-Microarrays berücksichtigt werden.

## 2. Aufgabenstellung

Ziel des Projektes war die Charakterisierung der genetischen Informationen, die auf dem Resistenzplasmid-Mobilom (Gesamtheit der Resistenzplasmide) von Kläranlagenbakterien lokalisiert sind. Die erhaltenen Informationen sollten auch mit Ergebnissen eines vorangehenden Projekts verglichen werden, das als Auftragsforschung in Zusammenarbeit mit dem Landesumweltamt Nordrhein-Westfalen betrieben wurde und die Entwicklung und Evaluierung eines Resistenz-Microarrays für die Detektion von Resistenzdeterminanten in Bakterien aus Kläranlagen zum Ziel hatte. Des Weiteren können die aus der Sequenzierung gewonnenen Daten zur Identifizierung weiterer Determinanten (Resistenz und Virulenz) verwendet werden. Diese wiederum könnten zur Ausweitung des Nachweisspektrums des bereits entwickelten Resistenz-Microarrays herangezogen werden.

## 3. Präparation von Gesamt-Plasmid-DNA aus Belebtschlamm Bakterien einer kommunalen Abwasserkläranlage für die Sequenzanalyse

Zur Isolierung der Gesamt-Plasmid-DNA, die für die Sequenzanalyse des Resistenzplasmid-Mobiloms benötigt wird, wurden die Bakterien des Belebtschlamm auf Luria Broth Medien mit unterschiedlichen Antibiotika (siehe Tabelle 1) ausplattiert und über 36 Stunden bei 30 °C inkubiert. Damit beschränkt sich der Ansatz auf die kultivierbare Fraktion der Belebtschlamm Bakterien. Anschließend wurden die gewachsenen Bakterien abgeschwemmt und einer Plasmidisolierung (Macherey-Nagel Kit, Düren, Deutschland) unterzogen. Auf diese Weise können Plasmide isoliert werden, die potentiell Resistenz gegen Antibiotika vermitteln. Diese Plasmide repräsentieren einen Teil des Kläranlagen-spezifischen Plasmid-lokalisierten Resistoms. Resistom definiert sich als die Gesamtheit der Resistenzgene in einem Habitat (also sowohl Plasmid- als auch chromosomal kodierte Gene).

Tabelle 1: Die eingesetzten Konzentrationen der zur Selektion von resistenten Bakterien verwendeten Antibiotika.

Antibiotikum	Konzentration [ $\mu\text{g} / \text{ml}$ ]
Ampicillin	100
Cefotaxim	1
Cefuroxim	15
Ciprofloxacin	1
Erythromycin	200
Gentamicin	20
Kanamycin	50
Norfloxacin	1
Rifampicin	30
Spectinomycin	100
Streptomycin	100
Tetracyclin	5

Das Luria-Broth Medium wurde zusätzlich mit 75  $\mu\text{g} / \text{ml}$  Cycloheximid versetzt, um das Pilzwachstum zu minimieren.



Die Kombination der alkalischen Lyse mit anschließender Anionentauscher-Chromatographie (Macherey-Nagel, Düren, Deutschland) hat sich als die effizienteste Methode zur Isolierung von Gesamt-Plasmid-DNA herausgestellt. Die so gewonnene Plasmid-DNA wurde weiter über Cäsiumchlorid-Dichtegradienten-Zentrifugation aufgereinigt. Ein Beispiel für das Ergebnis einer alkalischen Lyse (nach Macherey-Nagel) und einer nachgeschalteten Aufreinigung mittels Cäsiumchlorid-Dichtegradienten-Zentrifugation ist in Abbildung 1 dargestellt.

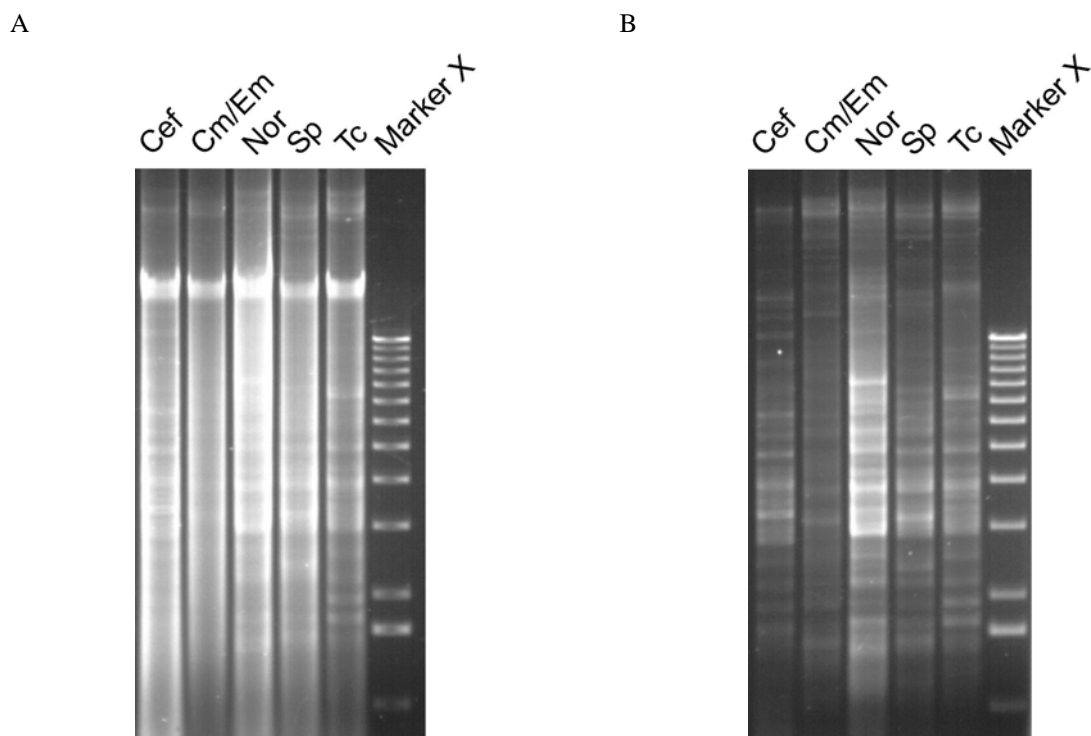


Abbildung 1: Die Überprüfung der Qualität von Gesamt-Plasmid-DNAs nach Isolierung mittels alkalischer Lyse (A) und nach einem weiteren Aufreinigungsschritt mittels Cäsiumchlorid-Dichtegradientenzentrifugation (B) in einem Agarosegel.

Abkürzungen: Cef – Cefuroxim; Cm – Chloramphenicol; Em – Erythromycin; Nor – Norfloxacin; Sp – Spectinomycin; Tc – Tetracyclin.

Es ist deutlich zu erkennen, dass der chromosomale Anteil der DNA (der weiße Schmier in Abbildung 1A) nach einer Dichtegradientenzentrifugation von der Plasmid-DNA (einzelne Banden in Abbildung 1B) getrennt wurde. Die Konzentrationen der isolierten Gesamt-Plasmid-DNAs sowie die eingesetzten Volumina für die Hochdurchsatz Sequenzierung mit der 454-Technologie können der Tabelle 2 entnommen werden.

Für die Sequenzierung mit der 454-Technologie wurden die Gesamt-Plasmid-DNAs in Abhängigkeit ihrer Konzentrationen vermischt (siehe Tabelle 2) und die gemischte Gesamt-Plasmid-DNA durch eine Isopropanol-Fällung auf ein Volumen von 25 µl mit einer DNA-

Menge von ca. 11 µg aufkonzentriert.

Tabelle 2: Konzentration der isolierten Gesamt-Plasmid-DNAs sowie die für die Sequenzierung eingesetzten DNA-Mengen.

Primärselektion der Bakterien auf	DNA-Konzentration [ng / µl]	Proben-Vol. [µl]	DNA-Menge [ng]
Ampicillin	47,15	27,95	1320
Cefotaxim	72,07	18,35	1320
Cefuroxim	78,86	16,72	1320
Ciprofloxacin	128,61	10,23	1320
Erythromycin	41,83	32,59	1320
Erythromycin + Rifampicin *	41,21	32,02	1320
Gentamicin	16,37	80,60	1320
Kanamycin	119,40	11,05	1320
Norfloxacin	16,51	79,98	1320
Spectinomycin	101,49	13,02	1320
Streptomycin	116,78	11,28	1320
Tetracyclin	22,64	58,30	1320

\* Die auf Erythromycin bzw. Rifampicin gewachsenen Kulturen wurden aufgrund der geringen Zahl wachsender Bakterienkolonien für die Isolierung der Gesamt-Plasmid-DNA vereinigt.

#### 4. Ergebnisse der Hochdurchsatz Sequenzierung mittels 454-Technologie

Die Sequenzierung des Belebtschlamm-bakterien-spezifischen Mobiloms, die durch 454 Life Sciences (Branford, CT, USA) durchgeführt wurde, ergab mit einem Sequenzierlauf 346.427 Reads mit einer durchschnittlichen Leseweite von 104 Basen. Damit ergibt sich eine Gesamtmenge von 36.071.493 Basen an auswertbarer DNA-Sequenz. Die Sequenzdaten wurden auf das eigene System übertragen und in einer Datenbank gespeichert (SAMS-Projekt: SAMS\_CEBITEC\_Mobilom).

##### 4.1 Kategorisierung der 346.427 Reads entsprechend der Gene Ontology Terminologie

Die 346.427 Reads wurden entsprechend der GO-Terminologie (GO – Gene Ontology) in die Bereiche zelluläre Komponente (*cellular component*), biologischer Prozess (*biological process*) und molekulare Funktion (*molecular function*) klassifiziert. Der Bereich 'zelluläre Komponente' beschreibt eine Komponente, die entweder zur anatomischen Struktur einer Zelle (z.B. Chromosom) oder zu einer Genprodukt-Gruppe gehört (z.B. Ribosom oder Proteindimer). Durch 'biologischer Prozess' werden Vorgänge beschrieben, die durch einen oder mehrere geordnete Bausteine molekularer Funktionen gebildet werden (z.B. Signaltransduktion, Pyrimidinmetabolismus, Zuckertransport). Die Kategorie 'Molekulare Funktion' beschreibt schließlich Aktivitäten (Katalyse oder Bindung), die auf molekularer Ebene stattfinden.

Im Bereich 'zelluläre Komponente' sind besonders die Komponenten hervorzuheben, die als

'*extrachromosomal circular*' DNA bezeichnet werden (siehe Tabelle A1). Dieser Term bezeichnet u.a. Plasmide. Die hohe Zahl an Reads (11.468), denen dieser Term zugeordnet wurde, zeigt, dass viele der Reads plasmidären Ursprungs sind. Dem gegenüber steht die Komponente '*chromosome*' mit 139 Reads. Dies zeigt, dass in der sequenzierten DNA nur wenig chromosomale Verunreinigung vorhanden war. Des Weiteren konnten insgesamt 2360 Reads als Komponenten identifiziert werden, die membrangebunden oder in die Membran integriert sind oder zur äußerer Membran gehören. In dieser Komponentenklasse befinden sich u.a. Transporter, die auch für Antibiotikaresistenz eine Rolle spielen können.

Im Bereich 'biologischer Prozess' konnten viele Reads den Begriffen DNA-Replikation (1809 Reads) und DNA-Replikationsinitiation (4889 Reads), sowie Plasmid-Erhaltung (4411 Reads) zugeordnet werden (Tabelle A1). Hierbei handelt es sich um Prozesse, die für den Erhalt und die stabile Vererbung von Plasmiden eine wichtige Rolle spielen. DNA-Transposition (8931 Reads), DNA-Rekombination (4321 Reads) und DNA-Integration (1670 Reads) stellen Prozesse dar, die überwiegend für die Mobilität von genetischen Elementen wichtig sind. Insgesamt konnten ca. 17.000 Reads Prozessen zugeordnet werden, die direkt mit DNA-Prozessierung zu tun haben.

Im Bereich 'molekulare Funktion' decken sich die Ergebnisse der Kategorisierung mit denen für 'biologischer Prozess' erhaltenen Informationen. So konnten z.B. 19.196 Reads der DNA-Bindung, 2310 der Rekombination und 1067 der Sequenz-spezifischen DNA-Bindung zugeordnet werden (insgesamt 22.573 Reads gegenüber 17.000 Reads aus dem 'biologischer Prozess' Bereich). Des Weiteren entspricht die Zahl der Reads, die als Transposaseaktivität zusammengefasst wurden (8754 Reads), fast genau der Zahl der Reads für DNA-Transposition (8931 Reads) des Bereichs 'biologischer Prozess'. Zusammenfassend lässt sich aus diesen Befunden bereits erkennen, dass Plasmide aus Klärschlamm Bakterien hauptsächlich aus Modulen aufgebaut sind, die für ihren Erhalt (Replikation), ihre stabile Vererbung und ihre Mobilität zuständig sind. Darüber hinaus verfügen die Plasmide über akzessorische Elemente, die zu einem großen Teil aus mobilen genetischen Elementen (Insertionssequenzen und Transposons) zusammengesetzt sind.

#### *4.2 Abgleich der einzelnen Reads gegen eine Plasmid- und eine Resistenzdatenbank*

Für eine Auswertung der genetischen Information, die in den 346.427 Reads enthalten ist, wurde ein Abgleich (Blast2n) der DNA-Sequenz von jedem Read gegen zwei Datenbanken mit Hilfe des SAMS-Programms (Sequenzanalyse und -management System) durchgeführt. Die erste Datenbank besteht aus kodierenden Regionen von ca. 870 bereits komplett

sequenzierten Plasmiden, die in der NCBI-Datenbank ([http://www.ncbi.nlm.nih.gov/genomes/static/eub\\_p.html](http://www.ncbi.nlm.nih.gov/genomes/static/eub_p.html)) zugänglich sind. Es konnten ca. 49.000 Reads auf die Plasmid-kodierten Gene der Datenbank kartiert werden. In Tabelle A2 (siehe Anhang) sind die Plasmid-kodierten Gene dargestellt, die durch mehr als 100 Reads getroffen wurden. Zu den am häufigsten getroffenen Genen gehören Replikations-, Partitions- und Mobilisierungsgene sowie Gene, die für Toxine, Resistenz, mobile genetische Elemente (Transposons und IS-Elemente) und hypothetische Proteine kodieren.

Die zweite Datenbank besteht aus den Resistenzgensequenzen, die für die Entwicklung des Resistenz-Microarrays verwendet wurden, das als Auftragsforschung in Zusammenarbeit mit dem Landesumweltamt Nordrhein-Westfalen entwickelt wurde (Projekt: Entwicklung und Evaluierung eines Resistenz-Microarrays für die Detektion von Resistenzdeterminanten in Bakterien aus Kläranlagen). Bei dem Vergleich der Plasmid-Mobilom Reads mit der Resistenzgensequenz-Datenbank wurden die DNA-Sequenzen von 66 Resistenzgenen und fünf Plasmid-spezifischen Genen durch insgesamt 3817 Reads getroffen. Die getroffenen Resistenzgene und deren Charakteristika sind in Tabelle A3 (siehe Anhang) dargestellt.

#### *4.3 Assemblierung der einzelnen Sequenzen und die Auswertung in SAMS*

Eine Assemblierung der einzelnen Sequenzen mit dem Newbler Assembler von 454 Life Sciences (20 Commercial Street, Branford, CT 06405, U.S.A.) führte zu 605 Contigs mit einer Mindestlänge von 500 Basen und maximal 4341 Basen. Die 605 Contigs wurden zwecks Annotation in das SAMS-Programm (Sequenzanalyse- und -management System) geladen. Das Programm ermöglicht die gleichzeitige Analyse von Sequenzen unter Verwendung verschiedener bioinformatischer Werkzeuge sowie eine Archivierung der erhaltenen Daten. Die Sequenzen der Contigs wurden gegen folgende Datenbanken abgeglichen:

- nt (non-redundant NCBI nucleotide database),
- COG (Cluster of Orthologous Groups),
- KEGG (Kyoto Encyclopedia of Genes and Genomes),
- SP (Swiss-Prot, annotated protein sequence database),
- nr (nucleotide database) und
- Interpro (database of protein families, domains and functional sites)

Aufgrund fehlender Personalmittel im Projekt war leider eine detaillierte Auswertung der Daten bisher nicht möglich. Nach erster Sichtung der Annotationsergebnisse kann festgehalten werden, dass viele der abgeleiteten Genprodukte (bzw. Teile von Genprodukten)

Ähnlichkeiten zu Replikations- und Mobilisierungsproteinen aufweisen. Diese auffällig hohe Häufigkeit der genannten Sequenzen ist wahrscheinlich auf den Selektionsdruck zurückzuführen, welcher auf dem Erhalt dieser für die stabile Erhaltung der Plasmide notwendigen Funktionen liegt. Des Weiteren zeigen viele abgeleitete Genprodukte Homologie zu mobilen genetischen Elementen (wie z.B. Transposons und IS-Elementen), was auf eine weite Verbreitung bestimmter mobiler genetischer Elemente schließen lässt. Weiterhin ist festzuhalten, dass die Referenzproteine fast ausnahmslos von Genen kodiert werden, die selbst auf Plasmiden lokalisiert sind (sowohl bekannten Resistenz- als auch Virulenzplasmiden).

#### 4.4 Translation der DNA-Sequenzen der 346.427 Reads und Abgleich gegen die Pfam-Datenbank

Die aus der Sequenzierung des Belebtschlamm-spezifischen Plasmid-Mobiloms erhaltenen 346.427 Reads wurden in allen sechs Leserastern in entsprechende Aminosäure-Sequenzen übersetzt und mit der Pfam- (Proteinfamilien-) Datenbank verglichen, um eine Kategorisierung der Reads in funktionelle Familien zu erhalten. Dabei wurde sichergestellt, dass ein Read nur einer Proteinfamilie zugeordnet wird. Insgesamt konnten 59.058 Reads rund 1050 Proteinfamilien zugeordnet werden. Die Tatsache, dass „nur“ ein Sechstel der Reads zu einer Proteinfamilie zugeordnet werden konnte ist einerseits auf die sehr kurze Länge der abgeleiteten Aminosäuresequenz zurückzuführen (u. U. ist die konservierte Domäne nicht vollständig und für eine Zuordnung zu kurz). Andererseits enthalten Plasmide oft Gene, die keine klassifizierten Domänen enthalten und somit nicht gruppiert werden können. Einige identifizierte Proteinfamilien sind in Tabelle A4 (siehe Anhang) dargestellt.

Die im Folgenden beschriebenen Funktionen *Replication* (Replikation), *Stability* (Stabilität), *Mobilization* (Plasmid-Mobilisierung) und *Transfer* (konjugativer Transfer), die zusammen durch 26.144 Reads repräsentiert werden, zählen zu dem sogenannten Plasmid-Rückgrat.

Unter der Rubrik *Replication* sind die Familien von Proteinen zusammengefasst, die in die Replikation der Plasmide involviert sind. Dazu gehören Proteine, die z.B. für die Initiation der Replikation, für die Erkennung des Ursprungs der vegetativen Replikation (*oriV*) und DNA-Einzelstrangbindung zuständig sind oder auch die Primasefunktion ausüben. Die große Anzahl der Reads, die in diese Kategorie fallen (13.449 Reads), ist nicht ungewöhnlich, da die entsprechenden Proteine eine für die Erhaltung der Plasmide essentielle Funktion (die Replikation nämlich) übernehmen.

Zu den Proteinfamilien, die direkt mit der Stabilität (*Stability*) der Plasmide zu tun haben,

konnten 4310 Reads zugeordnet werden. Die entsprechenden Gene kodieren u.a. für Plasmidpartition (stabile Verteilung der Plasmide auf Tochterzellen während der Zellteilung) oder für Toxin-Antitoxin Systeme. Diese Toxin/Antitoxin Systeme stellen sicher, dass nur die Tochterzellen überleben, die das Antitoxin-kodierende Plasmid nach der Zellteilung erhalten haben.

Zwei weitere Funktionen, die in Tabelle A4 dargestellt sind, nämlich *Mobilization* (Plasmid-Mobilisierung – 7081 Reads) und *Transfer* (konjugativer Transfer – 1304 Reads), können unter dem Oberbegriff Mobilität (insgesamt 8385 Reads) zusammengefasst werden. Während die Kategorie *Mobilization* die Mobilisierung von Plasmiden aus Donoren in Rezipienten durch sogenannte Helferplasmide beschreibt, bezieht sich *Transfer* auf Plasmide, die selbst für ihren konjugativen Transfer kodieren und als selbst-transmissible Plasmide bezeichnet werden. Die Mobilität von Plasmiden spielt eine wichtige Rolle für deren Verbreitung und somit auch für die Verbreitung von Resistenzdeterminanten. Der große Unterschied in der Zahl der Reads, die in die Bereiche Plasmid-Mobilisierung und konjugativer Transfer klassifiziert wurden, könnte als ein Hinweis auf eine größere Anzahl an mobilisierbaren Plasmiden in der Umwelt gedeutet werden.

Die akzessorischen Funktionen *MGE* (*Mobile Genetic Elements* - mobile genetische Elemente), *Resistance* (Resistenz), *Transporter* (Transporter), *Virulence* (Virulenz) und *Restriction/ Modification* (Restriktions- und Modifizierungssysteme) werden durch insgesamt 17.996 Reads repräsentiert (Tabelle A4). Reads, die zu den mobilen genetischen Elemente zugeordnet wurden, machen mit 13.102 den bei Weitem größten Anteil der Reads aus dem Bereich der akzessorischen Funktionen aus. Hier sind Proteinfamilien vorzufinden, die sowohl Transposon- als auch Insertionssequenz-spezifische Transposasen repräsentieren.

Zum Bereich *Resistance* konnten 3066 Reads zugeordnet werden. Zu dieser Kategorie gehören Proteinfamilien, die repräsentativ für Antibiotikaresistenz und Resistenz gegen Toxine und Schwermetalle sind.

Zu dem funktionellen Bereich *Transporter* konnten 1580 Reads zugeordnet werden. Hierzu gehören u.a. Proteine der ABC-Transporter (*ATP-binding cassette*), der MFS-Transporter (*major facilitator family*), der RND-Efflux Systeme (*resistance-nodulation-cell division*) und der Sulfattransporter. Es ist davon auszugehen, dass ein Teil dieser Transporter nicht nur für die Versorgung mit Nährstoffen bzw. den Export von Metaboliten sondern auch für den Export von Antibiotika zuständig ist. Innerhalb des *Virulence*-Bereichs konnten Proteinfamilien identifiziert werden, die u.a. für die Eisenaufnahme, Typ II und Typ III

Sekretionssysteme und Fimbrien charakteristisch sind. Es ist bemerkenswert, dass Virulenzdeterminanten aus Belebtschlamm-Bakterien isoliert werden konnten, obwohl keine spezielle Selektion zur Anzucht von Pathogenen vorlag. Entweder sind die identifizierten Determinanten chromosomal kodiert oder auf Plasmiden lokalisiert. Die Koexistenz von Resistenz- und Virulenzplasmiden in einer bakteriellen Zelle scheint möglich zu sein.

Unter der Rubrik Restriktions-/Modifikationssysteme (*Restriction/Modification* – 1754 Reads) sind Enzyme zusammengefasst, welche DNA restringieren bzw. vor der Wirkung der Restriktionsenzyme schützen, indem sie bestimmte Stellen der DNA (nämlich die Erkennungsstellen der Restriktionsenzyme) methylieren. Diese Systeme schützen die Plasmid-tragende Zelle vor Fremd-DNA, die in die Zelle aufgenommen wurde.

Zusammenfassend kann festgestellt werden, dass mehr Reads dem Plasmid-Rückgrat (26.144 Reads) als der Fracht (19.750 Reads) zugeordnet werden konnten. Einer der hierfür möglichen Gründe ist, dass das Plasmid-Rückgrat bei manchen Plasmid-Klassen (z.B. bei selbst-transmissiblen Plasmiden) viel größer als die Beladung sein kann. Des Weiteren ist es die genetische Information des Plasmid-Rückgrats, die für den stabilen Erhalt von Plasmiden absolut notwendig ist, während die Beladung hierfür entbehrlich ist. Hinzu kommt, dass Fracht-lose Plasmide in der Umwelt vorkommen, die potentiell mit akzessorischen Modulen ausgestattet werden können. Es ist wahrscheinlich, dass in den Plasmidisolierungen auch solche Plasmide vorhanden sind.

## **5. Ausblick**

Die erhaltenen Daten liefern sehr interessante Einblicke in die genetische Vielfalt der in der Kläranlage vorkommenden mobilen Elemente. Es steht außer Frage, dass die große Menge an Sequenzinformation weiter ausgewertet werden muss und dass auch weiterhin sehr interessante Ergebnisse zu erwarten sind. Eine Vervollständigung der Annotation der Contigs könnte z. B. mehr Informationen über die Plasmiddiversität und die genetische Fracht der Plasmide liefern. Ein Vergleich der Sequenzen der einzelnen Reads gegen die in der Mvir-Datenbank vorhandenen Sequenzen von Genen, die für Toxine, Virulenzfaktoren und Antibiotika-Resistenz kodieren, würde eine umfassende Übersicht über das Gefährdungspotential der genetischen Fracht der Kläranlagenplasmide liefern. Des Weiteren könnten die in einen Contig eingebauten Reads (siehe Punkt 4.3) sowie auf Plasmid-kodierte Gene kartierte Reads visualisiert werden, um so eine Aussage über die Abundanz der entsprechenden Sequenzen in Klärschlamm-Bakterien abzuleiten.

Die Analyse der vorliegenden Datenmenge ist jedoch sehr zeit- und personalintensiv und die entsprechenden bioinformatischen Werkzeuge sind nur ansatzweise vorhanden. Die Entwicklung neuer bioinformatischer Werkzeuge und Vorgehensweisen für die Prozessierung von 454-Sequenzdaten und die Bearbeitung spezifischer Fragestellungen wäre wünschenswert.



## 6. Literatur

**Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995).** Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**, 143-169.

**Davison, J. (1999).** Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73-91.

**Dröge, M., Pühler, A. & Selbitschka, W. (2000).** Phenotypic and molecular characterization of conjugative antibiotic resistance plasmids isolated from bacterial communities of activated sludge. *Mol Gen Genet* **263**, 471-482.

**Feuerpfeil, I., Lopez-Pila, J., Schmidt, R., Schneider, E. & Szewzyk, R. (1999).** Antibiotikaresistente Bakterien und Antibiotika in der Umwelt. *Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitsschutz* **42**, 37-50.

**Gordon, D., Abajian, C. & Green, P. (1998).** Consed: a graphical tool for sequence finishing. *Genome Res* **8**, 195-202.

**Gordon, D., Desmarais, C. & Green, P. (2001).** Automated finishing with autofinish. *Genome Res* **11**, 614-625.

**Guardabassi, L., Dalsgaard, A. & Olsen, J. E. (1999).** Phenotypic characterization and antibiotic resistance of *Acinetobacter* spp. isolated from aquatic sources. *J Appl Microbiol* **87**, 659-667.

**Hartmann, A., Alder, A. C., Koller, T. & Widmer, R. M. (1998).** Identification of fluoroquinolone antibiotics as the main source of *umuC* genotoxicity in native hospital wastewater. *Environmental Toxicology and Chemistry* **17**, 377-382.

**Hartmann, A., Golet, E. M., Gartsler, S., Alder, A. C., Koller, T. & Widmer, R. M. (1999).** Primary DNA damage but not mutagenicity correlates with ciprofloxacin concentrations in German hospital wastewaters. *Arch Environ Contam Toxicol* **36**, 115-119.

**Heuer, H., Szczepanowski, R., Schneiker, S., Pühler, A., Top, E. M. & Schlüter, A. (2004).** The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1 $\beta$  group without any accessory genes. *Microbiology* **150**, 3591-3599.

**Heuer, H., Krogerrecklenfort, E., Wellington, E. M. H., Egan, S., van Elsas, J. D., van Overbeek, L., Collard, J. M., Guillaume, G., Karagouni, A. D., Nikolakopoulou, T. L. & Smalla, K. (2002).** Gentamicin resistance genes in environmental bacteria: prevalence and transfer. *Fems Microbiology Ecology* **42**, 289-302.

**Kaiser, O., Bartels, D., Bekel, T., Goesmann, A., Kespohl, S., Pühler, A. & Meyer, F. (2003).** Whole genome shotgun sequencing guided by bioinformatics pipelines--an optimized approach for an established technique. *J Biotechnol* **106**, 121-133.

**Kümmerer, K. (2001).** Drugs in the environment: emission of drugs, diagnostic aids and disinfectants into wastewater by hospitals in relation to other sources--a review. *Chemosphere* **45**, 957-969.

**Kümmerer, K. (2003).** Significance of antibiotics in the environment. *J Antimicrob Chemother* **52**, 5-7.

**Kümmerer, K. & Henninger, A. (2003).** Promoting resistance by the emission of antibiotics from hospitals and households into effluent. *Clin Microbiol Infect* **9**, 1203-1214.

**Kümmerer, K., Al-Ahmad, A. & Mersch-Sundermann, V. (2000).** Biodegradability of some antibiotics, elimination of the genotoxicity and affection of wastewater bacteria in a simple test. *Chemosphere* **40**, 701-710.

**Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005).** Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

**Ohlsen, K., Ternes, T., Werner, G., Wallner, U., Löffler, D., Ziebuhr, W., Witte, W. & Hacker, J. (2003).** Impact of antibiotics on conjugational resistance gene transfer in *Staphylococcus aureus* in sewage. *Environ Microbiol* **5**, 711-716.

**Schlüter, A., Heuer, H., Szczepanowski, R., Forney, L. J., Thomas, C. M., Pühler, A. & Top, E. M. (2003).** The 64 508 bp IncP-1 $\beta$  antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1 $\beta$  group. *Microbiology* **149**, 3139-3153.

**Schlüter, A., Heuer, H., Szczepanowski, R., Poler, S. M., Schneiker, S., Pühler, A. & Top, E. M. (2005).** Plasmid pB8 is closely related to the prototype IncP-1 $\beta$  plasmid R751 but transfers poorly to *Escherichia coli* and carries a new transposon encoding a small multidrug resistance efflux protein. *Plasmid* **54**, 135-148.

**Seveno, N. A., Kallifidas, D., Smalla, K., van Elsas, J. D., Collard, J. M., Karagouni, A. D. & Wellington, E. M. H. (2002).** Occurrence and reservoirs of antibiotic resistance genes in the environment. *Reviews in Medical Microbiology* **13**, 15-27.

**Smalla, K. & Sobecky, P. A. (2002).** The prevalence and diversity of mobile genetic elements in bacterial communities of different environmental habitats: insights gained from different methodological approaches. *Fems Microbiology Ecology* **42**, 165-175.

**Szczepanowski, R., Krahn, I., Linke, B., Goesmann, A., Pühler, A. & Schlüter, A. (2004).** Antibiotic multiresistance plasmid pRSB101 isolated from a wastewater treatment plant is related to plasmids residing in phytopathogenic bacteria and carries eight different resistance determinants including a multidrug transport system. *Microbiology* **150**, 3613-3630.

**Szczepanowski, R., Braun, S., Riedel, V., Schneiker, S., Krahn, I., Pühler, A. & Schlüter, A. (2005).** The 120 592 bp IncF plasmid pRSB107 isolated from a sewage-treatment plant

encodes nine different antibiotic-resistance determinants, two iron-acquisition systems and other putative virulence-associated functions. *Microbiology* **151**, 1095-1111.

**Tauch, A., Schlüter, A., Bischoff, N., Goesmann, A., Meyer, F. & Pühler, A. (2003).** The 79,370-bp conjugative plasmid pB4 consists of an IncP-1 $\beta$  backbone loaded with a chromate resistance transposon, the *strA-strB* streptomycin resistance gene pair, the oxacillinase gene *bla*<sub>NPS-1</sub>, and a tripartite antibiotic efflux system of the resistance-nodulation-division family. *Mol Genet Genomics* **268**, 570-584.

**Tennstedt, T., Szczepanowski, R., Braun, S., Pühler, A. & Schlüter, A. (2003).** Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *Fems Microbiology Ecology* **45**, 239-252.

**Tennstedt, T., Szczepanowski, R., Krahn, I., Pühler, A. & Schlüter, A. (2005).** Sequence of the 68,869 bp IncP-1 $\alpha$  plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. *Plasmid* **53**, 218-238.

**Vorhölter, F. J., Thias, T., Meyer, F., Bekel, T., Kaiser, O., Pühler, A. & Niehaus, K. (2003).** Comparison of two *Xanthomonas campestris* pathovar *campestris* genomes revealed differences in their gene composition. *J Biotechnol* **106**, 193-202.

## 7. Anhang

### 7.1 Tabellen

Tabelle A1: Kategorisierung der Reads der *Gene Ontology* Terminologie entsprechend.

<b>Class</b>	<b>Term</b>	<b>Number of hits</b>
cellular component	extrachromosomal circular DNA	11468
cellular component	membrane	1588
cellular component	intracellular	963
cellular component	integral to membrane	586
cellular component	cytoplasm	198
cellular component	outer membrane	186
cellular component	chromosome	139
biological process	DNA transposition	8931
biological process	DNA replication initiation	4889
biological process	plasmid maintenance	4411
biological process	DNA recombination	4321
biological process	unidirectional conjugation	2107
biological process	DNA replication	1809
biological process	DNA integration	1670
biological process	regulation of transcription DNA-dependent	1528
biological process	DNA methylation	1218
biological process	electron transport	1195
biological process	response to antibiotic	1168
biological process	response to oxidative stress	1126
molecular function	DNA binding	19196
molecular function	transposase activity	8754
molecular function	DNA-directed DNA polymerase activity	4766
molecular function	recombinase activity	2310
molecular function	ATP binding	1301
molecular function	sequence-specific DNA binding	1067
molecular function	catalase activity	1016

Tabelle A2: Ähnlichkeit der Plasmid-Mobilom Reads zu bekannten Plasmid-lokalisierten Kodierregionen.

Accession	Zahl der Homologie zu Reads	
YP_133858.1	114	Pin, invertase/recombinase-like protein, pRSB101, uncultured bacterium
NP_857609.1	118	rep protein, pCR1, Corynebacterium renale
YP_213946.1	156	RepM, DNA replication protein, pMAC, Acinetobacter baumannii
YP_449007.1	163	replication protein, pMUR050, Escherichia coli
NP_710174.1	990	RepA, putative RepA protein, pAsal2, Aeromonas salmonicida subsp. salmonicida
NP_710167.1	1426	putative RepA protein, pAsal1, Aeromonas salmonicida subsp. salmonicida
NP_861554.1	1108	Rep, replication primase, pAsa1, Aeromonas salmonicida subsp. salmonicida A449
NP_861563.1	1280	Rep. replication primase, pAsa3, Aeromonas salmonicida subsp. salmonicida A449
YP_245435.1	1921	ParA, partitioning protein, Rms149, Pseudomonas aeruginosa
YP_133838.1	202	ParE-like, hypothetical protein, pRSB101, uncultured bacterium
YP_133839.1	175	ParD-like, hypothetical protein, pRSB101, uncultured bacterium
NP_046807.1	126	MobA, mobilisation protein, pUCD5000, Pantoea citrea
YP_245456.1	106	MobA, mobilisation protein, Rms149, Pseudomonas aeruginosa
NP_861556.1	303	MobC, mobilization protein, pAsa1, Aeromonas salmonicida subsp. salmonicida A449
NP_861557.1	626	MobA, mobilization relaxase, pAsa1, Aeromonas salmonicida subsp. salmonicida A449
NP_861558.1	205	MobB, mobilization protein, pAsa1, Aeromonas salmonicida subsp. salmonicida A449
NP_861565.1	401	MobC, mobilization protein, pAsa3, Aeromonas salmonicida subsp. salmonicida A449
NP_861566.1	812	MobA, mobilization relaxase, pAsa3, Aeromonas salmonicida subsp. salmonicida A449
NP_861568.1	114	MobD, mobilization protein, pAsa3, Aeromonas salmonicida subsp. salmonicida A449
NP_861572.1	161	MobC, mobilization protein, pAsa2, Aeromonas salmonicida subsp. salmonicida A449
NP_861573.1	750	MobA, mobilization relaxase, pAsa2, Aeromonas salmonicida subsp. salmonicida A449
NP_861574.1	125	MobB, mobilization protein, pAsa2, Aeromonas salmonicida subsp. salmonicida A449
NP_710164.1	769	MobA, putative MobA protein, pAsal1, Aeromonas salmonicida subsp. salmonicida
NP_710170.1	576	MobA, putative MobA protein, pAsal2, Aeromonas salmonicida subsp. salmonicida
NP_710171.1	202	MobB, putative MobB protein, pAsal2, Aeromonas salmonicida subsp. salmonicida
NP_710177.1	206	MobC, putative MobC protein, pAsal3, Aeromonas salmonicida subsp. salmonicida

NP_710178.1	621	MobA, putative MobA protein, pAsal3, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i>
NP_710179.1	125	MobB, putative MobB protein, pAsal3, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i>
YP_232873.1	133	MobB, mobilization protein B, pCCK381, <i>Pasteurella multocida</i>
NP_604399.1	468	MbeA, relaxase, pK, <i>Salmonella enteritidis</i>
NP_052607.1	128	TagA, ToxR-regulated lipoprotein, pO157, <i>Escherichia coli</i> O157:H7 str. Sakai
NP_861553.1	221	RelE, toxin, pAsa1, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
NP_861560.1	231	RelB, antitoxin, pAsa3, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
NP_861561.1	158	RelE, toxin, pAsa3, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
NP_710173.1	389	putative RelE protein, pAsal2, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i>
YP_449008.1	158	ATP-dependent Lon protease, bacterial type, pMUR050, <i>Escherichia coli</i>
NP_203135.1	109	putative IpaJ protein, pSFD10, <i>Salmonella choleraesuis</i>
NP_040356.1	261	Cea, colicin E1 protein, Cole1, <i>Escherichia coli</i>
YP_449009.1	314	putative macrolide 2'-phosphotransferase, pMUR050, <i>Escherichia coli</i>
YP_449010.1	454	macrolide-specific ABC-type efflux carrier, pMUR050, <i>Escherichia coli</i>
NP_775053.1	395	Mel, probable macrolide efflux protein, pCTX-M3, <i>Citrobacter freundii</i>
NP_775055.1	170	Mph2, hypothetical protein, pCTX-M3, <i>Citrobacter freundii</i>
NP_437580.1	161	AmpC, putative exported beta-lactamase protein, pSymB, <i>Sinorhizobium meliloti</i> 1021
NP_478073.1	148	QacEA1, Small Multidrug Resistance protein, pTet3, <i>Corynebacterium glutamicum</i>
NP_758542.1	225	Orf188, putative ABC-transporter subunit, pCAR1, <i>Pseudomonas resinovorans</i>
NP_068714.1	619	Rom protein, pKleB-k17/80, <i>Klebsiella pneumoniae</i>
YP_067858.1	799	methyl-accepting chemotaxis protein, pFBAOT6, <i>Aeromonas punctata</i>
NP_085428.1	180	TnpA, transposition transposase, pWR501-Tn501, <i>Shigella flexneri</i>
NP_598131.1	232	TnpA2, transposase, pB4, uncultured bacterium
NP_598137.1	142	TnpR, resolvase, pB4, uncultured bacterium
NP_598138.1	1078	TnpA, transposase, pB4, uncultured bacterium
NP_598177.1	156	TnpA, transposase, pB4, uncultured bacterium
NP_957548.1	555	TnpA, transposase, pC15-1a, <i>Escherichia coli</i>
NP_957549.1	127	TnpR, resolvase, pC15-1a, <i>Escherichia coli</i>
YP_067843.1	116	transposase, pFBAOT6-IS630-family, <i>Aeromonas punctata</i>
YP_067860.1	554	TnpA, transposase, pFBAOT6-Tn1721, <i>Aeromonas punctata</i>
YP_067891.1	436	TnpA, transposase, pFBAOT6-Tn3-family, <i>Aeromonas punctata</i>
YP_190208.1	215	TnpA, transposase, pAPEC-O2-R-Tn1721, <i>Escherichia coli</i>
YP_161722.1	124	TnpA, hypothetical transposase, pMOL28-Tn4378, <i>Ralstonia metallidurans</i> CH34

YP_581997.1	171	TnpA, transposase, plasmid-2-Tn3, <i>Ralstonia metallidurans</i> CH34
NP_943466.1	179	TnpA, transposase, pLVPK, <i>Klebsiella pneumoniae</i>
NP_569374.1	360	TnpA, transposase, pHCM1, <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18
NP_943128.1	108	TnpA, transposase, pND6-1, <i>Pseudomonas</i> sp. ND6
NP_862256.1	175	TnpA, transposase, pTP10, <i>Corynebacterium striatum</i>
NP_620616.1	197	transposase, pMGD2, <i>Klebsiella</i> sp. KCL-2
NP_639997.1	613	Orf36, transposase, Rts1, <i>Proteus vulgaris</i>
YP_209313.1	441	TnpA, transposase, pSC138, <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis str. SC-B67
NP_940688.1	125	TnpA, transposase, pPSR1, <i>Pseudomonas syringae</i> pv. <i>syringae</i>
NP_863363.1	148	TnpA, transposase, R64-Tn5393, <i>Salmonella typhimurium</i>
YP_213947.1	190	Orf2, hypothetical protein, pMAC, <i>Acinetobacter baumannii</i>
NP_049455.1	177	Orf5, hypothetical protein, pRAY, <i>Acinetobacter</i> sp. SUN
NP_049459.1	165	Orf3, hypothetical protein pRAY, <i>Acinetobacter</i> sp. SUN
NP_049460.1	1198	Orf2, hypothetical protein pRAY, <i>Acinetobacter</i> sp. SUN
NP_775056.1	102	Orf44, hypothetical protein, pCTX-M3, <i>Citrobacter freundii</i>
NP_862996.1	530	Orf33, hypothetical protein p1658/97, <i>Escherichia coli</i>
NP_861555.1	226	Orf1, hypothetical protein, pAsa1, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
NP_861564.1	648	Orf1, hypothetical protein, pAsa3, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
NP_861571.1	277	Orf1, hypothetical protein, pAsa2, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449
YP_618235.1	238	Orf1, hypothetical protein, pHW15, <i>Rahnella</i> sp. 'WMR15'
NP_045376.1	1042	OrfI, hypothetical protein, pBERT, <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Berta
NP_957538.1	192	Orf5, hypothetical protein, p9123, <i>Escherichia coli</i>
YP_245475.1	148	Ofn41, hypothetical protein, Rms149, <i>Pseudomonas aeruginosa</i>
NP_058421.1	111	R0210, hypothetical protein, R27, <i>Salmonella typhi</i>
NP_710168.1	216	Orf36, hypothetical protein, pAsal2, <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i>
NP_904288.1	112	Orf6, hypothetical protein, pUO1, <i>Delftia acidovorans</i>
YP_009034.1	431	hypothetical protein, pSERF1, <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis

---

Tabelle A3: Anzahl der Reads des Plasmid-Mobiloms, die mit der Sequenz ausgewählter Resistenzgene übereinstimmen.

Name	Genprodukt	Anzahl identischer Reads	Resistenz / Funktion #	Accession
<i>aacA</i>	aminoglycoside 6'-N-acetyltransferase	12	Km, Tob, Ak	M86913
<i>aacA4</i>	aminoglycoside 6'-acetyltransferase	106	Ak	AJ744860
<i>aacA29b</i>	aminoglycoside-6'-N-acetyltransferase	4	Ak, Km	AY139599
<i>aacC1</i>	aminoglycoside-3N-acetyltransferase	2	Gm	AY139604
<i>aacC2</i>	aminoglycoside-(3)-N-acetyltransferase	7	Gm	S68058
<i>aacC3</i>	aminoglycoside-(3)-N-acetyltransferase	2	Gm	X13543
<i>aadA1</i>	put. streptomycin 3"-adenylyltransferase	30	putative Sm, Sp	AJ744860
<i>aadA2</i>	put. streptomycin 3"-adenylyltransferase	2	putative Sm, Sp	AY115474
<i>aadA5</i>	streptomycin 3"-adenylyltransferase	2	Sm, Sp	AY139588
<i>aadA6/aadA10</i>	aminoglycoside (3")(9) adenylyltransferase	70	Sm, Sp	AM087405
<i>aadA8</i>	put. streptomycin 3"-adenylyltransferase	10	putative Sm, Sp	AY139603
<i>aadA11</i>	put. streptomycin 3"-adenylyltransferase	9	putative Sm, Sp	AJ567827
<i>aadA12</i>	put. streptomycin 3"-adenylyltransferase	13	putative Sm, Sp	AY665771
<i>aadA13</i>	streptomycin 3"-adenylyltransferase	28	Sm, Sp	AY940492
<i>aadA23</i>	put. streptomycin 3"-adenylyltransferase	8	putative Sm, Sp	AJ809407
<i>aadB</i>	aminoglycoside 6'-N-acetyltransferase	4	Gm	AY139594
<i>aph</i>	aminoglycoside 3'-phosphotransferase	24	Km, Neo	AJ851089
<i>aphA</i>	3'-aminoglycoside phosphotransferase	42	Km	AJ744860
<i>aphA-6</i>	3'-aminoglycoside phosphotransferase	3	Km, Ak	X07753
<i>strA</i>	aminoglycoside 3'-phosphotransferase	347	Sm	NC_004840
<i>strB</i>	aminoglycoside-6-phosphotransferase	349	Sm	NC_004840
<i>ges-3</i>	class A extended spectrum $\beta$ -lactamase	41	Titeracillin, Prl, CAZ, CTX, Atm, Ipm	AY494717
<i>per-1</i>	class A $\beta$ -lactamase	2	$\beta$ -lactams	Z21957
<i>shv-34</i>	class A $\beta$ -lactamase	8	CAZ, CTX	AY036620
<i>tem-1</i>	class A $\beta$ -lactamase	30	Amp, Pen-G	AJ851089
<i>tla-2</i>	class A extended spectrum $\beta$ -lactamase	5	Amo, Tic, Caz, Kf, Cxm, Fox, Ctx, Fep, Atm	NC_006385



<i>cmy-9</i>	class C $\beta$ -lactamase	27	$\beta$ -lactams	AB061794
<i>cmy-10</i>	class C $\beta$ -lactamase	13	$\beta$ -lactams	AF357597
<i>oxa-1</i>	class D $\beta$ -lactamase	55	$\beta$ -lactams	AY139600
<i>oxa-2</i>	class D $\beta$ -lactamase	47	$\beta$ -lactams	NC_007502
<i>oxa-5</i>	class D $\beta$ -lactamase	8	$\beta$ -lactams	X58272
<i>oxa-10</i>	class D $\beta$ -lactamase	20	$\beta$ -lactams	AY115475
<i>oxa-56</i>	class D $\beta$ -lactamase	2	$\beta$ -lactams	AY445080
<i>oxa-58</i>	class D $\beta$ -lactamase	19	Amo, Tic, Cpo, Prl, Ipm, Kf	AY665723
<i>cmlA1</i>	chloramphenicol efflux protein	2	Cm	NC_006388
<i>cmlA5</i>	chloramphenicol efflux protein	3	Cm	AY115475
<i>cat2</i>	chloramphenicol acetyltransferase	2	Cm	AY509004
<i>catA</i>	chloramphenicol acetyltransferase	6	Cm	AJ851089
<i>catB2</i>	chloramphenicol acetyltransferase	10	Cm	AY139601
<i>catB4</i>	chloramphenicol acetyltransferase	36	Cm	AF322577
<i>qnrS2</i>	quinolon resistance determinant	40	Cip, Nor, Nal	DQ460733
<i>mph(A)</i>	macrolide 2'-phosphotransferase I	11	Azi, Cla, Em, Rox	NC_006385
<i>mph</i>	macrolide 2'-phosphotransferase	550	Em	DQ839391
<i>erm(B)</i>	rRNA adenin N-6-methyltransferase	17	Em	M11180
<i>erm(GT)</i>	rRNA adenin N-6-methyltransferase	3	macrolide-lincosamide-streptogramin B	M64090
<i>mel</i>	macrolide efflux protein, macrolide-specific ABC-type efflux carrier	856	Azi, Cla, Em	DQ839391
<i>arr2</i>	put. rifampin ADP-ribosyl transferase	3	Rif	AF205943
<i>sulI</i>	dihydropteroate synthetase	122	Sul	NC_006388
<i>sulII</i>	dihydropteroate synthetase	94	Sul	AJ851089
<i>dfrV</i>	dihydrofolate reductase	3	Tp	AY139589
<i>dfrXII = eigentlich dfrVII</i>	dihydrofolate reductase	5	Tp	AY139605
<i>dhfrI</i>	dihydrofolate reductase	11	Tp	AJ698325
<i>dhfrVIII</i>	dihydrofolate reductase	3	Tp	U10186
<i>tet(A)</i>	MFS tetracycline efflux	46	Tc	NC_004840
<i>tet(A)</i>	MFS tetracycline efflux	9	Tc	NC_006388

<i>tet(D)</i>	MFS tetracycline efflux	3	Tc	L06798
<i>tetA(C)</i>	MFS tetracycline efflux	12	Tc	AJ851089
<i>tetA(E)</i>	MFS tetracycline efflux	20	Tc	L06940
<i>tet(39)</i>	MFS tetracycline efflux	417	Tc	AY743590
<i>tet(X)</i>	inactivation of tetracycline	27	Tc	M37699
<i>qacEΔ1</i>	small multidrug resistance protein, membrane transporter of cations and cationic drugs	65	quaternary ammonium compounds	AJ698325
<i>qacF</i>	small multidrug resistance protein, membrane transporter of cations and cationic drugs	3	quaternary ammonium compounds	AY139598
<i>qacH</i>	small multidrug resistance protein, membrane transporter of cations and cationic drugs	2	quaternary ammonium compounds	AY139595
<i>acrB</i>	RND family, acridine/multidrug efflux pump	4	multidrug efflux	M94248
<i>mexB</i>	cation/multidrug efflux pump, RND multidrug efflux transporter	4	multidrug efflux	L11616
<i>orf11</i>	ABC type permease	46	Nal, Nor	NC_006385
<i>rep</i>	replication initiation protein	2	IncA/C-specific gene	X73674
<i>repE</i>	replication initiation protein	9	IncFIA-specific replication gene	AJ851089
<i>kikA</i>	killing in <i>Klebsiella</i>	4	IncN-specific gene	AY046276
<i>trfA</i>	replication initiation protein	3	initiation of replication, IncP-specific gene	NC_004840
<i>oriV</i>	origin of the vegetative replication	13	IncQ-specific region	NC_001740

# Abbreviations: Ak - amikacin; Amo - amoxicillin; Amp - ampicillin; Atm - aztreonam; Azm - azithromycin; Car - carbenicillin; Caz - ceftazidim; Cec - cefaclor; Cfp - cefoperazon; Cft - ceftibuten; Cip - ciprofloxacin; Clr - clarithromycin; Cm - chloramphenicol; Cpo - cefpirom; Cro - ceftriaxon; Ctx - cefotaxim; Cxm - cefuroxim; Em - erythromycin; Fep - cefepim; Ffc - florfenicol; Fox - ceftoxitin; Gm - gentamicin; Ipm - imipenem; Kf - cephalothin; Km - kanamycin; Lev - levofloxacin; Met - methicillin; Mez - mezlocillin; Nal - nalidixic acid; Neo - neomycin; Nor - norfloxacin; Nv - novobiocin; Ob - cloxacillin; Ofx - ofloxacin; Pen-G - penicillin G; Prl - piperacillin; Rif - rifampicin; Rox - roxithromycin; Spar - sparfloxacin; Sm - streptomycin; Sp - spectinomycin; Sul - sulfonamides; Tc - tetracyclines; Tic - ticarcillin; Tob - tobramycin; Tp - trimethoprim; Ty - tylosin.

Tabelle A4: Einordnung der Reads des Plasmid-Mobiloms in Proteinfamilien.

<b>Pfam</b>	<b>Pfam No.</b>	<b>Description</b>	<b>Number of hits</b>
<b>Replication</b>			
RepA_C	04796	Plasmid encoded RepA protein	32
Rep_1	01446	Replication protein	599
Rep_2	01719	Plasmid replication protein	26
Rep_3	01051	Initiator Replication protein	4760
RepC	06504	Replication protein C (RepC)	230
Replicase	03090	Replicase family	4277
PriCT_1	08708	Primase C terminal 1 (PriCT-1)	2665
Rop	01815	Rop protein	262
Rep_trans	02486	Replication initiation factor	129
RHH_1	01402	Ribbon-helix-helix protein, copG family	426
SSB	00436	Single-strand binding protein family	21
TrfA	07042	TrfA protein	9
IncFII_repA	02387	IncFII RepA protein family	4
RP-C	03428	Replication protein C	3
RepA1_leader	08048	Tap RepA1 leader peptide	2
RepA_N	04057	Replication initiator protein A (RepA) N-terminus	1
RepL	05732	Firmicute plasmid replication protein (RepL)	2
L_lactis_RepB_C	06430	<i>Lactococcus lactis</i> RepB C-terminus	1
<b>Stability</b>			
ParBc	02195	ParB-like nuclease domain	42
ParB	08775	ParB family	2
ParG	09274	ParG	10
Plasmid_stabil	05016	Plasmid stabilisation system protein	2138
RelB	04221	RelB antitoxin	820
CbiA	01656	CobQ/CobB/MinD/ParA nucleotide binding domain	609
CcdA	07362	Post-segregation antitoxin CcdA	175
CcdB	01845	CcdB protein	128
PemK	02452	PemK-like protein	73
Plasmid_Txe	06769	Plasmid encoded toxin Txe	92
Plasmid_killer	05015	Plasmid maintenance system killer protein	113
Histone_HNS	00816	H-NS histone family	44
StbA	06406	StbA protein	28
KorB	08535	KorB domain	16
KorB_C	06613	KorB C-terminal beta-barrel domain	6
RepB	07506	RepB plasmid partitioning protein	4
MipZ	09140	ATPase MipZ	4
Epsilon_antitox	08998	Bacterial epsilon antitoxin	4
Zeta_toxin	06414	Zeta toxin	1
KicB	03882	KicB killing factor	1
<b>Mobilization</b>			
MobA_MobL	03389	MobA/MobL family	1986
MobC	05713	Bacterial mobilisation protein (MobC)	978

MbeD_MobD	04899	MbeD/MobD like	483
MbeB_N	04837	MbeB-like, N-term conserved region	477
TrwC	08751	TrwC relaxase	70
Mob_Pre	01076	Plasmid recombination enzyme	1799
Relaxase	03432	Relaxase/Mobilisation nuclease domain	1288

### **Transfer**

TraA	05513	TraA	13
TraB_pilus	06477	TraB pilus assembly protein	79
TraB_2	07186	TraB protein	1
TraC	07820	TraC-like protein	1
TraD	06412	Conjugal transfer protein TraD	1
TraE	05309	TraE protein	35
TraG	02534	TraG/TraD family	98
TraG_N	07916	TraG-like protein, N-terminal region	107
TraH	06122	TraH protein	92
TraI	07057	DNA helicase TraI	15
TraI_2	07514	Putative helicase	9
TraK	06586	TraK protein	37
TraL	07178	TraL protein	35
Tra_M	05261	TraM protein	10
TraN	06986	Mating pair stabilisation protein TraN	10
TraP	07296	TraP protein	4
TraT	05818	Enterobacterial TraT complement resistance protein	49
TraU	06834	TraU protein	108
TraX	05857	TraX protein	17
TraY	05509	TraY family	16
TrbD	05245	Conjugal transfer protein TrbD	5
TrbF	04585	Conjugal transfer protein	12
TrbH	07283	Conjugal transfer protein TrbH	2
TrbI	03743	Bacterial conjugation TrbI-like protein	23
TrbL	04610	TrbL/VirB6 plasmid conjugal transfer protein	102
TrbM	07424	TrbM	5
Pili_assembly_N	00345	Gram-negative pili assembly chaperone	6
Pili_assembly_C	02753	Gram-negative pili assembly chaperone	3
Pilin	00114	Pilin (bacterial filament)	1
PilM	07419	PilM	3
GSPII_E	00437	Type II/IV secretion system protein	70
SLT	01464	Transglycosylase SLT domain	69
CagE_TrbE_VirB	03135	CagE, TrbE, VirB family, component of type IV transporter system	52
CagX	03524	Conjugal transfer protein	50
HicB	05534	HicB family	37
Usher	00577	Fimbrial Usher protein	33
VirB3	05101	Type IV secretory pathway, VirB3-like protei	14
VirB8	04335	VirB8 protein	29
FINO	05286	Fertility inhibition protein (FINO)	27
PAP_PilO	06864	Pilin accessory protein (PilO)	17
T4SS	07996	Type IV secretion system proteins	7

## **MGE**

Transposase_2	01385	Probable transposase	36
Transposase_7	01526	Transposase	6913
Transposase_8	01527	Transposase	254
Transposase_9	01548	Transposase	9
Transposase_11	01609	Transposase DDE domain	918
Transposase_12	01610	Transposase	13
Transposase_14	01710	Transposase	5
Transposase_17	01797	Transposase IS200 like	31
Transposase_20	02371	Transposase IS116/IS110/IS902 family	36
Transposase_25	03050	Transposase IS66 family	134
Transposase_27	03400	IS1 transposase	295
Transposase_31	04754	Putative transposase, YhgA-like	55
Transposase_32	04986	Putative transposase	29
Transposase_34	05717	IS66 Orf2 like protein	95
Transposase_35	07282	Putative transposase DNA-binding domain	14
Transposase_Tn5	02281	Transposase Tn5 dimerisation domain	12
Transposase_mut	00872	Transposase, Mutator family	244
TniB	05621	Bacterial TniB protein	96
TniQ	06527	TniQ	78
Ins_element1	03811	Insertion element protein	199
IstB	01695	IstB-like ATP binding protein	258
IstB_N	08483	IstB-like ATP binding N-terminal	20
Resolvase	00239	Resolvase, N terminal domain	1878
rve	00665	Integrase core domain	1451
PIN	01850	PIN domain	29

## **Resistance**

Beta-lactamase	00144	Beta-lactamase	1168
Transpeptidase	00905	Penicillin binding protein transpeptidase domain	95
Lactamase_B	00753	Metallo-beta-lactamase superfamily	38
APH	01636	Phosphotransferase enzyme family	348
APH_6_hur	04655	Aminoglycoside/hydroxyurea antibiotic resistance	279
NTP_transf_2	01909	Nucleotidyltransferase domain	230
NTP_transferase	00483	Nucleotidyl transferase	29
Adenyl_transf	04439	Streptomycin adenylyltransferase	12
Pribosyltran	00156	Phosphoribosyl transferase domain	8
Multi_Drug_Res	00893	Small Multidrug Resistance protein	164
Mig-14	07395	Mig-14	160
Pterin_bind	00809	Pterin binding enzyme	148
Acetyltransf_1	00583	Acetyltransferase (GNAT) family	87
Antibiotic_NAT	02522	Aminoglycoside 3-N-acetyltransferase	8
CAT	00302	Chloramphenicol acetyltransferase	8
Glyoxalase	00903	Glyoxalase/Bleomycin resistance protein/Dioxygenase family	7
Pentapeptide	00805	Pentapeptide repeats (8 copies)	54
MerC	03203	MerC mercury resistance protein	13
MerE	05052	MerE protein	5
MerR	00376	MerR family regulatory protein	53

MarR	01047	MarR family	26
Colicin_im	03857	Colicin immunity protein	14
Colicin_Pyocin	01320	Colicin immunity protein / pyocin immunity	25
Microcin	03526	Colicin E1 (microcin) immunity protein	9
DHFR_1	00186	Dihydrofolate reductase	21
RrnaAD	00398	Ribosomal RNA adenine dimethylase	20
TelA	05816	Toxic anion resistance protein (TelA)	9
CopB	05275	Copper resistance protein B precursor (CopB)	7
CopC	04234	Copper resistance protein CopC	2
CopD	05425	Copper resistance protein D	4
ArsC	03960	ArsC family	4
ArsD	06953	Arsenical resistance operon trans-acting repressor	3
TerC	03741	Integral membrane protein TerC family	3
Cloacin_immun	03513	Cloacin immunity protein	3
PagP	07017	Antimicrobial peptide resistance and lipid A acylation protein PagP	1
OstA	03968	OstA-like protein	1

### **Restriction / Modification**

N6_N4_Mtase	01555	DNA methylase	507
DNA_methylase	00145	C-5 cytosine-specific DNA methylase	505
N6_Mtase	02384	N-6 DNA Methylase	143
Methylase_S	01420	Type I restriction modification DNA specificity domain	75
Methyltransf_11	08241	Methyltransferase domain	4
MethyltransfD12	02086	D12 class N6 adenine-specific DNA methyltransferase	63
Dam	05869	DNA N-6-adenine-methyltransferase (Dam)	1
BsuBI_PstI_RE	06616	<i>BsuBI/PstI</i> restriction endonuclease C-terminus	119
ResIII	04851	Type III restriction enzyme, <i>res</i> subunit	66
Endonuc-PvuII	09225	Restriction endonuclease <i>PvuII</i>	65
HSDR_N	04313	Type I restriction enzyme R protein N terminus	48
EcoRI	02963	Restriction endonuclease <i>EcoRI</i>	45
EcoRII-C	09019	<i>EcoRII</i> C terminal	7
Endonuc-MspI	09208	Restriction endonuclease <i>MspI</i>	24
HNH	01844	HNH endonuclease	6
Mrr_cat	04471	Restriction endonuclease	6
DpmII	04556	<i>DpmII</i> restriction endonuclease	5
NgoMIV_restric	09015	<i>NgoMIV</i> restriction enzyme	5
Endonuc-HincII	09226	Restriction endonuclease <i>HincII</i>	4
XhoI	04555	Restriction endonuclease <i>XhoI</i>	4
EcoRII-N	09217	Restriction endonuclease <i>EcoRII</i> , N-terminal	3
Endonuc-EcoRV	09233	Restriction endonuclease <i>EcoRV</i>	3
BamHI	02923	Restriction endonuclease <i>BamHI</i>	2
Endonuclease_1	04231	Endonuclease I	3
Endonuc-BglII	09195	Restriction endonuclease <i>BglII</i>	1
EcoEI_R_C	08463	<i>EcoEI</i> R protein C-terminal	1
Antirestrict	03230	Antirestriction protein	22

ArdA	07275	Antirestriction protein (ArdA)	17
------	-------	--------------------------------	----

### **Transporter**

ABC_tran	00005	ABC transporter	495
ABC_membrane	00664	ABC transporter transmembrane region	10
ABC2_membrane	01061	ABC-2 type transporter	19
ABC-3	00950	ABC 3 transport family	10
Sulfate_transp	00916	Sulfate transporter family	305
MFS_1	07690	Major Facilitator Superfamily	272
Cation_efflux	01545	Cation efflux family	211
ACR_tran	00873	AcrB/AcrD/AcrF family	130
OEP	02321	Outer membrane efflux protein	40
Chromate_transp	02417	Chromate transporter	29
MerT	02411	MerT mercuric transport protein	20
ArsB	02040	Arsenical pump membrane protein	13
BCCT	02028	BCCT family transporter	8
FtsX	02687	Predicted permease	7
OmpA	00691	OmpA family	4
OmpA_membrane	01389	OmpA-like transmembrane domain	3
Porin_2	02530	Porin subfamily	3
OprD	03573	outer membrane porin, OprD family	1

### **Virulence**

VapD_N	04605	Virulence-associated protein D (VapD) conserved region	60
HlyD	00529	HlyD family secretion protein	28
Ferritin	00210	Ferritin-like domain	23
Zot	05707	Zonular occludens toxin (Zot)	20
Big_2	02368	Bacterial Ig-like domain (group 2)	18
HIM	05662	Haemagglutinin	13
PilS	08805	PilS N terminal	11
Strep_67kDa_ant	06100	Streptococcal 67 kDa myosin-cross-reactive antigen like family	10
YadA	03895	YadA-like C-terminal region	9
VirE	05272	Virulence-associated protein E	8
Hep_Hag	05658	Hep_Hag	8
FTR1	03239	Iron permease FTR1 family	7
Peptidase_S6	02395	Immunoglobulin A1 protease	7
Pertactin	03212	Pertactin	6
IucA_IucC	04183	IucA / IucC family	5
Secretin	00263	Bacterial type II and III secretion system protein	5
Fimbrial	00419	Fimbrial protein	5
SIC	03482	sic protein	1
Fil_haemagg	05594	Haemagglutinin repeat	1
FAP	07174	Fibronectin-attachment protein (FAP)	1
Glug	07581	The GLUG motif	1
MCE	02470	mce related protein	1

## 7.2 Abkürzungsverzeichnis und Glossar

A	Adenosin
ABC	<i>ATP-binding cassette</i> - ATP-Bindekassette
Array	Zusammenstellung von Ziel-Sequenzen für ein Hybridisierungsexperiment auf einem Trägermaterial
ATP	Adenosintriphosphat
bp	Basenpaar(e)
Blast2n	Vergleich von Nukleinsäuresequenzen gegen Nukleinsäuresequenzen aus Datenbanken
bzw.	beziehungsweise
C	Cytidin
ca.	circa
COG	<i>Cluster of orthologous Groups</i>
DNA	Desoxyribonukleinsäure
<i>et al.</i>	<i>et alii</i> (und andere)
GO	<i>Gene Ontology</i> Datenbank
IS	Insertionssequenz
kb	Kilobasen (Längenmaß einzelsträngiger DNA oder RNA)
MFS	<i>major facilitator family</i>
<i>Microarray</i>	Array mit einer hohen Dichte an Zielsequenzen
Pfam	Proteinfamilien
Read	Sequenz eines DNA-Fragments ermittelt in einem Sequenzierlauf
Replikon	Replikationseinheit
RND	<i>Resistance-nodulation-cell division</i>
sp.	Spezies / Art (Einzahl)
spp.	Spezies / Arten (Mehrzahl)
u.a.	unter anderem
u.U.	unter Umständen
z.B.	zum Beispiel