GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

Veröffentlichung der Ergebnisse von Forschungsvorhaben im BMBF-Programm

# B I O L O G I E

**Forschungsvorhaben:**

Der Energie- und Biosynthesestoffwechsel von *Bacillus licheniformis*:
eine komparative Analyse der regulatorischen Netzwerke von
Produktions- und Wildtypstämmen

**Förderkennzeichen:** 0315387

**Zuwendungsempfänger:** Georg-August-Universität Göttingen,
37070 Göttingen

**Ausführende Stelle:** Georg-August-Universität Göttingen - Biologische Fakultät -
Institut für Mikrobiologie und Genetik, Grisebachstr. 8, 37077 Göttingen

**Projektleitung:** Dr. Heiko Liesegang

**Laufzeit**: 01.03.2009 bis 31.05.2015

"Das diesem Bericht zugrundeliegende BMBF-Forschungsvorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 0315387 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor".

# Schlussbericht

*(gemäß BNBest-BMBF 98 (Stand: April 2006) Nr. 3.2)*

| **Zuwendungsempfänger:** | **Förderkennzeichen:** |
|---|---|
| Georg-August-Universität Göttingen<br>Dr. Heiko Liesegang | 0315387 |
| **Vorhabenbezeichnung:**<br>Der Energie- und Biosynthesestoffwechsel von *Bacillus licheniformis*: eine komparative Analyse der regulatorischen Netzwerke von Produktions- und Wildtypstämmen | |
| **Laufzeit des Vorhabens:**<br>01.03.2009 - 31.05.2015 | |

## I. Kurzdarstellung

### 1. Aufgabenstellung

Die Aufgabenstellung des Projektes war die Identifikation und Analyse der regulatorischen Netzwerke des Biosynthese- und Energiestoffwechsels von *Bacillus licheniformis* Wildtyp und Produktionsstämmen. Zum Hauptziel der Förderinitiative, die Optimierung von biotechnologischen Wertschöpfungsketten, sollte dieses Projekt einen Beitrag zur Identifizierung von produktionsrelevanten Regulatoren zwecks Prozess- und Stammoptimierung leisten. Dabei sollte der Fokus auf der transkriptionalen und posttranskriptionalen Kontrolle des zentralen Metabolismus von *B. licheniformis* während einer produktiven Fermentation liegen. Hierzu sollten die sogenannten -omiks-Technologien eingesetzt werden. Als Ziele wurden dabei festgelegt:

a) Die Bestimmung der aktiven Transkriptionsstartpunkte
b) Die Identifizierung von 5'UTR kodierten Riboswitches
c) Die Identifizierung von kleinen nicht Protein kodierenden RNAs (ncRNAs)
d) Die Untersuchung von Promotor und Operatorregionen.

während der verschieden Fermentationszeitpunkte zu denen Proben genommen werden sollten. Auf Basis dieser Daten sollten die aktiven regulatorischen Netzwerke modelliert werden, um Regulatoren zu identifizieren, die Ansatzpunkte zur Prozess- bzw. Stammoptimierung liefern könnten.

### 2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Zu Beginn des Projektes waren die Genomsequenzen der zu untersuchenden Wildtyp- und Produktionsstämmen bekannt, tatsächlich wurden die dafür notwendigen wissenschaftlichen Arbeiten an der Universität Göttingen im Genomlabor unter Mitwirkung des Antragsstellers durchgeführt. Der Fermentationsprozess für die

Produktion von Proteasen sowie die dazugehörigen Produktionsstämme lagen im Besitz des Industriepartners Henkel. Die Wildtypstämme dagegen waren und sind öffentlich frei verfügbar. Voraussetzung und Basis dieses Universität/Industrie Kooperationsprojektes waren a) die an der Georg-August Universität aufgebaute Kompetenz bezüglich der –omiks-Technologien und der Bioinformatik und das im Industrieunternehmen Henkel etablierte Produktionsverfahren sowie die dort entwickelten und vorhandenen Produktionsstämme. Auf Basis dieser Situation sind bereits in den vorangehenden Jahren eine Reihe erfolgreicher Forschungskooperationen zwischen den beiden Partnern durchgeführt worden. In diesem Projekt wurde daher eine Vereinbarung getroffen, dass die Universität Göttingen das für das Projekte notwendige *know-how* über eigenes Personal bzw. durch wissenschaftliche Kooperationen mit Dritten sowie die Labor und die IT-Infrastruktur zur Verfügung stellt bzw. aufbaut, und die Firma Henkel das Produktions-verfahren sowie die Produktionsstämme einbringt. Es versteht sich, dass gesonderte Geheimhaltungsabkommen bezüglich des vertraulichen Umgangs mit den Stämmen und dem Verfahren aus dem Besitz der Firma geschlossen wurden.

## 3. Planung und Ablauf des Vorhabens

Zunächst wurden die experimentellen Voraussetzungen für die RNA-seq basierte Transkriptomanalyse für Stämme der Spezies *B. licheniformis* etabliert und getestet. Für die Durchführung des Projektes wurde ein Plan für die Etablierung der Probennahmestrategie, die Durchführung von industrienahen Fermentationen sowie die Probennahme durch einen Doktoranden aus dem Förderprojekt mit dem Industriepartner Henkel erstellt. Das Herzstück des Projektes, die Erzeugung und die Analyse von RNA-seq basierter Transkriptomik sowie die Entwicklung entsprechender bioinformatischer Analyseprogramme wurde für die Universität Göttingen geplant und erfolgreich etabliert, dafür wurden zwei der drei geförderten Doktoranden eingesetzt. Für die Ergänzung der Transkriptomdaten mittels Proteom- und Metabolomdaten wurden mit den universitären Partnern aus Greifswald Kooperationsverträge geschlossen, ein Zeitplan erstellt und erfolgreich durch einen der Doktoranden umgesetzt. Die Proteomanalyse wurde erfolgreich von einem der geförderten Doktoranden beim Partner in Greifswald durchgeführt. Eine Metabolomanalyse erwies sich unter den Bedingungen der industrienahen Fermentation aufgrund der Komplexität des eingesetzten Industriemediums als nicht durchführbar. Die integrative Analyse der Daten wurde für den Arbeitsgruppenleiter sowie die drei geförderten Doktoranden geplant und durchgeführt und führte zu mehreren Publikationen (siehe II.6). Für den Abschluss des Vorhabens sollte die experimentelle Bestätigung erbracht werden, dass auf Basis der identifizierten Regulatoren die Produktivität von Protease sekretierenden Stämmen erhöht werden kann. Zu diesem Zweck wurden durch den dritten geförderten

Doktoranden entsprechende molekulare Werkzeuge erstellt und publiziert. Abschließend wurde durch den Arbeitsgruppenleiter sowie den dritten geförderten Doktoranden gezeigt, dass in der Tat eine Produktivitätssteigerung durch Inaktivierung eines im Projekt identifizierten neuartigen RNA-Repressors möglich ist. Die Ergebnisse dieser Arbeiten werden derzeit durch den Industriepartner sowie die Patentabteilung der Universität Göttingen auf Schutzwürdigkeit überprüft und gegebenenfalls patentiert und anschließend publiziert.

4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Wissenschaftlich und technisch war es bei Beginn des Projektes möglich, RNA-seq Analysen basierend auf 454-Technologie für etablierte Labororganismen durchzuführen. Für Produktionsorganismen waren die Technologien und Verfahren nicht etabliert.

5. Zusammenarbeit mit anderen Stellen

Im Rahmen dieses Projektes wurde mit der Firma Henkel und der Arbeitsgruppe Hecker der Universität Greifswald experimentell zusammengearbeitet. Die Entwicklung der Methoden sowie die Analyse der Ergebnisse wurde durch Zusammenarbeit mit der Arbeitsgruppe Süß der Universität Frankfurt und der Arbeitsgruppe Liebl der TU München unterstützt. Die Analyse der Schutzwürdigkeit einer Erfindung wurde in Kooperation mit der Firma Henkel sowie der MBM-Sciencebridge (der zuständigen Institution der Universität Göttingen) durchgeführt.

## II. Eingehende Darstellung

1. Verwendung der Zuwendung und des erzielten Ergebnisses im Einzelnen, mit Gegenüberstellung der vorgegebenen Ziele

   Die Zuwendung wurde für die wissenschaftliche Arbeit wie geplant eingesetzt. Beschäftigt wurden drei Doktoranden, ein wissenschaftlicher Mitarbeiter, ein Arbeitsgruppenleiter sowie zwei studentische Hilfskräfte. Um die unter I.1. definierte Projektziele zu erreichen, wurden 8 Arbeitspakete definiert:

   1. Erstellung und Sequenzierung von cDNA-Bibliotheken
   2. Implementierung einer SQL-Datenbank für die Integration aller Daten
   3. *Set up* und Pflege der Daten
   4. Identifikation von Riboswitches, Promotor- und Operatorregionen
   5. Analyse von Riboswitches
   6. Metabolomanalysen von Regulationsmutanten
   7. Fermentationen

   Die geplanten Arbeitspakete sind in Tabelle 1: zeitliche Planung der Arbeitspakete im Projekt. dargestellt.

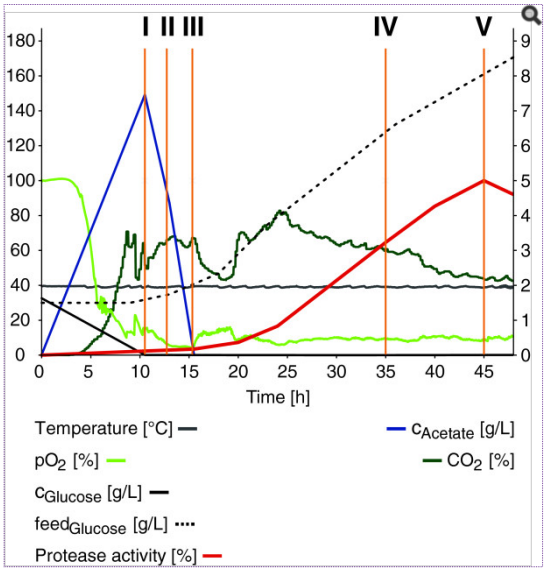
*Tabelle 1: zeitliche Planung der Arbeitspakete im Projekt.*

| Jahr 1 | | | Jahr 2 | | | | Jahr 3 | | | | Jahr 4 | | | | Jahr 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AP1 | | | | | | | | | | | | | | | | | | |
| AP2 | | | | | | | | | | | | | | | | | | |
| | AP3 | | | | | | | | | | | | | | | | | |
| | | | AP4 | | | | | | | | | | | | | | | |
| | | | | | AP5 | | | | | | | | | | | | | |
| | | | | | | AP6 | | | | | | | | | | | | |
| | | | | | | | | AP7 | | | | | | | | | | |
| | | | | | | | | | | AP8 | | | | | | | | |

**Im ersten Jahr** des Projektes sollte die RNA-seq-Methodik für *Bacillus licheniformis* anhand des Wildtypstamms DSM13 im Göttinger Genomlabor etabliert werden. Dazu wurde zunächst Frau Dipl. Ing. Sandra Wiegand als Doktorandin eingestellt. Da das eigentliche Industriemedium der Geheimhaltung unterliegt, wurde einem vorangegangenen BMBF-geförderten namens GenomikDesign ein synthetisches, dem Industriemedium ähnliches Medium, das sogenannte 20 Aminosäuremedium, entwickelt. In diesem Medium wurden Anzuchten im Schüttelkolben durchgeführt und zu geeigneten Zeitpunkten Proben für die Aufbereitung von RNA, Proteinen und

Metaboliten entnommen. Zusätzlich wurden geeignete Probennahmen und Lagermethoden für das biologische Material etabliert, um eine spätere Durchführung der verschiedenen Experimente zu ermöglichen. Die Aufbereitung der Proben wurde soweit etabliert, dass eine erste RNA-Sequenzierung durchgeführt werden konnte. Da im Laufe des ersten Jahres die auf Illumina-Technologie basierte NGS-Sequenzierung mit erheblich größerem *output* pro Experiment für RNA-Sequenzierung verfügbar wurde, wurde entgegen der Planung aus dem Antrag entschieden, statt der 454-Technologie die Illumina-Technologie einzusetzen. Ein halbes Jahr versetzt zum Beginn der experimentellen Arbeiten wurde der Doktorand Herr Dipl. Biol. Sascha Dietrich für die bioinformatischen Analysen eingestellt. Im Zuge seiner Arbeit wurde zunächst ein geeigneter Server für die Analysen der Daten angeschafft und eingerichtet. Nach Beschaffung, Installation und Inbetriebnahme des Servers begann Herr Dietrich mit dem Design und der Entwicklung einer integrativen Datenbank zur Aufnahme und Analyse von RNA-seq Daten im genomischen Kontext. Eine erste Version der Datenbank wurde zeitgleich mit der Erstellung der ersten Test-Datensätze aus RNA-seq Experimenten von Frau Wiegend in Betrieb genommen. Somit wurde die Meilensteinplanung aus Tabelle 1 eingehalten.

**Im zweiten Jahr** des Projektes wurde entschieden, die ursprünglich im Göttinger Genomlabor geplanten Fermentationsexperimente beim Kooperationspartner Henkel durchzuführen, da somit sehr viel näher am eigentlichen Produktionsprozess gearbeitet werden konnte. Zusätzlich vereinfachte dies den Umgang mit Material, für das besondere Geheimhaltungsbedingungen erforderlich waren. Somit konnte ein Produktionsplasmid der Firma Henkel eingesetzt werden. Frau Wiegand hat daher bei der Firma Henkel jeweils drei parallele industrienahe Fermentationen im originalen Industriemedium mit einem zur Publikation geeigneten Derivat des Wildtypstammes DSM13 sowie einem Produktionsstamm der Firma Henkel im Fermentationslabor des Kooperationspartner Henkel durchgeführt. Alle Proben wurden zur weiteren Bearbeitung in das Göttinger Genomlabor verbracht. Auf Basis des im ersten Jahr entwickelten Verfahrens wurde aus den Proben RNA isoliert und für jeweils 20 *deep sequencing*-Experimente verwendet. Insgesamt wurden fünf Fermentationszeitpunkte beprobt, wobei für jeden Zeitpunkt drei Replikate in RNA-seq Experimenten und ein zusätzliches dRNA-seq-Experiment explizit zur Bestimmung von Transkriptionsstartpunkten durchgeführt wurden. Die Probenzeitpunkte wurden anhand der Prozessdaten der Fermentation festgelegt (Abbildung 1).




**Protease production and process parameters.** Process parameters are shown for fermentation L (the parameters for the replicate fermentations R and M are corresponding, data not shown). Temperature T [°C], oxygen partial pressure $pO_2$ [%], glucose concentration $c_{Glucose}$ [g/L], supplied glucose $feed_{Glucose}$ [g/L] and normalized protease activity [%] are displayed on left y-axis, whereas acetate concentration $c_{Acetate}$ [g/L] and carbon dioxide content $CO_2$ [%] are scaled on the right y-axis. Process time t [h] is given on the x-axis. The sampling points I to V are indicated by orange lines.

*Abbildung 1: Probenzeitpunkte für RNA-seq Experimente.*

Im Gegensatz zu klassischen mikrobiologischen Laborversuchen mit klaren Medien kann im trüben Industriemedium die sonst üblicherweise verwendete optische Dichte nicht für die Verfolgung des Wachstums herangezogen werden. Entscheidend für die Festlegung der Probenzeitpunkte waren daher die Fermentationsparameter Acetat-Produktion, Glucose-feed und Proteaseaktivität. Basis war die Überlegung, solche Zeitpunkte zu wählen, an denen eine möglichst große Zahl von unterschiedlichen, produktionsrelevanten, physiologischen Aktivitäten zu erwarten ist. Gewählt wurde daher ein Probenzeitpunkt am Beginn des Wachstums, zum Maximum des sogenannten Acetat-overflow Metabolismus, bei Beginn der Proteaseproduktion, in der exponentiellen Phase der Proteaseproduktion und am Maximumpunkt der Produktion, welcher gleichzeitig das Ende der Fermentation darstellt. Die für den Stamm *B. licheniformis* DSM13 spoD4 erreichten RNA-seq Ergebnisse sind in Tabelle 2 dargestellt.

*Tabelle 2: Ergebnisse der RNA-seq Experimente auf B. licheniformis DSM13 spoΔ4.*

| Sample | ∑ reads | reads against tRNA or rRNA genes | multihit reads | unmapped reads | plasmid-mapping reads | chromosome-mapping reads |
|---|---|---|---|---|---|---|
| L-I | 34.553.369 | 28.969.888 | 32.013 | 3.424.018 | 174.110 | 1.953.340 |
| R-I | 34.491.633 | 30.642.606 | 13.956 | 2.148.822 | 161.458 | 1.524.791 |
| M-I | 23.583.582 | 17.140.657 | 138.359 | 3.157.703 | 309.415 | 2.837.448 |
| | | | | | | |
| L-II | 34.935.270 | 28.636.703 | 60.336 | 2.963.179 | 189.401 | 3.085.651 |
| R-II | 36.024.461 | 29.680.210 | 98.178 | 2.203.136 | 289.793 | 3.753.144 |
| M-II | 28.018.016 | 20.903.311 | 81.313 | 3.689.827 | 235.161 | 3.108.404 |
| | | | | | | |
| L-III | 34.152.665 | 27.024.138 | 32.914 | 4.084.170 | 154.977 | 2.856.466 |
| R-III | 38.187.230 | 30.719.713 | 81.366 | 2.746.909 | 183.528 | 4.455.714 |
| M-III | 28.174.975 | 21.615.689 | 49.622 | 3.099.499 | 149.313 | 3.260.852 |
| | | | | | | |
| L-IV | 38.200.610 | 28.824.001 | 22.027 | 5.878.876 | 579.663 | 2.896.043 |
| R-IV | 42.959.023 | 35.377.857 | 150.077 | 4.734.060 | 327.074 | 2.369.955 |
| M-IV | 36.414.040 | 27.221.034 | 23.275 | 6.205.296 | 523.290 | 2.441.145 |
| | | | | | | |
| L-V | 35.656.360 | 27.619.551 | 30.045 | 4.611.200 | 578.175 | 2.817.389 |
| R-V | 31.798.624 | 26.061.998 | 280.073 | 3.075.680 | 224.030 | 2.156.843 |
| M-V | 34.376.322 | 23.540.363 | 28.510 | 7.807.537 | 362.828 | 2.637.084 |

Aufgrund der leistungsfähigen Illumina-Technologie wurden pro Sequenzierung zwischen ~23 Mio. Illumina reads und ~43 Mio. Illumina reads produziert. Damit wurde für die eingesetzten Investitionsmittel eine erheblich höhere Sequenziertiefe erreicht als die im Antrag angestrebte Tiefe von ca. 140.000 454-reads pro Probe. Die Ergebnisse für den Produktionsstamm waren entsprechend. Auf Basis der RNA-seq Ergebnisse wurde, wie in den Arbeitspaketen vorgesehen, eine Transkriptionsstartbestimmung und zusätzlich eine Genstartkorrektur durchgeführt. Auf Basis dieser Arbeiten, aber auch durch Einbeziehung der seit der Erstveröffentlichung zugänglichen neuen Daten, wurde eine komplette Reannotation des *B. licheniformis* Typstammes DSM13 durchgeführt und der Allgemeinheit durch Submission des Datensatzes bei GenBank unter dem *accesskey* GCA_000008425.1 zur Verfügung gestellt.

Schließlich wurde der dritte Doktorand, Herr Dipl Biol. Robert Hertel eingestellt, der a) an der Reannotation des Typstammgenoms mitarbeitete und b) umgehend mit der Entwicklung von Reporterplasmiden für die Untersuchung von Riboswitches und Promotoren/Operator-Fusionen in *B. licheniformis* begann. Bei der Reannotation wurde jedem der drei Doktoranden ein Drittel des Genoms zugeordnet, wobei Herr Hertel sich schwerpunktmäßig auf sogenannte Prophagenregionen konzentrierte, da diese laut Hinweis des Kooperationspartners Henkel ein potentielles Produktionshindernis

darstellen. In diesem Fall besteht ein möglicher Nutzen nicht in der Steigerung der Produktion, sondern in der möglichen Stabilisierung des Produktionsprozess unter Stressbedingungen, die zur Induktion solcher Prophagen führen können und laut den Erfahrungen unseres Industriepartners in der Vergangenheit in der Tat für Produktionseinbußen verantwortlich waren.

**Im dritten Jahr** des Projektes wurden korrespondierend zu den RNA-seq Experimenten mit den Proben beim Kooperationspartner Greifswald durch Frau Wiegand die Arbeiten zu den geplanten Proteomanalysen durchgeführt. Parallel dazu begann Herr Dietrich mit den Arbeiten zum effizienten Umgang mit multiplen RNA-seq Datensätzen, sowie der Entwicklung eines interaktiven Interfaces für die Visualisierung und Analyse der im zweiten Jahr erstellten Datenbank. Zum Zeitpunkt dieser Arbeiten war es mit den verfügbaren Servern und Analysetools möglich, maximal einen Datensatz für eine RNA-seq Analyse gleichzeitig zu verarbeiten. Der Hauptgrund für diese Beschränkung ergab sich aus dem mit der Zahl der reads steigenden Speicherbedarf. Herr Dietrich entwickelte und implementierte ein Datenmodell zur Verarbeitung von RNA-seq Daten, das linear mit der Größe des Genoms und unabhängig von der Zahl der gemappten reads skaliert. Basierend auf diesem Datenmodell wurde die Entwicklung des Programms TraV gestartet. Dies ermöglichte die simultane Bearbeitung aller zwanzig RNA-seq Datensätze pro Stamm.

Zur Erreichung der Ziele des Projektes sollten Transkriptionsstartpunkte oder *transcription start sites* (TSS), Riboschalter und ncRNAs identifiziert, sowie eine vergleichende Aktivitätsbestimmung von metabolischen und biosynthetischen Operons durchgeführt werden. Herr Dietrich entwickelte dazu vier entsprechende Analysetools und integrierte diese in TraV. Bei der ersten Sichtung der RNA-seq Daten im genomischen Kontext wurde eine überraschend hohe Anzahl von transkriptionalen *anti-sense* Aktivitäten gefunden. Darauf reagierend hat Herr Dietrich ein fünftes Analysetool entwickelt und ebenfalls in TraV integriert.

Es erfolgte dann eine vergleichende Evaluierung und Analyse aller RNA-seq Daten durch die drei Doktoranden und die Arbeitsgruppenleitung, die zu diesem Zeitpunkt aus Frau Dr. Sonja Volland und dem Antragsteller bestand. Basierend auf den Ergebnissen der Sichtung entwickelte Frau Wiegand ein Klassifizierungsschema aus 9 Transkriptionsmustern (Abbildung 2 A). Ein hervorstechendes Ergebnis der quantitativen Analyse ist, wie bereits erwähnt, die unerwartet hohe Zahl an *anti-sense* Aktivität. Hier ist auffallend, dass mit fast 90% der Großteil der gefundenen freien oder ncRNA-Muster in die *anti-sense* Klassen fallen (Abbildung 2 B). Im Gegensatz dazu ist der Großteil der in UTR-Regionen lokalisierten transkriptionalen Aktivität nicht *anti-sense* zu anderen genomischen Elementen. Die Evaluierung der quantitativen

transkriptionalen Aktivität des Genoms ergab, dass insgesamt mindestens 87% des Genoms signifikant aktiv ist (Abbildung 2 C).

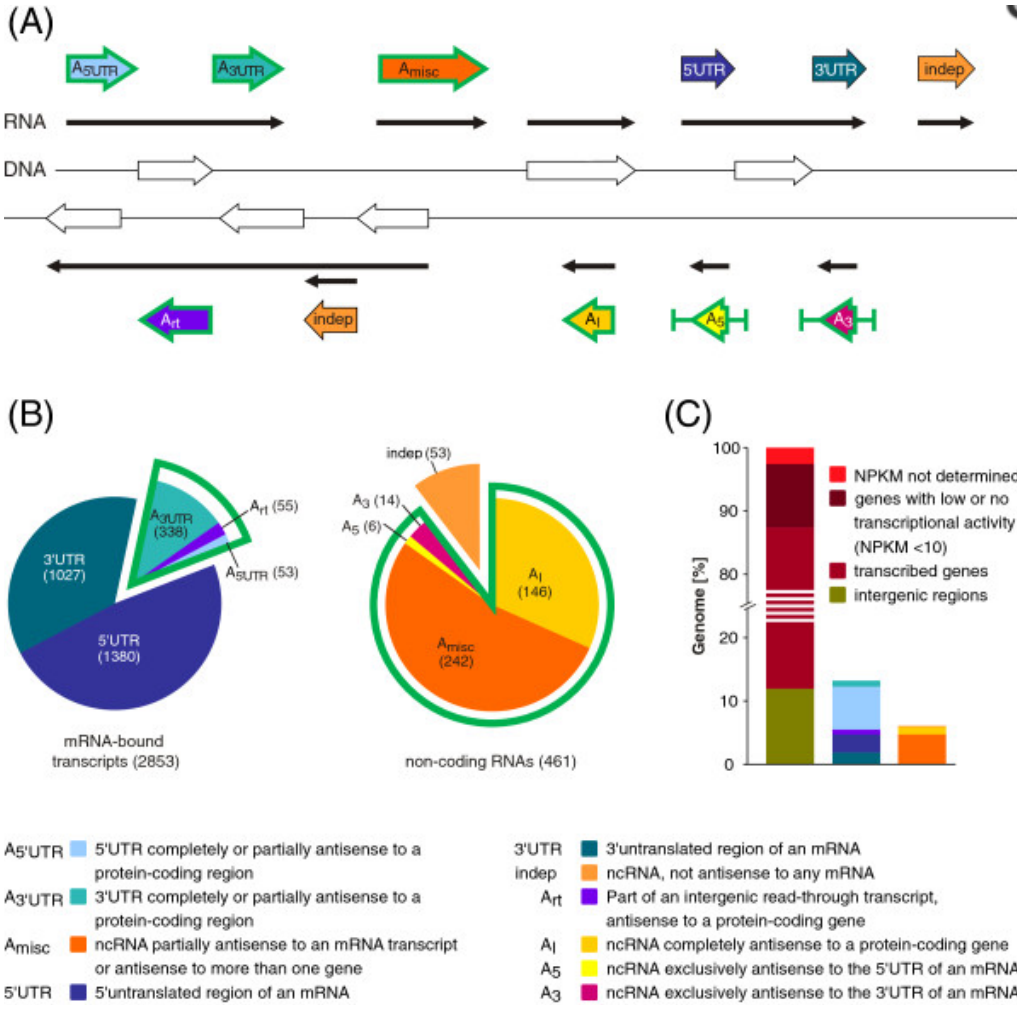



*Abbildung 2: Klassen transkriptionaler Aktivität in B. licheniformis.*

Als nächstes stellte sich die Frage, ob es bestimmte Bereiche im Genom von *B. licheniformis* gibt, die in besonderem Maße während der Fermentation aktiv bzw. inaktiv sind. Hierzu erfolgte eine regionale Zuordnung der transkriptionalen Aktivität zur aktuellen Genomannotation, die in Abbildung 3 dargestellt ist. Diese regionale Analyse ergab, dass unter den beobachteten Bedingungen acht Regionen bemerkenswert inaktiv waren. Interessanterweise ergab der Abgleich mit der Genomannotation, dass alle annotierten Prophagenregionen sowie ein sekundärmetoblischer Gencluster dazu zählten. Somit konnte gezeigt werden, dass die Gene der Prophagen keinen Beitrag zur Produktivität leisten und unterstreichen damit, dass diese Regionen gute Deletionsziele für eine Genomreduktion im Sinne des rationalen Stammdesigns darstellen.

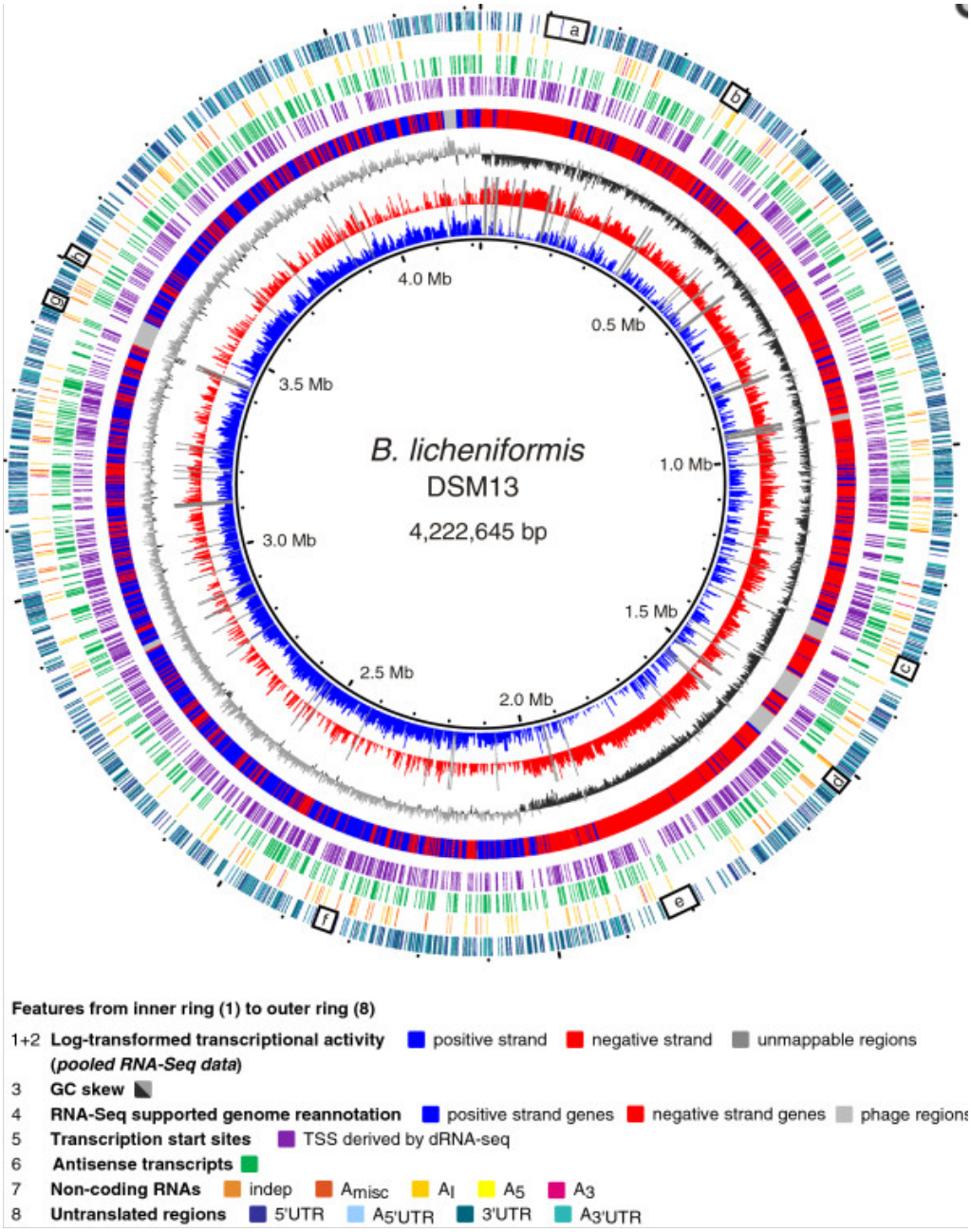



*Abbildung 3: Trasnkriptionale Aktivitäten des B. lichenifomris Genoms während ener produktiven Fermentation.*

**Im vierten Jahr** des Projektes wurden die RNA-seq Daten aus dem Jahr drei zusammengefasst und als Basis der ersten Publikation verwendet (1). Als Autoren war hier das gesamte aus der Förderung finanzierte Team sowie die Kooperationspartner aus der Firma Henkel beteiligt.

Im nächsten Schritt erfolgte ein Abgleich der Transkriptomdaten mit den Proteomdaten in den verschiedenen Fermentationsphasen. Die integrative Analyse der Daten zeigte unter anderem die Reihenfolge der verwendeten Substrate auf. Da der Produktionsprozess die Reaktion des Organismus auf die Substratversorgung ausnutzt, schließlich ist eine Protease eines der zentralen Fütterenzyme der saprophytisch lebenden Spezies, könnten sich hier Ansätze für eine Verbesserung der

Prozessführung ergeben. Die Beobachtung, dass zur Hauptproduktionsphase der Fermentation die stärkste transkriptionale Aktivität eines Proteingens vom Lichenisin-Gen erfolgt, zeigt, dass der Organismus hier auf Zelldichtestress reagiert. Dies ist eine Reaktion, die im natürlichen Habitat der Spezies beim Wettbewerb um Ressourcen mit Nahrungskonkurrenten sinnvoll ist. Während einer Industriefermentation ist dies natürlich kontraproduktiv. Hier ergibt sich ein weiterer Ansatz zur Stammverbesserung. Insgesamt ergaben die Untersuchungen eine Reihe von physiologischen und auch regulatorischen Phänomenen, die als Vorschläge zur Prozess- und Stammoptimierung sinnvoll erscheinen. Alle für die Energieversorgung und die Biosynthese von Aminosäuren wichtigen aktiven Stoffwechselwege wurden in ein allgemeines Stoffwechselmodell integriert. Hierbei zeigte sich, wie wichtig die Integration von Proteom- und RNA-seq Daten für das Verständnis der Physiologie ist. Aufgrund der unterschiedlichen Halbwertzeiten von RNA- und Protein-Molekülen wären einige Stoffwechselwege fälschlicherweise als inaktiv betrachtet worden, wenn man nur die RNA-Daten zur Verfügung gehabt hätte.

Alle hier erwähnten Daten und Vorschläge zur Stammverbesserung wurden in Wiegand et al. 2013b (2) gemeinsam mit den Kooperationspartnern aus der Universität Greifswald und der Firma Henkel publiziert. Frau Wiegand wendete sich dann den geplanten Arbeiten zur Metabolomanalyse zu. Hierbei zeigte sich allerdings, dass sich die Zellen aus den Proben wegen des komplexen Industriemediums nicht vom Medium trennen ließen. Damit wurde klar, dass die Metabolomexperimente in der geplanten Form prinzipiell unmöglich sind. Konsequenterweise mussten die Arbeiten eingestellt werden. In den vorbereitenden Experimenten mit dem komplexen 20 Aminosäuremedium stellte dies kein Problem dar. Frau Wiegand beendet erfolgreich Ihre Dissertation und schied aus dem Projektteam turnusgemäß aus.

**Im fünften Jahr** des Projektes hatte Herr Hertel die Entwicklung der Reporterplasmide abgeschlossen und selbige mit im dritten und vierten Jahr identifizierten Regulatoren getestet und evaluiert. Gemeinsam mit der Projektleitung wurde die entsprechende Publikation verfasst und veröffentlicht (3). Herr Dietrich hat parallel dazu das RNA-seq Visualisierungstool TraV auf Basis der gemeinsam durchgeführten Analysen aus dem vierten Projektjahr optimiert und finalisiert. Hier erfolgte in Zusammenarbeit mit Frau Wiegand und dem Antragsteller die Erstellung eines Manuskriptes, das ebenfalls unter Dietrich et. al 2014 publiziert wurde (4). Herr Dietrich beendet auf Basis dieser Ergebnisse und Publikationen erfolgreich seine Dissertation.

Aus den RNA-seq Daten ging hervor, dass die Prophagenregionen in dem von uns untersuchten Stamm nicht vollständig inaktiv waren. Eine *en detail*-Analyse der Phagenregionen zeigte, dass es sehr wohl einzelne aktive Gene gab. Hierbei handelte es sich sogar um relativ starke Aktivitäten der Phagenrepressorgene. Anscheinend

waren die Prophagen in einer Art „*standby*“ Modus reguliert und könnten demnach aktiv werden, wenn die Repressoren ihre Arbeit einstellen. Eine solche Aktivität hätte eine Lyse der Zellen zur Folge und wäre demnach mit einer massiven Reduktion der Produktivität verbunden. Ein solches Szenario würde die Beobachtung unserer industriellen Kooperationspartner erklären, die Prophagen als ein potentielles Produktivitätsproblem gefunden haben. Da dem Hauptziel der Förderinitiative, nämlich der Verbesserung der Produktivität von biotechnologischen Prozessen, hier gedient werden kann, wurde die Untersuchung der Regulation und der Aktivität der Prophagenregion als Nebenziel festgelegt. Herr Hertel hat hier, unter anderem mit Unterstützung der Masterstudentin Stefania Neuber, eine NGS-Methode zur Bestimmung der Aktivität von Prophagenregionen in Produktions- und Wildtypstämmen entwickelt, welche um den Faktor hundert sensitiver ist als alle bisher bekannten Methoden zur Bestimmung von Phagenregionen. Eine wichtige Erkenntnis war hierbei, dass von den sieben im Genom des Wildtypstammes annotierten Prophagenregionen in der Tat drei in der Lage sind, Phagenpartikel zu produzieren. Somit wurde gezeigt, dass diese Prophagen in der Lage sind, einen Effekt wie den von unseren Industriepartnern beobachteten Schaden innerhalb einer Fermentation zu erzeugen. Dieses Nebenergebnis wurde in Hertel et al. 2014 publiziert (5).

Ziel der Identifizierung neuartiger RNA-basierter Regulatoren und ihre Evaluation im Stoffwechselmodel war letztlich die Stammoptimierung. Herr Hertel begann die letzte Phase seiner Doktorarbeit mit der Evaluierung der in 2013 publizierten Regulatoren aus dem Blickwinkel möglicher Ansatzpunkte für eine Stammoptimierung. Als vielversprechendster Kandidat ergab sich dabei eine *anti-sense* Aktivität zur 3’Region des Gens *apr* der Protease Carlsberg, welche im weiteren AprA für „apr anti-sense Aktivität“ genannt wird. Daher wurde das *aprA*-Gen als Gen zum „*proof of principle*“ des Projektes ausgewählt. Herr Hertel erzeugte hierzu eine *knock down* Mutante, indem er den Promotor der aktiven Region durch eine alternative Sequenz ersetzte. Für die weiteren Untersuchungen wurden studentische Hilfskräfte als Unterstützung für Herrn Hertel eingesetzt. Bei der Charakterisierung der Mutante zeigte sich zunächst auf Plattentests und darauf folgend auch in flüssigem Medium eine erhöhte Proteaseaktivität. Dies war ein erster Hinweis darauf, dass in der Tat auf Basis unserer neuartigen Regulatoren Stammverbesserungen möglich sind. Somit war das Hauptziel des Projektes bestätigt.

Die Kooperationsvereinbarung mit dem Industriepartner besagt unter anderem, dass Ergebnisse, die potentiell interessant für Schutzrechte, sprich Patente, sind, vor einer möglichen Veröffentlichung mit dem Industriepartner evaluiert werden. Trotz der zwischenzeitlichen Einstellung der Proteaseproduktion des Projektpartners Henkel mit dem Stamm und der Lizensierung des Verfahrens an die Firma BASF, wurde

grundsätzliches Interesse gezeigt. Daraufhin wurde vereinbart, dass entsprechende Untersuchungen zur Patentierbarkeit in Göttingen an den Wildtypstämmen und in Düsseldorf bei der Firma Henkel an den Produktionsstämmen durchgeführt werden sollten. Im weiteren wurden durch Robert Hertel sowie durch die von ihm angeleitete Masterstudentin Sandra Meyerjürgens Mutanten erzeugt und in Bezug auf ihre Produktivitätssteigerung, sowie den konkreten Mechanismus der möglichen Erfindung hin untersucht.

**In der kostenneutralen Verlängerung** des Projektes wurden die Experimente zur Patentwürdigkeit der Manipulation des *aprA*-Gens erfolgreich abgeschlossen. Es konnte eindeutig gezeigt werden, dass die Produktivität der in Göttingen untersuchten Wildtypstämme um ca. 100% in Bezug auf die Protease gesteigert werden kann. Zugleich ist dies der erste Fall, in dem ein RNA-Repressor für ein Hauptprodukt eines Bakterienstammes entdeckt wurde. Damit stellt das Phänomen sowohl in wirtschaftlicher als auch in wissenschaftlicher Hinsicht eine wichtige neue Erkenntnis dar. Leider konnte der Industriepartner den Effekt auf den Produktionsstämmen nicht reproduzieren. Möglicherweise haben wir hier einen der bisher unbekannten Faktoren identifiziert, der einen Wildtyp- von einem Produktionsstamm unterscheidet. Vielleicht ein lohnendes Ziel bei der Konvertierung neuer Wildtypisolate als Grundlage für die Konvertierung zu Produktionsstämmen. Letztlich wurde nach mehrmonatigen Verhandlungen zwischen der MBSR-Sciencebridge (für Patentfragen der Universität Göttingen zuständig) mit der für Patente zuständigen Person aus der Rechtsabteilung Firma Henkel und dem Lizenznehmer BASF von Henkel und BASF deutlich gemacht, dass kein Interesse mehr an einer Patentierung besteht. Derzeit wird an der Uni-Göttingen an einer baldigen Veröffentlichung der Untersuchungen gearbeitet.

**Fazit**

Generell haben die Mitglieder der geförderten Juniorgruppe alle geplanten Arbeiten (mit Ausnahme der Metabolomanalyse) durchgeführt und alle Ziele erreicht. Auf Basis der erzielten Ergebnisse existiert nun ein umfassendes Modell der Stoffwechselprozesse und der damit verbundenen RNA-basierten regulatorischen Netzwerke von *B. licheniformis*-Stämmen während einer industrienahen produktiven Fermentation. Auf Basis dieses Modells wurde eine Reihe bis dato unbekannter neuartiger Regulatoren und auch Regulationsprinzipien identifiziert. Darüber hinaus konnte exemplarisch gezeigt werden, dass die Manipulation eines solchen Regulators in der Tat die Effizienz eines produktiven Stammes verdoppeln kann.

Alle im Rahmen dieses Projekts beschäftigten Doktoranden wurden promoviert. Insgesamt wurden drei erfolgreiche Doktorarbeiten sowie eine Masterarbeit und eine Bachelorarbeit erstellt.

Beim wissenschaftlichen Equipment wurde der beantragte Server sowie die Computer für die grafische Analyse der Ergebnisse wie geplant erworben. Die Maschinen wurden für die Durchführung der geplanten Analysen, die Entwicklung der notwendigen Analysetools sowie die Erstellung der Publikationen eingesetzt. Im Unterschied zur ursprünglichen Planung wurde der beantragte Fermenter nicht angeschafft. Die ursprünglich damit geplanten Ergebnisse wurden dennoch erzielt, da der Industriepartner Henkel die nötige Fermenterkapazität zur Verfügung gestellt hat. Für die ursprünglich für den Fermenter bewilligten Mittel wurde beim Projektträger ein Mittelumwidmungsantrag gestellt und genehmigt, und ein Server für die dauerhafte Nutzung der im Projekt erstellten öffentlichen Daten und Tools durch die Projektpartner sowie die wissenschaftliche Gemeinschaft angeschafft und implementiert.

## 2. Wichtigste Positionen des zahlenmäßigen Nachweises

a) Personalmittel

- ein Projektleiter für die Anleitung aller Arbeiten und die finale Auswertung über die Projektlaufzeit
- ein Postdoc für die Organisation der Datenanalysen und als Unterstützung für die Publikationen für letzten zwei Jahre
- ein Doktorand, der die Probennahmen, die RNA-seq Durchführung sowie die Proteomanalysen etabliert hat
- in Doktorand für die Integration sämtlicher Daten, der eine bioinformatische Analyseplattform entworfen, programmiert und implementiert hat
- ein Doktorand, der molekularbiologische Tools zur Untersuchung neuartiger Regulatoren entwickelt und exemplarisch an neugefunden Regulatoren evaluiert hat
- zwei studentische Hilfskräfte, die bei der experimentellen Überprüfung von Regulationssystemen und Produktionshemmnissen geholfen haben

b) Investitionsmittel

Zwei Server für (i) die Analyse und Integration der gewonnenen RNA-seq- und Proteomdaten in den genomischen Kontext und (ii) eine dauerhafte Nutzung der Daten und entwickelten Tools.

c) Auftragsmittel

- 20 Illumina basierte *deep sequencing* Experimente
- 15 Proteomanalysen

3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die im Projekt geleistete Arbeit diente dem Erreichen der im Projektantrag aufgeführten Ziele, wie in diesem Bericht zusammenfassend dargestellt wurde. Da alle Ziele (bis auf die Metabolomanalyse) im Wesentlichen erreicht wurden, waren die durchgeführten Arbeiten sowohl notwendig als auch dem Umfang nach angemessen.

4. Voraussichtlicher Nutzen, insbesondere der Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans

Wie im Projekt gezeigt wurde lassen sich auf Basis der gewonnenen Erkenntnisse Produktionssteigerungen erzielen. Die Vielzahl der erstmals beschriebenen neuartigen Regulatoren verspricht eine Fülle von neuen Ansatzpunkten für das Verständnis der Stoffwechseldynamik unter produktiven Bedingungen. Zugleich können die Daten auch im wissenschaftlichen Sinne neue Ansatzpunkte zum Verständnis der Physiologie sowohl der Spezies *Bacillus licheniformis* als der Gattung *Bacillus* insgesamt bieten. In unserer Arbeitsgruppe wird auf jeden Fall eine andauernde Verwertung der Ergebnisse in Form von Ideen für neue Forschungsvorhaben stattfinden.

Da die im Projekt erstellten Veröffentlichungen den Status „highly accessed“ erreicht haben, ist davon auszugehen, dass andere Wissenschaftler und Arbeitsgruppen von diesen Möglichkeiten ebenfalls Gebrauch machen. Insbesondere das integrierte Stoffwechselmodel bietet bisher unzugängliche Daten für dynamische Substratverwertungsketten, die neben der Stammoptimierung auch Ansatzpunkte für eine Prozessoptimierung erlauben. Zudem ermöglicht eine der beobachteten Stressreaktionen Einblick in kompetitive Fähigkeiten der Wildtypstämme. Damit ergeben sich Ansatzpunkte für eine erweiterte Betrachtung der ökologischen Funktion von *Bacillus licheniformis* an seinem natürlichen Standort als Saprophyt. Die veröffentlichten Ergebnisse stellen somit eine exzellente Grundlage für weiteres rationales Stammdesign dar.

Darüber hinaus werden die entwickelten und publizierten Tools, sowohl das Programm TraV als auch die Reportersysteme, bereits von anderen Gruppen in DFG- und ERA-Net-Projekten sowie BMBF geförderten Kooperationsprojekten eingesetzt. Somit ist bereits gezeigt, dass die Ergebnisse dieses Projektes im Sinne des fortgeschriebenen Verwertungsplanes eingesetzt werden.

5. Während der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Projektlaufzeit wurden keine wesentlichen Fortschritte auf der Thematik an anderen Stellen veröffentlicht.

## 6. Erfolgte Veröffentlichungen des Ergebnisses


1. Rachinger M, Volland S, Meinhardt F, Daniel R, Liesegang H. (2013) First insights into the completely annotated genome sequence of *Bacillus licheniformis* strain 9945A. Genome Announc 1:e00525-13.
2. Rachinger M, Bauch M, Strittmatter A, Bongaerts J, Evers S, Maurer KH, Daniel R, Liebl W, Liesegang H, Ehrenreich A (2013) Size unlimited markerless deletions by a transconjugative plasmid-system in *Bacillus licheniformis*. J Biotechnol 167:365–369
3. Wiegand S, Dietrich S, Hertel R, Bongaerts J, Evers S, Volland S, Daniel R, Liesegang H (2013) RNA-Seq of *Bacillus licheniformis*: active regulatory RNA features expressed within a productive fermentation. BMC Genomics 14:667.
4. Wiegand S, Voigt B, Albrecht D, Bongaerts J, Evers S, Hecker M, Daniel R, Liesegang H (2013) Fermentation stage-dependent adaptations of *Bacillus licheniformis* during enzyme production. Microb Cell Fact 12:120.
5. Wiegand S, Rabausch U, Chow J, Daniel R, Streit WR, Liesegang H (2013) Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. Genome Announc 1:e00092-13.
6. Hertel R, Volland S, Liesegang H. (2015) Conjugative reporter system for the use in *Bacillus licheniformis* and closely related Bacilli. Lett Appl Microbiol. 60:162-167.
7. Dietrich S, Wiegand S, Liesegang H (2014) TraV: a genome context sensitive transcriptome browser. PLoS One 9:e93677.
8. Handtke S, Volland S, Methling K, Albrecht D, Becher D, Nehls J, Bongaerts J, Maurer K-H, Lalk M, Liesegang H, Voigt B, Daniel R, Hecker H (2014) Cell physiology of the biotechnological relevant bacterium *Bacillus pumilus* – an omics-based approach. J Biotechnol 192:204-214.
9. Jakobs M, Hoffmann K, Grabke A, Neuber S, Liesegang H, Volland S, Meinhardt F (2014) Unravelling the genetic basis for competence development of auxotrophic *Bacillus licheniformis* 9945A strains. Microbiology 160:2136-2147.
10. Hertel R, Rodríguez DP, Hollensteiner J, Dietrich S, Leimbach A, Hoppert M, Liesegang H, Volland S (2015) Genome-based identification of active prophage regions by Next Generation Sequencing in *Bacillus licheniformis* DSM13. PLOS One 10:e0120759.
11. Jakobs M, Hoffmann K, Liesegang H, Volland S, Meinhardt F (2015) The two putative *comS* homologs of the biotechnologically important *Bacillus licheniformis* do not contribute to competence development. Appl Microbiol Biotechnol 99:2255-2266.


Eine Publikation zum Effekt des *aprA*-Gens ist in Vorbereitung. Darüber hinaus gibt es zwei weitere wissenschaftliche Fragestellungen, die auf die Untersuchung weiterer identifizierter neuartiger Regulatoren (Riboswitch, ncRNA-Stressregulator) und deren Veröffentlichung abzielen.