AGES

BUNDESMINISTERIUM FÜR GESUNDHEIT

Mosaic Genes

Impact of Mosaic Genes on the Risk Assessment of GMOs

Impressum

Publisher, Mediaowner
Bundesministerium für Gesundheit,
Radetzkystraße 2, 1031 Wien

Team
Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH (AGES)
Spargelfeldstraße 191, 1220 Wien

Project Management
Johann Steinwider[2)]

Scientific Concept and Coordination
Markus Wögerbauer[2)]

Authors
Sara Domingues[1)]
Klemens Fuchs[2)]
Ian Kopacka[2)]
Melanie Kuffner[2)]
Johann Steinwider[2)]
Markus Wögerbauer[2)]

Scientific Advisor
Kaare M. Nielsen[3)]

Suggested Citation
Woegerbauer M., Kuffner M., Kopacka I., Domingues S., Steinwider J., Nielsen K. M., and K. Fuchs. Impact of mosaic genes on the risk assessment of GMOs. Federal Ministry of Health, 2015. pp. 1 – 268; Vienna, Austria.

Affiliations
1) Department of Microbiology, Health Sciences Campus, Faculty of Pharmacy
University of Coimbra
Azinhaga de Santa Comba
3000-548 Coimbra, Portugal

2) Integrative Risk Assessment – Data – Statistics
Austrian Agency for Health and Food Safety - Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH (AGES)
1220 Wien, Spargelfeldstrasse 191

3) Department of Pharmacy, University of Tromsö, 9037 Tromsö, Norway

Release Date
March 2015

ISBN 978-3-903099-00-5

Content

List of Figures:

List of Tables:

List of Abbreviations:

APH	aminoglycoside phosphotransferase (enzyme)
ARM	antibiotic resistance marker
ATP	adenosine triphosphate
BD413	*Acinetobacter baylyi*, strain designation
BLAST	basic local alignment tool
bp	base pair
CDC	Center for Disease Control
CIN	competene induction box (genetic element)
com	competence (gene, locus or regulon)
CP4	*Agrobacterium tumefaciens,* strain designation
CRP	cAMP receptor protein
CRY	crystal toxins, *Bacillus thuringiensis*
CSP	competence stimulating peptide
DNA	deoxyribonucleic acid
EFSA	European Food Safety Authority
EPSPS	5-enolpyruvylshikimate-3-phosphate synthase
ET	electrophoretic type defined by multilocus enzyme electrophoresis
GARD	genetic algorithm for recombination detection
GC	guanosin-cytosin
GIT	gastrointestinal tract
GM	genetically modified
GMO	genetically modified organism
GVO	gentechnisch veränderter Organismus
HFIR	homology-facilitated illegitimate recombination
HGT	horizontal gene transfer
ISI	Institute for Scientific Information
LHS	left hand side
Log	logarithmic
MSA	multiple sequence alignment
NCBI	National Center for Biotechnology Information
nptII	neomycin phosphotransferase II (= *aph(3')-IIa*), gene
NJ	neighbor joining
NS	non-significant
nt	nucleotide
ORF	open reading frame
PAT	phosphinotricin acetyltransferase, glyphosate tolerance
PBP	penicillin binding protein
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PGT	partial horizontal gene transfer
p-value	probability value
RDP	recombination detection program
RHS	right hand side
sp.	species (singular)

spp.	species (plural)
ssDNA	single stranded DNA
TCS	two-component regulatory system
T4P	Type IV pili
tfp	transformation pili

Executive Summary

Mosaic genes are genetic entities consisting of DNA segments of different phylogenetic origin leading to sequence patterns which encode new phenotypic properties. Such intragenic recombination accelerates bacterial evolution and adaptability to new environmental conditions. Its deleterious impact on human and animal health is exemplified with mosaic penicillin binding proteins of *Streptococcus pneumoniae* which rendered this pathogen resistant to multiple ß-lactam antibiotics. Mosaic gene formation relies primarily on the uptake of free DNA from the environment by competent bacteria via natural genetic transformation and subsequent integration of the incoming DNA fragment into the bacterial genome via homologous recombination. Although most of the transgenic plants currently approved for marketing and/or cultivation contain transgenes of bacterial origin (e.g. *cry*, *epsps*, *pat* etc...) or carry prokaryotic antibiotic resistance marker genes (e.g. *aph(3')-IIa* / *nptII*) or fragments of the vector backbone used for genetic manipulations the formation of mosaics with similar chromosomal sequences in competent soil or gastrointestinal bacteria has not been in the focus of risk assessment bodies so far. According to EFSA mosaic gene formation would be an exotic phenomenon demonstrated only in a few highly transformable species for a limited number of genes under severe selection pressure.

The present report provides evidence for the relevance of mosaic gene formation for GMO risk assessment:

Mosaic gene formation is not restricted to highly transformable bacterial species, but appears to occur throughout the whole bacterial kingdom and to affect genes of all cellular functions. Natural competence for DNA uptake is being observed in an ever growing number of bacterial species totaling currently (i.e. November 2014) at more than 130 bacterial species experimentally proven to be competent under specific conditions. Moreover functional competence genes are continuously being discovered in novel bacterial genome sequences. Natural transformability is tightly regulated via complex regulatory circuits affecting the induction of competence. Stimuli inducing competence in certain species include e.g. DNA or cell wall damage, starvation, nucleotide and oxygen supply, pH, temperature, proton motive force and cell density. Mosaic gene formation relies on recA dependent homologous recombination guiding the single stranded form of the intruding DNA molecule to its chromosomal target. The rate of mosaic gene formation is therefore dependent on the extent of sequence similarity among the involved DNA molecules which decreases in a log-linear relationship with increasing sequence divergence and is falling below the limit of detection if nucleotide sequence divergence exceeds 25 - 30%. Plant DNA fragments with high sequence dissimilarity to bacterial receptor sequences are, thus, per se poor substrates for homologous recombination with bacterial chromosomes. Non-homologous DNA fragments may be also integrated into bacterial genomes via transformation - however at lower rates - by homology-directed illegitimate recombination requiring only a short homologous anchor sequence and regions of microhomology of 3 - 10 bp at the opposite end of the incoming strand of the foreign DNA with the chromosomal target. There appears to be no strict boundaries for DNA fragment lengths to be integrated into bacterial chromosomes by natural transformation. Transferred segment lengths between 12 bp and > 1 megabasepairs have been observed. Hot spots of mosaic gene formation are expected in environments with high bacterial cell densities and in locations reported to support high rates of horizontal gene or gene fragment transfer especially mediated by natural transformation. Important examples are the rhizosphere, phyllosphere and plant tissue, manured soil, the mammalian gastrointestinal tract, and sewage plants. Natural transformation is also reported in sediments, in the water column and in the food

matrix (e.g. dairy products). Mosaic gene formation is therefore also to be expected to occur in the food/feed chain. The mammalian oropharynx and the upper respiratory and the urogenital tract are carriers of highly transformable bacteria (e.g. *Neisseria spp.*, *Streptococcus spp.*) and, thus, environments supportive for mosaic gene formation.

To gain information about the potential involvement of plant derived antibiotic resistance marker and a glyphosate tolerance mediating transgene - both of bacterial origin - in the formation of mosaic genes a public sequence database (GenBank) was screened for *aph(3')-IIa* and CP4 *epsps* homologs. The resulting sequence collections were analyzed with sophisticated recombination detection algorithms to identify possible mosaic variants of *aph(3')-IIa* and CP4 *epsps* already present in natural environments. The overall variability of available *aph(3')-IIa* homologs was low: 48 of a total of 51 homologs shared more than 99% sequence identity with *aph(3')-IIa* from the transposon Tn5 (GenBank accession# V00618). For CP4 *epsps,* 144 homologs were deposited in GenBank with sequence similarities of 70 - 87% to the original CP4 *epsps* gene and of 62 - 74% to coCP4*epsps*, the codon optimized version of the transgene. Sequence similarities among the CP4 *epsps* homologs reflected the phylogenetic relatedness of the source organisms, suggesting primarily a vertical inheritance pattern of this gene. However, one recombination event was identified among *aph(3')-IIa* homologs with high confidence and several recombination events were detected among CP4 *epsps* homologs. Thus, both *aph(3')-IIa* and CP4 *epsps* appear to be astonishingly stable genetic entities in naturally occurring bacterial populations but a priori susceptible to intragenic recombination. In addition a general survey of literature and sequence databases for characterized mosaic genes revealed numerous hits covering literally all functional categories of genes and occurring in diverse bacterial phyla.

To test whether *aph(3')-IIa / nptII* has the potential to form mosaics with similar aminoglycoside phosphotransferase genes an experimental *Acinetobacter baylyi* model (carrying *aph(3')-Va* as DNA fragment acceptor aminoglycoside phosphotransferase gene) was developed. An anchor sequence mediated integration of the antibiotic resistance marker gene *aph(3')-IIa* / *nptII* into *aph(3')-Va* leading to a characteristic alteration of the antibiotic resistance profile of an *Acinetobacter baylyi* recipient strain could be shown. The frequency of this homology-directed illegitimate recombination was approximately 10^{-7}; the length of the terminal region of microhomology was 10 – 12 nucleotides. Formation of mosaic patterns among *nptII* (DNA fragment donor) – *aph(3')-Va* (acceptor gene) could not be demonstrated in the *in vivo* model. However, our results substantiate the importance of anchor sequence mediated gene transfer by homology-directed illegitimate recombination for the risk assessment of GMOs considering the fact that most genetically modified plants contain transgenes or vector backbone sequences of microbial origin which could function as homologous recombination anchors with chromosomal sequences already present in competent bacteria.

In a bioinformatic probabilistic modeling approach, two models were developed to investigate the probability and the impact of a successful mosaic gene formation event in the intestinal tract of pigs fed with transgenic plants and in manure. The simulations were run for various bacterial generation times, recombination rates and selection coefficients, taking into account the bacterial cell number, growth rate and movement in the gut content, DNA degradation and DNA diffusion, and the number of ingested transgenic plant DNA molecules. The gut model investigated the time until the first successful mosaic gene formation occurs in the intestinal tract of pigs, and the probability of such an event to occur within the life span of a single pig and within the entire Austrian pig population. The results of the probabilistic framework showed that the expected time for a mosaic gene formation event far exceeds the life span of a single pig when applying realistic (i.e. very low) recombination

rates. However, under strong positive selection pressure a mosaic gene formation event becomes likely in the life span of a single pig at recombination rates of 10^{-12} to 10^{-11} and in the total annual Austrian pig production at recombination rates as low as 10^{-18} to 10^{-16}. The second model focused on mosaic gene formation in an agricultural liquid manure tank. Formation of a mosaic gene was assumed to have had occurred in the digestive tract of a pig and that manure with bacteria carrying the mosaic gene is introduced into the tank, where the novel gene provides a selective advantage. The results show that with sufficiently large selection coefficients, the multiplication effect of the novel genetic material within the bacterial host population becomes rather strong, resulting in a considerable number of bacterial cells carrying the novel gene which are eventually released into the environment. Both models identified selection pressure as the key driver for mosaic gene formation and fixation.

Conclusions

The likelihood for the formation of mosaic antibiotic resistance genes with transgenic plant DNA involvement is low. It is justified to assume that similar mosaic gene formation processes will occur with a significantly higher rate already naturally in bacterial populations compared to processes involving transgenic plant DNA (i.e. the naturally occurring background rate will be comparatively high). It is therefore questionable whether the relative contribution of transgenic plant DNA for the formation of mosaic genes will be strong enough to be of any biological relevance. However, this assumption remains to be verified experimentally. There are substantial knowledge gaps concerning the actual frequency of mosaic gene formation and the kind of selection pressure prevailing in natural habitats. Moreover, a low likelihood of such an event is not predictive for an absence of any adverse long-term effects induced by this event. For example devastating adverse effects on public health of an extremely unlikely mosaic penicillin resistance gene formation was documented in a clinical setting. It is therefore advisable that risk assessment bodies put the formation of mosaic genes on their agenda and take it into account on a routine basis for the risk assessment of transgenic plants containing microbial-derived DNA. Additionally, it would be advisable to increase research efforts on this topic to narrow the still prevailing knowledge gaps and reduce the uncertainties currently linked with the risk assessment of mosaic genes.

Zusammenfassung

Mosaikgene sind genetische Einheiten, die aus DNA-Segmenten verschiedenen phylogenetischen Ursprungs bestehen, welche wiederum zu Sequenzmustern führen, die neue phenotypische Eigenschaften kodieren. Eine derartige intragene Rekombination beschleunigt die bakterielle Evolution und Adaptationsfähigkeit an neue Umweltbedingungen. Als Beispiel für ihre schädigende Auswirkung auf die menschliche und tierische Gesundheit können penicillin-bindende Mosaikproteine von *Streptococcus pneumoniae* angeführt werden, die zu einer Resistenz dieses Pathogens gegenüber multiplen ß-lactam Antibiotika geführt haben. Die Mosaikgen-Bildung beruht primär auf der Aufnahme von freier DNA aus der Umwelt durch kompetente Bakterien über eine natürliche genetische Transformation und nachfolgende Integration des aufgenommenen DNA-Fragments in das Bakteriengenom via homologe Rekombination. Obwohl die meisten transgenen Pflanzen, die gegenwärtig für das Inverkehrbringen und/oder die Kultivierung zugelassen sind, Transgene bakteriellen Ursprungs (z.B. *cry*, *epsp*, *pat*, etc.) enthalten oder Träger von prokaryotischen Antibiotikaresistenz-Markergenen (z.B. *aph(3´)-IIa / nptII*) oder von Fragmenten des Vektorgerüstes sind, das für genetische Manipulationen verwendet wird, befand sich die Bildung von Mosaiken mit ähnlichen chromosomalen Sequenzen in kompetentem Boden oder gastrointestinalen Bakterien bislang nicht im Fokus institutionalisierter Risikobewerter. Gemäß EFSA sei die Mosaikgen-Bildung ein exotisches Phänomen, das nur in einigen wenigen hoch-transformierbaren Spezies für eine begrenzte Anzahl von Genen unter massivem Selektionsdruck demonstriert wurde.

Der gegenständliche Bericht liefert den Nachweis für die Relevanz der Mosaikgen-Bildung für die GVO-Risikobewertung:

Die Mosaikgen-Bildung ist nicht auf hoch-transformierbare Bakterienspezies beschränkt, sondern scheint sich durch das ganze Bakterienreich zu ziehen und Gene für alle zellulären Funktionen zu betreffen. Die natürliche Kompetenz für eine DNA Aufnahme kann in einer stetig wachsenden Anzahl von Bakterienspezies beobachtet werden, die gegenwärtig (Stand: November 2014) mehr als 130 Bakterienspezies umfasst, bei denen der experimentelle Beweis für Kompetenz unter spezifischen Bedingungen erbracht wurde. Darüber hinaus werden funktionelle Kompetenz-Gene laufend in neuen bakteriellen Genomsequenzen entdeckt. Die natürliche Transformierbarkeit wird eng über komplexe Regulationskreisläufe, die die Kompetenzinduktion beeinflussen, gesteuert. Stimuli, die Kompetenz in bestimmten Spezies induzieren, schließen z.B. DNA- oder Zellwand-Schädigung, Ernährungsmangel, Nukleotid- und Sauerstoff-Versorgung, pH, Temperatur, Protonengradient und Zelldichte ein. Die Mosaikgen-Bildung beruht auf einer recA abhängigen homologen Rekombination, die die Einzelstrangform des intrudierenden DNA Moleküls zu seinem Chromosomentarget führt. Das Verhältnis der Mosaikgen-Bildung ist daher abhängig vom Ausmaß der Sequenzähnlichkeit unter den involvierten DNA-Molekülen, die in einem log-linearen Verhältnis mit steigender Sequenzdivergenz abnimmt und unter die Bestimmungsgrenze fällt, wenn die Divergenz der Nukleotidsequenz 25 bis 30% überschreitet. Pflanzliche DNA-Fragmente mit hoher Sequenz-Dissimilarität gegenüber bakteriellen Rezeptor-Sequenzen sind daher per se nur mäßige Substrate für eine homologe Rekombination mit Bakterienchromosomen. Nicht-homologe DNA-Fragmente können ebenfalls in Bakteriengenome via Transformation integriert werden – wenn auch in geringerem Ausmaß -, und zwar durch homologie-geleitete illegitime Rekombination, die nur eine kurze homologe Ankersequenz und Regionen der Mikrohomologie von 3 – 10 bp mit dem chromosomalen Target am gegenüberliegenden Ende des hereinkommenden Stranges der Fremd-DNA erfordert. Für die Integration von DNA ins bakterielle Genom via naturlicher Transformation scheint es keine strikten

Grenzen bezüglich der DNA-Fragment-Längen zu geben. Tansferierte Segmentlängen zwischen 12 bp und > 1 Mbp wurden beobachtet. Hot spots der Mosaikgen-Bildung können in Umgebungen mit hohen Bakterienzelldichten und an Orten erwartet werden, bei denen berichtet wurde, dass sie hohe Raten an horizontalem Gen- oder Genfragment-Transfer unterstützen, besonders mediiert durch natürliche Transformation. Bedeutende Beispiele sind die Rhizosphäre, Phyllosphäre und Pflanzengewebe, gedüngter Boden, der Gastrointestinaltrakt von Säugetieren sowie Kläranlagen. Natürliche Transformation wird auch berichtet in Sedimenten, in der Wassersäule und in der Lebensmittelmatrix (z.B. Milchprodukte). Mosaikgen-Bildung kann daher auch in der Lebensmittel-/Futtermittelkette erwartet werden. Oropharynx-, oberer Respirations- und Urogenitaltrakt von Säugetieren sind Träger von hochtransformierbaren Bakterien (z.B. *Neisseria spp.*, *Streptococcus spp*.) und daher vom Umfeld her für eine Mosaikgen-Bildung begünstigend.

Um Information über eine mögliche Beteiligung eines von Pflanzen abstammenden Antibiotikaresistenz- Markers und eines Glyphosat-Toleranz vermittelnden Transgens – beide bakteriellen Ursprungs – an der Bildung von Mosaikgenen zu erhalten, wurde eine öffentliche Sequenzdatenbank (GenBank) auf *aph(3´)-IIa* und CP4 *epsps* Homologe gescreent. Die resultierenden Sequenzsammlungen wurden mit Hilfe ausgeklügelter Rekombinations-Detektions-Algorithmen analysiert, um eventuelle Mosaik-Varianten von *aph(3´)-IIa* und CP4 *epsps*, die bereits in natürlichen Milieus vorkommen, zu identifizieren. Die Gesamtvariabilität der verfügbaren *aph(3´)-IIa* Homologe war niedrig: 48 von insgesamt 51 Homologen teilte mehr als 99% Sequenz-Identität mit *aph(3´)-IIa* aus dem Transposon Tn5 (GenBank accession# V00618). 144 Homologe waren in GenBank mit Sequenz-Ähnlichkeiten von 70 – 87% gegenüber dem originalen CP4 *epsps* Gen und von 62 – 74% gegenüber *co*CP4*epsps*, der Codon-optimierten Version des Transgens, abgelegt. Sequenzähnlichkeiten unter den CP4 *epsps* Homologen widerspiegelten die phylogenetische Verwandtschaft der Quellorganismen und deuteten primär ein vertikales Vererbungsmuster dieses Gens an. Dennoch wurde mit hoher Wahrscheinlichkeit ein Rekombinationsereignis unter *aph(3`)-IIa* Homologen identifiziert, außerdem wurden mehrere Rekombinationsereignisse unter CP4 *epsps* Homologen detektiert. Somit scheinen sowohl *aph(3´)-IIa* als auch CP4 *epsps* genetisch erstaunlich stabil in natürlich auftretenden Bakterienpopulationen zu sein. Sie sind aber a priori empfänglich für intragenetische Rekombination. Darüber hinaus erbrachte eine allgemeine Suche in Literatur- und Sequenz-Datenbasen nach charakterisierten Mosaikgenen zahlreiche Treffer, die buchstäblich alle funktionalen Kategorien abdeckten und in verschiedenen Bakterien-Phyla auftraten.

Um zu testen, ob *aph(3´)-IIa* das Potential besitzt, Mosaike mit ähnlichen Aminoglykosid-Phosphotransferase-Genen zu bilden, wurde ein experimentelles *Acinetobacter baylyi* Modell, das *aph(3´)-Va* als DNA-Fragment- Acceptor-Aminoglykosid-Phosphotransferase-Gen trägt, entwickelt. Eine Ankersequenz mediierte Integration des Antibiotikaresistenz-Marker-Gens *aph(3´)-IIa* in *aph(3´)-Va,* die zu einer charakteristischen Veränderung des Antibiotikaresistenz-Profils eines *Acinetobacter baylyi* Rezpienten-Stamms führte, konnte gezeigt werden. Die Frequenz dieser Homologie-geleiteten illegitimen Rekombination betrug ungefähr 10^{-7}; die Länge der terminalen Region der Mikrohomologie umfasste 10 – 12 Nukleotide. Die Bildung von Mosaikmustern zwischen *nptII* (DNA-Fragment-Donor) und *aph(3´)-Va* (Akzeptorgen) konnte im *in vivo* Modell nicht gezeigt werden. Nichtsdestotrotz unterstreichen die Resultate die Bedeutung des Anker-Sequenz-vermittelten Gentransfers durch Homologie-geleitete illegitime Rekombination für die Risikobewertung von GVOs, wenn man die Tatsache berücksichtigt, dass die meisten genetisch modifizierten Pflanzen Transgene oder Vektor-Sequenzen mikrobiologischen Ursprungs enthalten, die als homologe-Rekombinations-

Anker mit chromosomalen Sequenzen, die in kompetenten Bakterien bereits vorhanden sind, fungieren können.

In einem bioinformatischen probabilistischen Modellierungsansatz wurden zwei Modelle entwickelt, um die Wahrscheinlichkeit und den Impakt eines erfolgreichen Mosaikgen-Bildungs-Ereignisses im Intestinaltrakt von Schweinen, die mit transgenen Pflanzen gefüttert worden waren, sowie in Gülle zu erforschen. Die Simulationen wurden bei verschiedenen bakteriellen Generationszeiten, Rekombinationsraten und Selektionskoeffizienten durchgeführt, wobei die bakterielle Zellzahl, Wachstumsrate und Motilität im Darminhalt, DNA Abbau und Diffusion sowie die Anzahl der aufgenommenen transgenen Pflanzen-DNA-Moleküle berücksichtigt wurden. Das Darmmodell untersuchte den Zeitraum, bis die erste erfolgreiche Mosaikgen-Bildung im Intestinaltrakt von Schweinen auftritt, sowie die Wahrscheinlichkeit, inwieweit ein derartiges Ereignis innerhalb der Lebenszeit eines einzelnen Individuums aber auch der gesamten österreichischen Schweinepopulation zu erwarten ist. Die Resultate des probabilistischen Konstrukts zeigten, dass die zu erwartende Zeit für ein Mosaikgen-Bildungs-Ereignis bei weitem die Lebensdauer eines Einzelindividuums überschreitet, wenn man realistische (d.h., sehr niedrige) Rekombinationsraten in Betracht zieht. Aber unter stark positivem Selektionsdruck wird ein Mosaikgen-Bildungsereignis durchaus wahrscheinlich, und zwar im Lebenszeitraum eines Einzelindividuums bei Rekombinationsraten von 10^{-11} bis 10^{-12}, und in der gesamten österreichischen Schweineproduktion bei Rekombinationsraten, die so niedrig wie 10^{-16} bis 10^{-18} sind. Das zweite Modell war auf eine Mosaikgen-Bildung in einem landwirtschaftlichen Flüssiggülle-Tank fokussiert. Dabei wurde angenommen, dass die Mosaikgen-Bildung im Verdauungstrakt eines Schweines geschah und dass Gülle mit Bakterien, die das Mosaikgen tragen, in den Tank eingebracht wurde, wo das neuartige Gen einen selektiven Vorteil bietet. Die Ergebnisse zeigen, dass mit ausreichend großen Selektionskoeffizienten der Multiplikationseffekt des neuartigen genetischen Materials innerhalb der bakteriellen Trägerpopulation ziemlich stark wird, was in einer beträchtlichen Anzahl von Bakterienzellen resultiert, die das neuartige Gen beherbergen und letztendlich in die Umwelt freigesetzt werden. Beide Modelle identifizierten den Selektionsdruck als Schlüsselfaktor für eine Mosaikgen-Bildung und –Fixierung.

Schlussfolgerungen

Die Wahrscheinlichkeit für die Bildung von Mosaik-Antibiotikaresistenz-Genen unter Beteiligung von transgener Pflanzen-DNA ist gering. Man kann davon ausgehen, dass die Mosaikgenbildung in signifikant höherer Rate bereits natürlicherweise innerhalb von Bakterienpopulationen vorkommt als in Prozessen unter Beteiligung transgener Pflanzen-DNA (d.h., die natürlich vorkommende Hintergrundrate wird vergleichsweise hoch sein). Es ist daher fraglich, ob der relative Beitrag transgener Pflanzen-DNA zur Bildung von Mosaikgenen groß genug ist, um von biologischer Relevanz zu sein. Jedoch bleibt diese Annahme experimentell abzuklären. Wesentliche Wissenslücken bestehen betreffend die tatsächliche Frequenz einer Mosaikgen-Bildung im natürlichen Habitat und die Art des Selektionsdrucks, der dort vorherrscht. Darüber hinaus ist eine niedrige Wahrscheinlichkeit für ein derartiges Ereignis nicht prädiktiv für eine Abwesenheit jeglicher adverser Langzeiteffekte, die durch dieses Ereignis hervorgerufen werden könnten. Die negativen Effekte einer extrem unwahrscheinlichen Mosaik-Penicillin-Resistenzgen-Bildung in Bezug auf die Volksgesundheit konnten in einem klinischen Setting dokumentiert werden. Es ist daher anzuraten, dass Risikobewertungsinstitutionen die Bildung von Mosaikgenen auf ihre Agenda setzen und auf Routinebasis in der Risikobewertung transgener Pflanzen, die mikrobielle DNA enthalten, berücksichtigen. Zusätzlich wäre es ratsam, die Forschungsanstrengungen hinsichtlich dieses Themas

zu erhöhen, um die immer noch herrschenden Wissenslücken zu verringern und die Ungewissheiten, die gegenwärtig mit der Risikobewertung von Mosaikgenen verbunden sind, zu reduzieren.

1. Mosaic Genes – Current State of Knowledge

1.1 Terms of Reference

Mosaic genes are involved in the formation of genetic entities coding for altered proteins which allow bacteria to develop novel resistances to antimicrobial agents, to escape the immune defence of the eukaryotic host, to increase their virulence or to help pathogens to escape vaccination schemes. The risk of mosaic genes to induce adverse effects on human and animal health is, thus, not restricted to the issue of antibiotic resistance development and spread; however, the impact of mosaic gene formation in the area of antibiotic resistance is more comprehensively documented in the scientific literature and the effects of mosaic gene formation are worsening a generally already deteriorating situation in antibiotic resistance and public health.

The present analysis, therefore, discusses the issue of mosaic gene formation with a focus on antibiotic resistance development involving plant derived transgenic DNA.

The phenomenon of mosaicism also occurs on the genome level being the cause for mosaic chromosomes, for variations in mitochondrial DNA and is prevalent in eukaryotic genes. However, these observations are not discussed and out of scope of this report.

1.2 Summary

Mosaic genes are genetic entities consisting of DNA segments of different phylogenetic origin leading to sequence patterns which usually encode new phenotypic properties. The reshuffling and recombination of DNA fragments increases the genetic plasticity of the targeted genome and substantially accelerates bacterial evolution and adaptability to new environmental conditions. The deleterious impact of this process on human and animal health is representatively exemplified with mosaic penicillin binding proteins of *Streptococcus pneumoniae* which renders this pathogen resistant to various ß-lactam antibiotics. This development has resulted in an increase in morbidity and mortality in patients suffering from infectious pneumonia and to a substantial financial burden for public health. The process of mosaic gene formation relies primarily on the uptake of free DNA from the environment by competent bacteria via natural genetic transformation and subsequent integration of the incoming DNA fragment into the bacterial genome via homologous recombination. The efficiency of DNA segment integration is dependent on sequence similarity between the involved DNA strands working most efficiently between mostly identical DNA molecules. The simultaneous introduction of a few single nucleotide polymorphisms encoded on the incoming DNA strand might be sufficient to change the binding site of the antibiotic in pencillin binding proteins or to evade the host immune system via subtle changes in antigenic surface proteins.

The strict dependence on sequence similarity between incoming and recipient DNA is relieved by homology-directed illegitimate recombination (as shown in *S. pneumoniae* and *Acinetobacter baylyi*) requiring only a short stretch of homologous DNA on one side of the incoming DNA fragment and a region of microhomology on the opposite terminus of the molecule. This homologous “anchor” sequence of high sequnce similarity with a chromosomal target region can be as short as 153 bp, the target site for microhomologies between incoming and chromosomal DNA may encompass only 3 – 10 bp.

Although most of the transgenic plants currently approved for marketing and/or cultivation contain transgenes of bacterial origin (e.g. *cry*, *epsps*, *pat* etc...) or carry prokaryotic antibiotic resistance marker genes (e.g. *aph(3')-IIa* / *nptII*) or fragments of the prokaryotic vector backbone used for genetic manipulations the formation of mosaics with similar chromosomal sequences in comptent soil or gastrointestinal bacteria was not considered as a hazard and is still not in the focus of risk assessment bodies. Mosaic genes would be an exotic phenomenon demonstrated only in a few highly transformable species for a limited number of genes under severe selection pressure was the main line of argumentation for ignoring the phenomenon for GMO risk assessment.

The present report provides evidence in support of the relevance of mosaic gene formation also for GMO risk assessment.

Mosaic gene formation is not restricted to a small group of highly transformable bacteria nor is it confined to bacterial genes under high selection pressure. On the contrary it is demonstrated to be a widespread phenomenon involving transformable and non-transformable bacteria throughout the whole bacterial kingdom and genes coding for all cellular functions. It is reasonable to assume that DNA fragment exchange is a frequent process in bacterial cells because it is a major component of the DNA repair machinery. Moreover, many of these DNA fragment replacements do not lead to changes in the bacterial phenotype and, thus, are not readily detectable for the experimenter. This may be the cause for underestimating the actually occuring frequency for DNA fragment exchange under natural conditions. To date it appears that only two factors are limiting the formation of mosaic genes: i) an increasing sequence divergence among donor and recipient DNA is reducing the efficiency of homologous recombination in a log-linear relationship to the point where it falls below the limit of detection – which is usually the case when dual sequence identity drops below 70%; ii) the second requirement is the presence of competent bacteria primed for the uptake of free DNA. We provide data for more than 130 bacterial species experimentally proven to be naturally compentent for the uptake of free DNA under distinct environmental conditions. This is an increase in abundance of approximately 50 taxa since the last review on this topic as published by Johnston et al. in Nature Reviews Microbiology in 2014.[3] There are indications that many more bacterial species may become competent under certain environmental conditions because the genetic signature for functional competence genes is present in several bacterial phyla/classes (e.g. in all γ-proteobacteria). For many of these strains the conditions necessary for the induction of competence simply have not been discovered in laboratory experiments, yet.

Natural transformability is tightly regulated via complex regulatory circuits involved in the fine tuning of competence induction: A central master regulator of competence is responsible for the activation of competence regulons which basically encode the components for the bacterial DNA uptake machinery - the transformasome. This primary control mechanism for the induction of competence is astonishingly conservered over all analysed bacterial phyla. However, the expression of the central master regulator of competence proteins is itself under control of a plethora of different species-specific regulatory cascades linked to intracellular signaling pathways which sense intra and extra cellular stimuli like DNA or cell wall damage, nutrient starvation, nucleotide and oxygen supply, pH, temperature, proton motive force and cell densities in the environment as most prominent examples. This extreme fine tuning of competence induction is necessary to allow i) rapid adaptation to various changing environmental parameters (i.e. alternating selection pressure) and ii) to avoid deleterious effects of transformation like chromosome instability and asynchronicity of replication and cell division after uptake of foreign DNA. As the formation of mosaic genes is inherently linked to

natural transformation the generation of mosaic genes appears to be the product of a tightly regulated process in response to changing intra- and extracellular conditions and adaption to selection pressure. Mosaic gene formation relies on recA dependent homologous recombination guiding the single stranded form of the intruding DNA molecule to its chromosomal target. The rate of mosaic gene formation is therefore dependent on the extent of sequence similarity among the involved DNA molecules which decreases in a log-linear relationship with increasing sequence divergence and is falling below the limit of detection if nucleotide sequence divergence exceeds 25 - 30%. Plant DNA fragments with high sequence dissimilarity to bacterial receptor sequences are, thus, per se poor substrates for homologous recombination with bacterial chromosomes. However, completely non-homologous DNA fragments may be integrated - at substantially lower frequencies than homologous counterparts - into bacterial genomes by homology-directed illegitimate recombination requiring only a short homologous anchor sequence and regions of microhomology of 3 - 10 bp at the opposite end of the incoming strand of the foreign DNA. There appears to be no strict boundaries for DNA fragment lengths to be integrated into bacterial chromosomes by natural transformation. Transferred segment lengths between 12 bp and > 1 megabasepairs have been observed.

Hot spots of mosaic gene formation are to be expected in environments with high bacterial cell densities and in locations reported to support high rates of horizontal gene or gene fragment transfer especially mediated by natural transformation. Important examples are the rhizosphere, phyllosphere and plant tissue, manured soil, the mammalian gastrointestinal tract, and sewage plants. Natural transformation is also reported in sediments, in the water column and in the food matrix (e.g. dairy products). The mammalian oropharynx and upper respiratory and the urogenital tract are carriers of highly transformable bacteria (e.g. *Neisseria spp.*, *Streptococcus spp.*) and therefore also targets for mosaic gene formation.

Summarizing the available data the likelihood for the formation of mosaic antibiotic resistance genes with transgenic plant DNA involvement is low. It is justified to assume that similar mosaic gene formation processes will occur with a significantly higher rate already naturally in bacterial populations compared to processes involving transgenic plant DNA (i.e. the naturally occurring background rate will be comparably high). It is therefore questionable whether the relative contribution of transgenic plant DNA for the formation of mosaic genes will be high enough to be of biological relevance. However, this assumption remains to be verified experimentally. There are substantial knowledge gaps concerning the actual frequency of mosaic gene formation and the kind of selection pressure prevailing in natural habitats. Moreover, a low likelihood of such an event is not predictive for an absence of adverse long-term effects induced by this event. Severe adverse effects on public health of an extremely unlikely mosaic resistance gene formation could have been already demonstrated in a clinical setting for mosaic pencillin binding proteins of *Streptococcus pneumoniae*. It is therefore recommended to try to increase the awareness of risk assessment bodies for this phenomenon and to take mosaic gene formation into account on a routine basis for the risk assessment of transgenic plants containing microbial-derived DNA. Additionally, it would be necessary to increase research efforts on this topic to narrow the still prevailing knowledge gaps and reduce the uncertainties currently linked with the risk assessment of mosaic genes.

1.3 Background

During the recent years of collaboration with the EFSA GMO Panel to improve the risk assessment of GMOs it became obvious that the issue of mosaic gene formation was not in the focus of this risk assessment body and was not considered as relevant on a routine basis.[4-6] Mosaic gene formation - especially in the case of *aph(3')-IIa / nptII* applied as antibiotic resistance marker genes in transgenic plants – was according to EFSA of no relevance because mosaic gene formation would be (personal communication, 2013):

a) an exotic phenomenon only demonstrated in a small number of highly transformable bacterial strains for
b) a limited number of bacterial genes under heavy selection pressure like surface antigens (selection by host immune system) or certain antimicrobial resistance determinants (selection by antibiotics) and
c) there would be no scientific information available reporting an involvement of *nptII* in the formation of mosaic genes.
We provide evidence that the formation of mosaic genes is a basic element in bacterial evolution providing a universally available tool for rapid adaptation to changing environments but also fulfilling crucial functions in the general cellular metabolism: "housekeeping genes" providing such essential functions for cellular survival as DNA replication, recombination and repair, cell division, chromosome partitioning, transcription, translation, ribosomal structure and biogenesis, secondary metabolites biosynthesis, transport and catabolism also take part in the exchange of gene fragments allowing them to produce a versatile and efficient response to changes in the intra- and extracelluar environment.[7]

A stringent bioinformatic analysis for the presence of gene mosaics performed by Zheng et al. revealed that at least 10 - 20% of the protein coding regions in all of the currently sequenced microbial genomes represent mosaic genes.[7]

The formation of mosaic genes is certainly not restricted to highly transformable bacteria because bacterial species usually classified as naturally non-transformable carry mosaic genes (e.g. *E. coli, Salmonella spp*.), as well as bacteriophages, viruses, and eukaryotes.[8-12] The formation of mosaic genes is not restricted by the kind of horizontal gene transfer mechanims – transformation and conjugation may serve equally well as mediators of the process; however, strains capable to take up free DNA from the environment are thought to utilize predominantly this pathway for providing the substrate for mosaic gene formation.[13, 14]

1.4 Aims

To provide an overview over the state of knowledge concerning mosaic genes literature databases (PubMed, ISI Web of Science/Scopus) are to be screened for relevant information with a focus on transferable antibiotic resistance. Additionally up-to-date information on natural transformation, competence and homologous recombination is to be collected on mosaic gene formation.

1.5 Introduction

Genetic recombination leading to mosaic patterns in antibiotic resistance genes result in therapy failure of infectious diseases in clinical settings.[15, 16] This observation is of particular concern because

pathogens may rapidly respond in this way to alternating antibiotic selection pressure by generating new resistance determinants with alternative targets or substrate specificities.

The majority of the commercially available GMOs contain prokaryotic transgenes like antibiotic resistance markers or other metabolically advantageous genes of bacterial origin.[17] These transgenes will be part of the DNA released into the environment (e.g. soil, animal gastrointestinal tract) upon decay of GMOs.[18] Natural transformation is a mechanism that allows competent bacteria to acquire free DNA horizontally.[19] Such DNA can be of a size that includes intact genes, or if further fragmented, of sub-gene size. Irrespective of size, such DNA may transfer into competent bacteria and recombine with their sequence-similar counterparts in naturally occurring bacterial populations. If the recombining regions are short, the resulting gene acquires a mosaic structure (= mosaic gene), in which sections of native gene sequences have been replaced by foreign sequence elements. Such mosaic genes may potentially generate a new phenotype that provides a benefit to the bacterial host under certain selective environmental conditions.

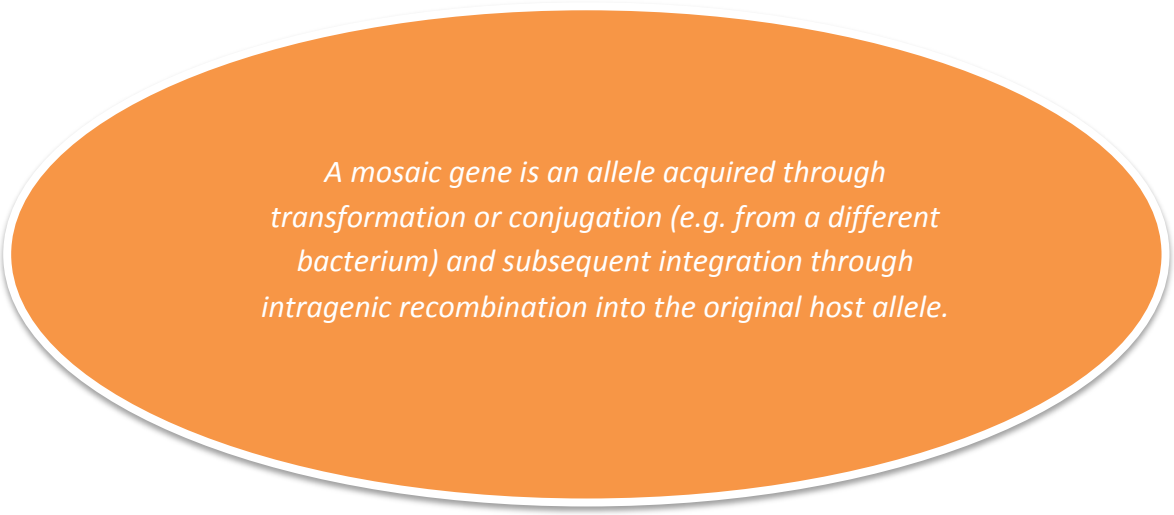
The transfer of plant DNA to soil or gut bacteria in natural environments is considered to be an extremely unlikely event.[5, 6, 20-24] Experimental evidence for plant transgene transfer outside the laboratory under field conditions is still lacking.[20, 25] However, it could be shown that there are massive deficiencies in the currently available methodology for the detection of rare horizontal gene transfers in the environment.[16, 26] Frequency estimates of bacterial horizontal gene transfers appear to be not informative for the prediction of long–term effects on animal or human health [27] and there are indications that also extremely rare events can be of substantial relevance for the risk assesssment if the selection pressure is strong enough to support the fixation of the trait in the respective bacterial population.[28, 29] The prediction of fitness adavantages for bacterial transformants is not straightforward due to the limited understanding of selection pressures active in natural environments.[30]

It is hypothesized that antibiotic resistance marker (ARM) genes (or fragments thereof) originating from genetically modified organisms (GMOs) have the potential to be involved in the formation of new antibiotic resistance determinants based upon the recombination and re-shuffling of ARM gene fragments with endogenously present bacterial resistance genes of sufficient sequence homology. Regulatory risk assessment bodies have not routinely considered this potential hazard and, consequently, have not included the formation of mosaic genes in the risk scenario analysis of commercialized GMOs. However, there is currently no experimental evidence available that can unambiguously support or disprove the hypothesis that ARM genes can be involved in the formation of mosaic genes. This knowledge gap should be narrowed by a detailed analysis of the available literature concerning the genetics, prevalence and ecology of mosaic genes in naturally occurring bacterial populations.

1.6 Current State of Knowledge

1.6.1 Mosaic Genes – Definition

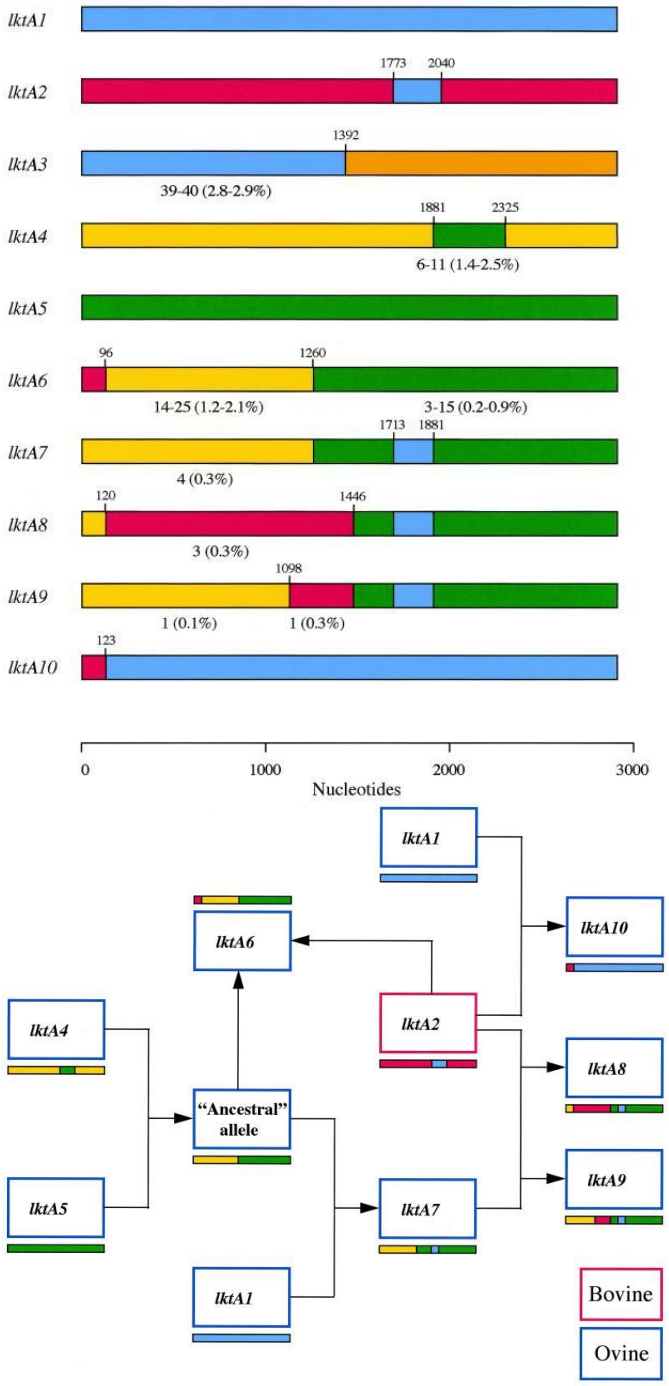
For the presented project the definition for a ”mosaic gene” as proposed by Alix Boc and Vladimir Makarenkov is applied:[31]



The term “mosaic” stems from the pattern of interspersed blocks of sequences having different evolutionary histories but found combined in the resulting allele subsequent to recombination events. The recombined segments can be derived from other strains of the same species or from other more distant bacterial relatives or of viral origin.[32-34]

The formation of mosaic genes is inherently connected with the concept of partial gene transfer in contrast to the traditional model of complete gene transfer where the incoming gene either replaces the orthologous gene of the recipient genome, or is added to it.[35]

A schematic representation of a typical mosaic gene structure is depicted in Figure 1A. The evolutionary relationship of the mosaic-like leukotoxin genes as shown in Figure 1A is exemplarily delineated in Figure 1B.



A) B)

Figure 1: Mosaic gene structure of leukotoxin A in *Pasteurella haemolytica*.

A) Schematic representation of the mosaic structures of alleles representative of the major allelic groups *lkt*A1 to *lkt*A10. The different colors indicate sequence identity and the likely origins of recombinant segments. The number of sites different from those in the corresponding region of the likely donor allele(s) and the degree of divergence are indicated below certain recombinant segments. All other segments exhibited 100% sequence identity to the corresponding regions of the donor alleles. Numbers above the proposed recombination sites indicate the position of the last nucleotide at the downstream end of the recombinant segment.
B) Proposed sequence of recombination events in the evolution of *lkt*A leading to the formation of *lkt*A8 and *lkt*A10 type alleles in the ovine-specific lineages represented by ETs 12 to 14 and 19 to 22. The central role of the bovine *lkt*A2 allele in the evolution of ovine alleles *lkt*A6, *lkt*A8, *lkt*A9, and *lkt*A10 is clearly seen. The mosaic structures of the alleles are as shown in Figure 1A (data from Davies et al.[36]).

1.6.2 Mosaic Genes – Horizontal gene transfer - Lateral sequence transfers

Horizontal or lateral gene transfer is dominant for various groups of genes in prokaryotes[31, 37], is a key driver of bacterial adaptation and evolution[33, 38-40] and plays a key role in the formation of bacterial species.[41, 42] A lack of genetic exchange is an unusual situation usually confined to a small number of genetically monomorphic pathogens.[43] It is scientific consensus that a significant amount of the bacterial genome is affected by full and partial horizontal gene transfers.[44-46]

During horizontal gene transfer exogenous genetic material is first transferred into the recipient cell and then integrated into the new host via recombination. Recombination events are characterized to introduce changes to regions of contiguous stretches of DNA and have the potential to introduce several mutations simultaneously in a single event in contrast to spontaneous point mutations.[47] The integrated genetic material may consist in general of stretches of noncoding DNA, fragments of genes (forming subsequently mosaics), entire genes, multiple (entire or fragmentary) adjacent genes, operons, transposable chromosomal elements, plasmids and naturally occurring extrachromosomal elements, and pathogenicity islands.[47] It is important to note that the lateral transfer of DNA sequences is not primarily linked to a particular size or function and is not restricted to gene boundaries. A large number of random DNA sequence transfers appear to occur independently of sequence length: Fragment lengths between 12 bases to > 1Mb have been documented.[48]

Units of DNA sequence transfer

There is little reason to assume that the units of transfer and recombination correspond to entire, intact genes.[49] Protein domains which have been suspected to constitute boundaries for efficient DNA transfers - because these would be the smallest functional units for selection - are units of function, but not modules of transfer and recombination. It was demonstrated that lateral sequence transfer can remodel even the most functionally conservative modules within genomes and lateral genetic transfer more commonly interrupts genes than preserves them intact.[49] All genes of the bacterial pangenome are proposed to have undergone lateral genetic transfers at some point of their phylogentic histories.[50]

For a stable integration of transferred sequences resulting in mosaic gene structures the following requirements have to be met:[48]

1. There must be a physical opportunity to encounter other species for the exchange of sequence elements, which is favored in bacterial communities with a sympatric lifestyle over cells living isolated allopatric lifestyles.

2. The recipient cell must have the ability to chromosomally integrate foreign DNA.

3. The available tRNA repertoire must allow the translation of the transferred sequences to result in gene expression

4. The expressed new gene product must allow positive selection of the new trait.

Detection of segmented DNA sequence transfers (mosaic genes)

The detection of segmented lateral sequence transfers is no trivial task. These transfers are usually masked by mutations and by species-specific codon optimization. However, bioinformatic algorithms have been developed to characterize these sequence transfers in selected sequence datasets based upon either atypical sequence features which exploit different nucleotide compositions (including the GC content) between native and incoming DNA fragments or employing phylogentic methods.[13, 14, 35, 48, 51, 52] All these bioinformatic approaches are affected with uncertainties therefore the application and combination of several different algorithms is advisable to minimize misleading results.[53]

1.6.3 Mosaic Genes – The paradigm of segmentally variable genes

Although it recently became clear that there are no physical boundaries concerning potential units of DNA sequence transfer[48, 49] an interesting metagenomic approach for the identification of mosaic genes using microbial whole genome sequence data is presented in this chapter.[7] To the best knowledge the publication of Zheng et al. was the first effort trying to collect quantitative information about the prevalence of mosaic genes in microbial genomes in a structured and comprehensive fashion and providing evidence for the involvement of mosaic gene formation in a broad spectrum of functionally divergent protein families besides antibiotic resistance and surface antigens.

Bioinformatic analysis as performed by Zheng et al. revealed that 10 – 20% of all bacterial protein coding genes may be classified as “mosaic genes” according to a stringent classification developed by the authors.[7] The authors developed a functional annotation concept for the definition of mosaic genes by requesting the variable DNA region of the mosaic gene to contain at least a contiguous stretch of 210 bp coding for a minimum of 70 amino acids flanked by conserved constant regions. This minimum length of the amino acid chain would permit folding into functional active protein domains.

Mosaic genes as defined by Zheng et al. usually consist of one or more highly variable region(s) which are interspersed between well-conserved sequence elements (Figure 2).

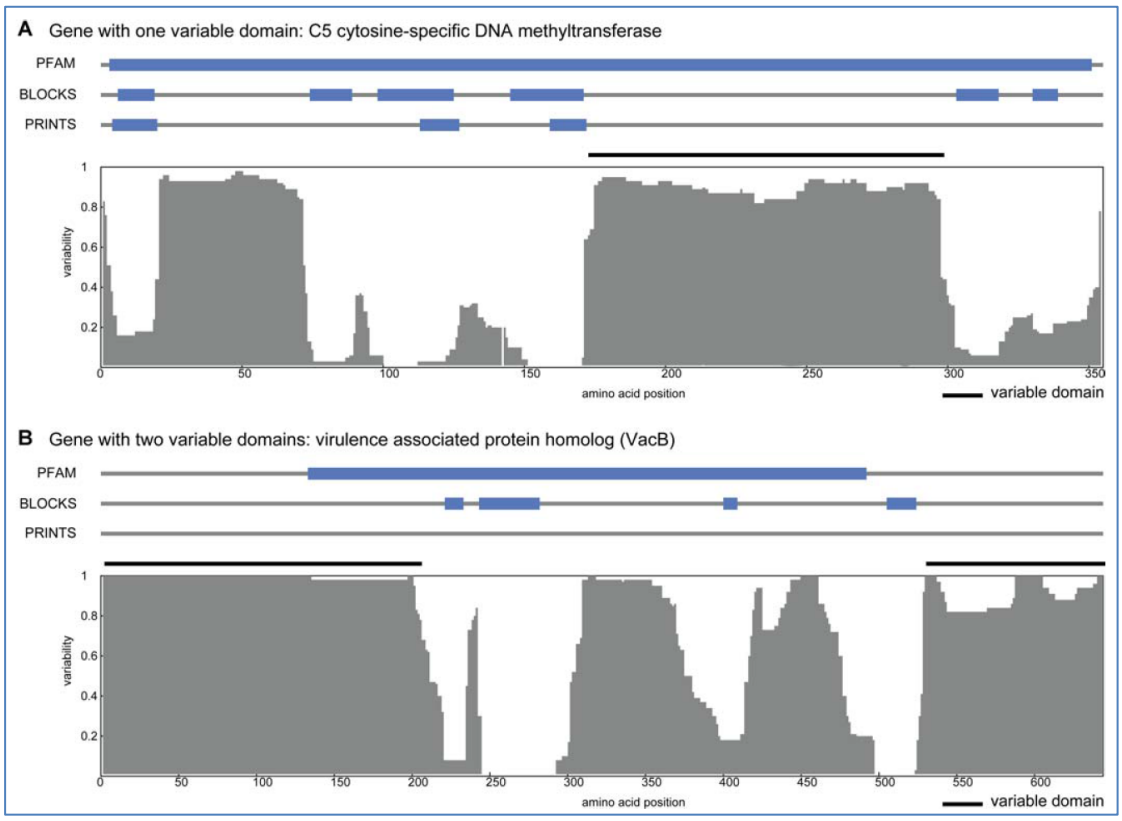


Figure 2: Examples of mosaic genes as defined by Zhen et al.

5'C-DNA methyltransferase and VacB of *Helicobacter pylori* are depicted. Variable domains are marked by black lines above the variabilty graph (data from Zhen et al.[7]).

The "mosaic gene" concept of Zheng et al. is not completely congruent with the definition proposed by Boc et al.[31], who do not restrict the exchange of gene fragments to defined variable regions (compare with Figure 1). However, as Zheng et al. provide a sophisticated bioinformatic analysis substantiating the issue of mosaic genes in an overall approach the data is discussed in this context. The shuffling and genetic recombination of whole functional genes between prokaryotic genetic entities are not considered as mosaic genes.[7]

An analysis of the functional category of the proteins encoded by mosaic genes revealed an overrepresentation of mosaic genes in the following cluster of orthologous groups[54]:

1. DNA replication, recombination and repair (in *Neisseria menigitidis*)
2. Cell division and chromosome partitioning (*Streptococcus pneumoniae*)
3. Secondary metabolites biosynthesis, transport and catabolis (*Helicobacter pylori*)

All three categories represent functions inherently connected with the concept of "housekeeping" genes usually thought to constitute stable – non-variant – entitities because they perform essential celluar functions necessary for the survival of the cell and, thus, are under strong selection pressure for functionality.[55] Housekeeping genes are usually conserved over family and even kingdom boundaries in the tree of life and are expected to form the genetic backbone of living organisms. Figure 3 lists functional categories where mosaic genes have been identified by Zhen et al.[7]

	COG functional category
Info store	Translation, ribosomal structure and biogenesis
	Transcription
	DNA replication, recombination and repair
Cellular processes	Cell division and chromosome partitioning
	Posttranslational modification, protein turnover, chaperones
	Cell envelope biogenesis, outer membrane
	Cell motility and secretion
	Inorganic ion transport and metabolism
	Signal transduction mechanisms
Metabolism	Energy production and conversion
	Carbohydrate transport and metabolism
	Amino acid transport and metabolism
	Nucleotide transport and metabolism
	Coenzyme metabolism
	Lipid metabolism
	Secondary metabolites biosynthesis, transport and catabolism
?	General function prediction only
	Function unknown

Figure 3: Functional classification of proteins encoded by mosaic genes.

The data are modified from Zheng et al.[7] COG: cluster of orthologous groups[54]

Zheng et al. point to the fact that their 70 amino acid coding capacity limit for mosaic genes was arbitrarily set.[7] This implies that the frequency of mosaic genes as defined by Boc et al. will be actually substantially higher. However the frequency data for mosaic genes by Zheng et al. are a good indication for their pivotal role in the general cellular metabolism, besides their commonly known function as surface antigens and antibiotic resistance determinants involved in host - pathogen interaction, defense mechanisms and intracellular responses to environmental changes.

An important question would be to determine the minimum length of the transferred DNA sequence which still should fall under the definition of a mosaic gene. If only a single base pair is resubstituted in a recipient genome by an external fragment it would become difficult to differentiate between horizontal gene transfers and single nucleotide polymorphisms induced by chance or mutagenic agents. This question is not yet solved; however, the minimum fragment length required for successful recombination would be an attractive option.[56]

The variety of mosaic genes is displayed in Annex 1, which contains a representative selection of these genes currently described in the recent scientific literature (as of November 2014). Antibiotic resistance determinants encoded by mosaic genes are elaborately studied with tetracycline (e.g. tet(O), tet(M), tet(W)), ß-lactam (penicillin bindin proteins), and macrolide (mefE, erythromycin) resistances. The overwhelming majority of scientific papers, however, is dealing with different proteins.

It is also noteworthy that DNA fragment exchange does not only take place in coding but also in intergenic regions, which may have effects on gene regulation and DNA uptake specificity.[47]

1.6.4 Mosaic Genes – The paradigm of pencillin binding proteins of *Streptococcus pneumoniae*

Streptococcus pneumoniae strains are carriers of five penicillin binding proteins (PBP) which have the potential to mediate clinically relevant resistance to ß-lactam antibiotics. The most important PBPs – altered variants of PBP2x, 2a and 2b – inducing high-level resistance to penicillins have been shown to be affected by a series mutations which have been caused by the introduction of contiguous segments of DNA originating from divergent streptococcal donor species (e.g. from *S. mitis*; Figure 4).

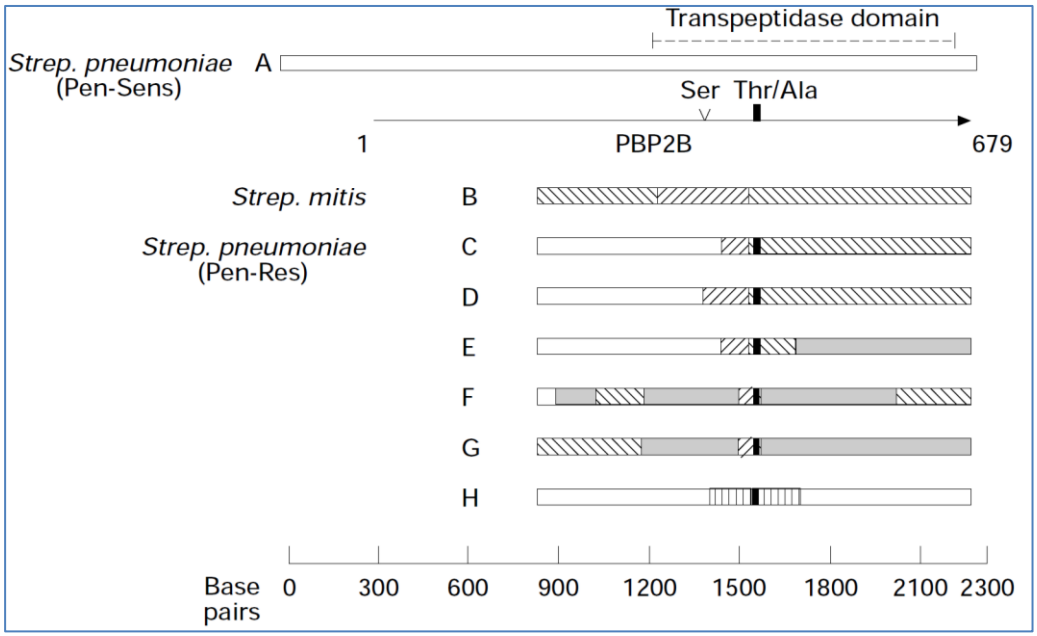


Figure 4: Mosaic penicillin binding proteins (PBP2P) of *Streptococcus pneumoniae*.

Streptococcus mitis has donated PBP2B genes in the evolution of penicillin resistance in *S. pneumoniae*. The PBP2B gene from a penicillin susceptible isolate of *S. pneumoniae* is represented by A. The line ending in an arrow shows the extent of the coding region. The location of the penicillin sensitive transpeptidase domain is shown by the dashed line. The sequenced regions of PBP2B genes from a penicillin susceptible strain of *S. mitis* are shown in B. C–H represent regions of PBP2B from various penicillin-resistant *S. pneumoniae*. The shading represents differing degrees of nucleotide divergence from the penicillin-susceptible *S. pneumoniae* strain A ranging from 4% to 21%. *S. pneumoniae* strains C–G possess blocks of nucleotides originally derived from *S. mitis*. The position of the active site serine residue is marked, as is the threonine/alanine alteration; the presence of the alanine, represented by a bar in strains C–H, is thought to be important in the development of a low affinity PBP2B (data from Dowson et al.[57]).

The resulting mosaic pattern of the gene for the affected penicilin binding protein lead to the formation of *S. pneumoniae* strains highly resistant to several types of ß-lactam antibiotics with deleterious effects for the treatment of patients suffering from bacterial pneumonia. Replacement of endogenous chromosomal sequences by external DNA fragments originating from similar penicillin binding protein genes of phylogenetically related streptococci is a process assumed to appear continuously in bacterial populations which share the same habitat. However, the result of this DNA fragment exchange process is usually invisible to most analyses because the sequence rearrangements are in many cases silent without causing a detectable phenotype or lethal. However, this process is the basis for an impressive genetic variability of the chromosomal apparatus of streptococci providing the means to allow rapid adaptation of the bacterial population to changing

environmental conditions i.e. in this case to alternating selection pressure induced by ß-lactam antibiotics. The process of mosaic gene formation provides the means for the affected bacterial populations to constantly develop new strategies to evade selection imposed by the application of alternative antimicrobials. The requirement for the application of ever new therapy regimens for the treatment of streptococcal infections in addition to higher morbidity, and mortality rates and prolonged hospitalizations of the affected patients leads to a substantial financial burden in public health. Approximately 21.5% of all *S. pneumoniae* isolates in the United States were determined to show high-level penicillin resistance in 1999 – 2000.[58] It, thus, took ca. fifty years of selection pressure induced by the application of penicillins to select for clinically significant levels of antibiotic resistance induced by mosaic genes. Although it was initially thought that mosaic gene formation is restricted to highly transformable species like *S. pneumoniae* it is now clear that many other resistance determinants in various bacterial pathogens are target for mosaic gene formation (Table 12 and Annex 1: Table 28). It appears to be that mosaic genes are rather the norm than exotic genomic aberrations in the antibiotic resistant microbial flora.[16, 59]

Ecological lag time

The time period until a new trait emerges and shows an observable effect – the ecological lag time - is dependent on the adaptive value of the new mosaic gene and on the strength of an appropriate selection pressure. Moreover, the new mutant line must successfully compete with the existing dominant flora of the habitat and it must be taken into account that a new trait may only become effective if the environment changes. The lag time between the generation of a recombinant – mosaic – phenotype and the impact of the novel phenotype on the environment or animal and human health is very difficult to predict and variable. Moreover it is important to realize that the kind of resistance which may arise from mosaic gene fragment recombinations is also not predictable. The strength of selection or its absence cannot always be known in advance.[60, 61]

Another difficulty for measuring the ecological lag time is the fact that when genes evolve by horizontal gene transfer rather than by vertical – clonal – reproduction neither the generation time nor the geographical range of the organisms necessarily limit the lag time.[16] This point is particularly relevant for attempts to measure horizontal gene transfers resulting in mosaic genes in field trials e.g. in soil bacterial communities: the combinatorial development of mosaic genes in decade time scales follows from the flow of genes across the globe, not through the generation of variability within plots.[16] Once a mosaic gene is made this new genetic element can be transferred with a much higher efficiency by homologous recombination as could combinations of recombinant genes be assembled in one or more different strains.[16, 62] The speed of penicillin resistance spread accelerated exponentially due to horizontal gene transfers after mosaics have had been formed. This development is far exceeding the speed for the dissemination of emergent clonal lines.[16]

Detection limits

The predicted probability for one recombination event per pbp gene to form a clinically relevant mosaic in *S. pneumoniae* was calculated by Heinemann et al. to be 1×10^{-24}.[16] The detection limit of horizontal gene transfers in soil bacterial communities is approximately between 10^{-8} and 10^{-11}.[63-68] The transfer of plant derived transgen DNA to soil bacteria was calculated to occur at a rate of 10^{-17} or even lower.[69, 70]

Implications

The data presented above allow the following interpretation:

1. An extremely rare event - like the mosaic gene formation in pbp genes of *S. pneumoniae* – nevertheless does occur and may result in a phenomenon of significant clinical relevance.[16]

2. The frequencies of plant to bacteria gene transfer in natural environment are so low that they cannot be detected with the current arsenal of methodologies.[16, 26]

3. A low likelihood of an event is not predictive for no adverse long term effects on animal, human or environmental health.[27]

1.6.5 Natural Transformation

Mosaic gene formation is inherently linked to natural genetic transformation which is a horizontal DNA transfer process characterized by the uptake of free DNA by a competent bacterium, its chromosomal integration or extrachromosomal stabilization, and its expression, which leads to a new phenotype of the recipient.[8, 71, 72] It is to be discriminated from artificial transformation which is used for genetic modifcations in bacterial strains which are not competent for DNA uptake per se but can only incorporate foreign DNA after harsh physical or physicochemical treatments (e.g. electroporation, heatshock, unphysiologically high concentrations of divalent cations etc…).[8] Natural transformation is not only responsible for the horizontal transfer of intact genes but also thought to be the primary route for the exchange of gene fragments and the formation of mosaic genes.[73]

Natural transformation is considered to be the simplest mechanism for heritable DNA acquisition in bacteria because no direct cell-to-cell contacts between donor and DNA recipient (as is the case for conjugation) or the involvement of vector systems (i.e. bacteriophages for transduction) is required.[74] Transformation is a parasexual process involving the recipient cell and free donor DNA (in contrast to the eukaryotic mating process which requires a spatio-temporally coordinated interaction between two cells). In contrast to conjugation and transduction transformation is directed entirely by the recipient cell.[3, 75] The binding of double stranded DNA to cell surface receptors (i.e. extruding pili) is followed by the transfer through the cell wall and the membranes. Upon entry into the cytoplasm one strand of DNA is degraded resulting into a single stranded DNA molecule which is the substrate for homologous recombination and final integration into the bacterial genome.[71]

For a long time it was believed that the unit of transfer in natural transformation equals to a genetic unit coding for a protein (i.e an open reading frame or a gene)[76] or at least to a protein domain.[49] It is now clear that no such restrictions exist: Protein domains are indeed units of function but not modules of transfer and homologous recombination.[49] Any kind of dsDNA - even damaged - of any length (i.e. at least in the range of 20 to several 100 000 bp) are eligible for horizontal transfer via natural transformation irrespective of representing a coding region or not.[77] It was also suspected that the incoming DNA fragment is being integrated exclusively as contiguous molecule into the receiving chromosome. But there are indications that the donor DNA fragment itself may be target for segmentation and may be integrated in a dispersed pattern into the bacterial genome.[78] This observation highlights the potential for extreme versatility of the genetic processes underlying the formation of mosaic genes providing the basis for flexible genetic variability and improved adaptability to changing environmental conditions. The most prominent and vicious examples in this field are the rapid development of resistance to antibiotics[15, 79] and the appearance of bacterial strains evading novel vaccine treatment schemes in clinical settings.[80]

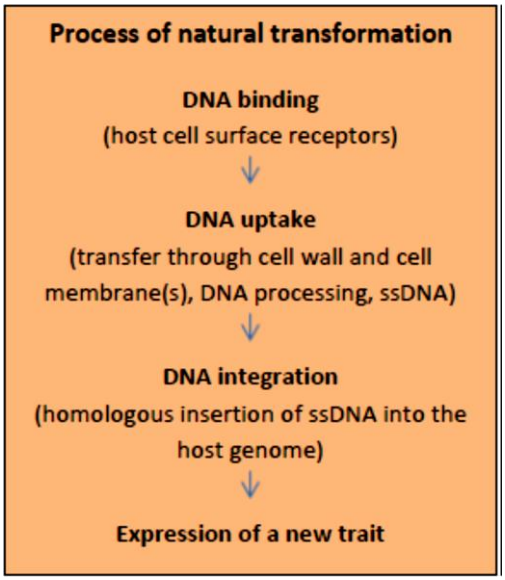
Gram positive and Gram negative bacteria share many common features in natural genetic transformation: The DNA uptake and processing machinery is astonishingly conserved over several bacterial phyla relying on similar mechanisms and involvement of homologous genes (“com regulons”).[3] However, the regulation of competence induction (i.e. starting the expression of the genes present in the com regulons) varies considerably between bacterial species. Divergent species-specific central competence regulator proteins (i.e. “master regulator of competence”) responsible for competence induction are themselves regulated by a variety of activators which allows a focused fine tuning of DNA uptake according to the prevailing growth conditions in the ecological niche and stress dependent requirements of the affected bacterial populations.[3] Competence is therefore

tightly regulated and is usually only maintained in a short window phase during growth for a limited fraction of the bacterial population. The involved regulatory circuits are extremely versatile and complex rendering natural transformation much more complex than originally thought. The positive effects of bacterial transformability is the increase in genome plasticity and environmental adaptability but DNA uptake by natural genetic transformation can also lead to deleterious mutations and to chromosomal instability of the affected genome.[74]

Most information about the mechanisms and genes involved in natural transformation were initially obtained from *Streptococcus pneumoniae* and *Bacillus subtilis* as model systems representative for the situation in Gram positive bacteria and from *Haemophilus influenzae* and *Neisseria gonorrhoeae* as representatives for Gram negative bacteria.[1, 8, 71, 81, 82] The data obtained from these model systems were supplemented by information from *S. thermophilus* and *Staphylococcus aureus* (both Gram positive) and from *Helicobacter pylori* and *Vibrio cholerae*.[3, 83-88]

Approximately 80 species so far have been reported to be naturally transformable in the most recent review on bacterial transformation in Nature Reviews Microbiology.[3] Gram positive and Gram negative bacteria appear to be equally represented in this collection although a strong selection bias (according to the research interests of the involved scientists) has to be taken into account.[89] But also the number of reported archaeal bacterial species capable for natural transformation is increasing.[90-92] It is intriguing that phylogenetically distant bacterial species share a conserved DNA uptake and processing machinery but rely on divergent central competence regulator proteins. The main biological function of natural transformation is the production of genetic diversity and chromosome repair. An additonal function for transformation as acquisitor of DNA as nutrient for the replenishment of the intracellular nucleotide pool is feasible[93, 94] but under dispute.[3, 95] An analysis of the proteins involved in competence regulation revealed that transformation is an ancient process in evolutionary terms, possibly inherited from a common ancestor.[3]

The process of natural genetic transformation



All bacteria rely on a common set of genes (i.e. the com regulons) necessary to render bacteria capable to take up free DNA from the environment (i.e. to induce competence)(Figure 5). For *Haemophilus influenzae* it has been shown that the com regulon contains 17 essential genes (of a total of 26) which are absolutely necessary for transformation.[96] In contrast the com regulon of *Bacillus subtilis* comprises of 25 essential competence genes out of a total of 100 genes present in the respective com regulon.[95] It is intriguing that some of these competence regulons contain so many genes obviously not involed in providing structural information for the assembly of the DNA uptake apparatus. The number of genes constituting a com regulon varies substantially between species and the role of the additional genes included is not clear, yet.[3]

Conserved DNA uptake

The DNA uptake machineries of Gram positive and Gram negative bacteria are very similar: both rely on the formation of DNA uptake or transformation pili (Tfp)(Figure 6). DNA uptake may be sequence specific relying upon special uptake sequences as utilized in *Neisseria gonorrhoeae* and *Haemophilus influenzae* which leads to the preferential incorporation of homospecific DNA.[97, 98] A large number of bacterial species does not use DNA uptake sequences and, thus, does not discriminate between DNA from the same species or foreign DNA as is the case for instance with *Streptococcus pneumoniae, Bacillus subtilis* and *Acinetobacter baylyi*.[71]

Gram positive bacteria form transformation pili mainly consisting of comGC subunits (“competence pseudopili”)(Figure 6). Gram negative bacteria build pilE containing pili which belong to the category of type IV pili (T4P) and are functionally related to the type II secretion system. Both pili structures convey the incoming DNA to the primary DNA receptor comEA (Gram positive) and comE (Gram negative). As Gram negative bacteria posess an outer membrane they additionally use a secretin channel comprising of pilQ elements which facilitates the passing of double stranded DNA through this lipid bilayer (Figure 8).

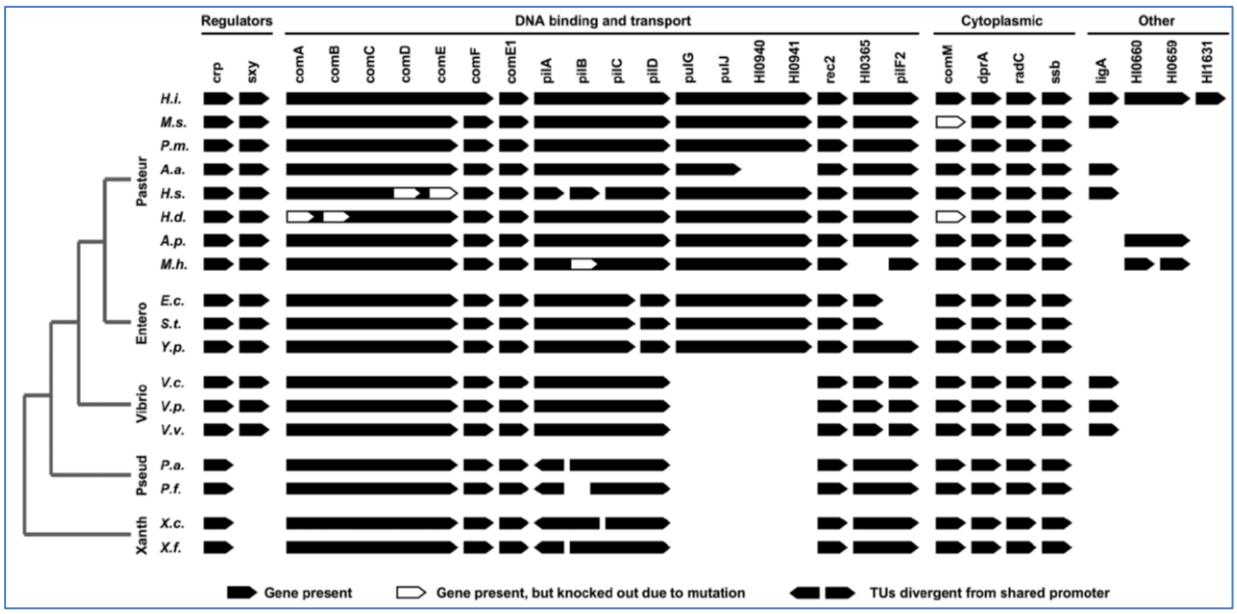


Figure 5: Competence regulons in Gram negative and Gram positive bacteria.

Solid lines depict transcriptional units (gene lengths not to scale). Pasteur, *Pasteurellaceae*; Entero, *Enterobacteriaceae*; Vibrio, *Vibrionaceae*; Pseud, *Pseudomonadaceae*; Xanth, *Xanthomonadaceae*; H.i., *H. influenzae*; M.s., *M. succiniciproducens*; P.m., *P. multocida*; A.a., *A. actinomycetemcomitans*; H.s., *H. somnus*; A.p., *A. pleuropneumoniae*; M.h., *M. haemolytica*; E.c. *E. coli*; S.t., *S. typhimurium*; Y.p., *Y. pestis*; V.c., *V. cholerae*; V.p., *V. parahaemolyticus*; V.v., *V. vulnificus*; P.a., *P. aeruginosa*; P.f., *P. fluorescens*; X.c., *X. campestris*; X.f., *X. fastidiosa* (data from Cameron et al.[99]).

Conserved DNA processing

Double stranded DNA is converted into a single stranded form by an endonuclease (endA) in *Streptococcus pneumoniae* and transferred via the comEC channel (Gram positive) or comA (Gram negative) into the cytoplasm (Figure 8). For Gram negative bacteria a special endonuclease enzyme is still hypothetical. For Gram positive bacteria it could be shown that the transfer of ssDNA through the channel is a polar process in 3' → 5' direction. In the cytoplasm the single stranded DNA is covered immediately by ssDNA binding proteins like DprA, SsbB and RecA and protected from degradation.[3] DprA has no other function than ssDNA protection and guiding the loading of this single stranded DNA with recA. DprA is a transformation specific protein and may, thus, be used as a signature for transformability in *in silico* genetic screens.[3] RecA is the core protein of homologous recombination by effecting the search for homologus regions between incoming single stranded DNA and the target genome.[100, 101] Homologous recombination is initiated by ssDNA in all organisms.[102] For an efficient functioning of RecA cofactors like DprA are necessary.[101]

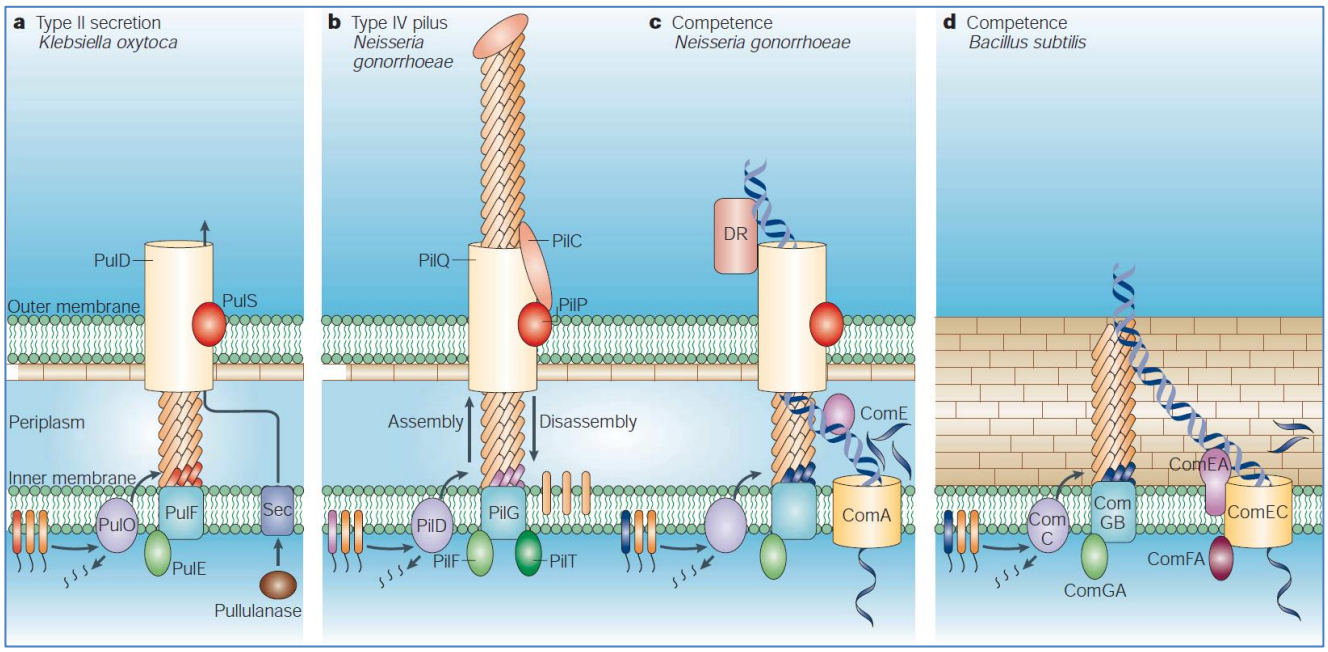


Figure 6: Comparison of type II secretion, type IV pilus formation and transformation.

a) A schematic model for type II secretion, based on the pullulanase secretion system (Pul) from *Klebsiella oxytoca*. The pseudopilins, both major (PulG; orange) and minor (PulH,-I,-J and -K; red), are processed by the prepilin peptidase (PulO), and assembled into the pseudopilus. **b)** A schematic model for type IV pilus formation, based on the *Neisseria gonorrhoeae* pilus. The major pilin (PilE; orange) and minor pilin (PilV; magenta) are processed by the prepilin peptidase (PilD). **c)** A schematic model for the competence pseudopilus and DNA translocase in *N. gonorrhoeae*. Assembly of the pseudopilus requires the same components as the type IV pilus (shown in part **b**). The major pilin (PilE; orange) and minor pilin (ComP; blue) are processed by the prepilin peptidase (PilD), and assembled into the pseudopilus. The polytopic membrane protein (PilG) and the traffic NTPase (PilF) participate in this process, as well as PilC (not shown). The specific sequence in the exogenous DNA that is required for efficient uptake is recognized by its postulated, but as-yet-unidentified, receptor (DR). The incoming DNA is transported across the outer membrane through a channel that is formed by the secretin (PilQ), with the assistance of its pilot protein (PilP). The periplasmic DNA-binding protein (ComE) is involved in uptake, and delivers the DNA to the channel at the cytoplasmic membrane (ComA). One strand enters the cytosol; the other is degraded and the degradation products are released into the periplasmic space. **d)** A schematic model for the competence pseudopilus and DNA translocase in *Bacillus subtilis*. The major pseudopilin (ComGC; orange) and minor pseudopilins (ComGD, -GE and -GG; blue) are processed by the prepilin peptidase (ComC), and assembled into the pseudopilus. The polytopic membrane protein (ComGB) and the traffic NTPase (ComGA) participate in this process. The pseudopilus allows the exogenous DNA to access its membrane-bound receptor (ComEA), which delivers the bound DNA to the channel at the cytoplasmic membrane (ComEC). An ATP-binding protein (ComFA) is involved in DNA transport across the membrane. One strand enters the cytosol, while the other is degraded and the degradation products are released into the extracellular milieu (data from Chen et al.[1]).

Divergent central competence regulators

Although DNA uptake and processing relies on an astonishingly similar arsenal of competence genes the expression of these com regulons is controlled by a variety of evolutionary different central competence regulators. Regulation occurs during specific bacterial growth phases, is essentially species-specific and varies considerably among transformable species.

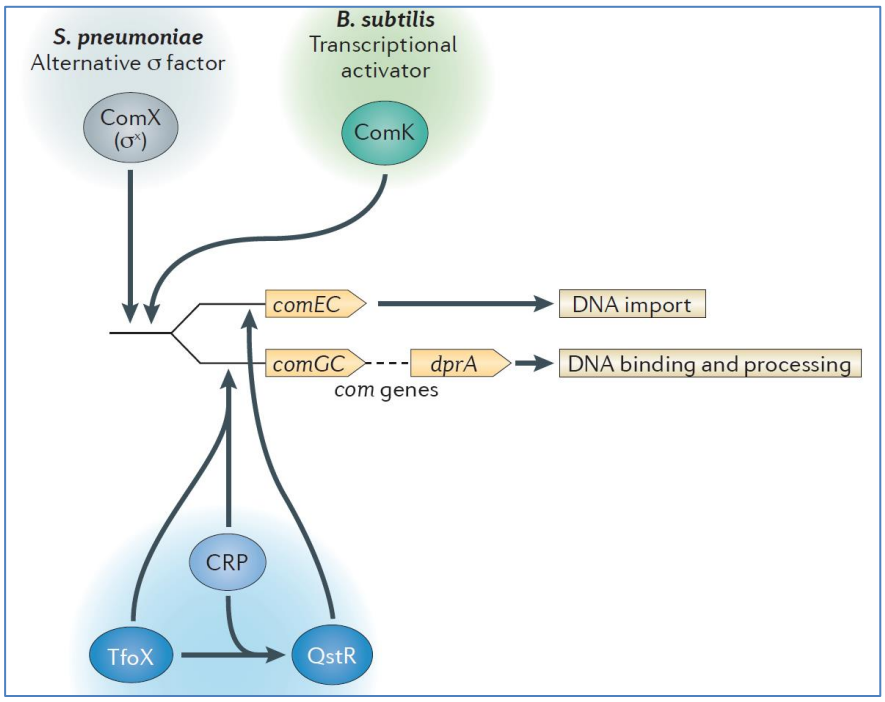


Figure 7: Regulation of comptence by specific central competence regulators.

Competence genes can be regulated by alternative σ factors (for example, σ^X in *Streptococcus pneumoniae*), transcription factors (such as ComK in *Bacillus subtilis*) and transcription co-regulators (such as the cAMP receptor protein (CRP) cofactor TfoX in *Vibrio cholerae*). Although TfoX directly regulates most of the *com* genes in *V. cholerae*, it indirectly activates *comEA* and *comEC* by regulating the expression of *qstR*[85] (data from Johnston et al.[3]).

The best studied central competence regulators to date are alternative sigma factors like Sigma^X (*S. pneumoniae*)[103], transcription activators like ComK (*Bacillus subtilis*)[104] or transcription co-regulators like Sxy (*H. influenzae*)[105] and TfoX (*Vibrio cholerae*)[106](Figure 7). These central competence regulators are themselves controlled by species-specific regulatory cascades responding to a plethora of environmental stimuli like nutrient depletion, changes in pH or temperature, DNA damage or the presence of certain competence inducers like chitin. But they are also responding to endogenous triggers like autoinducers and pheromone-like peptides like CSP which is involved in cell-to-cell signalling in *S. pneumoniae* and induces competence in neighbouring cells.

Chromosomal DNA integration - Homologous recombination

The recombinase RecA, which is recruited by the DNA processing protein A DprA, covers the incoming single DNA strand, executes homology search along the chromosomal DNA and catalyzes DNA strand exchange (Figure 8). The resulting heteroduplex DNA can be a completely homologous recombination intermediate which eventually leads to the replacement of an existing chromosomal region or the insert consists of a heterologous core sequence flanked by sequences homologous to the chromosoma target region. Resolution of the recombination complex leads in this case to the addition of new DNA. The new incoming DNA is usually unmethylated and target for the cellular restriction – modification system. Unmethylated DNA is prone for degradation by restriction endonucleases, which may kill transformants. This mechanism may restrict heterologous transformation.[3]

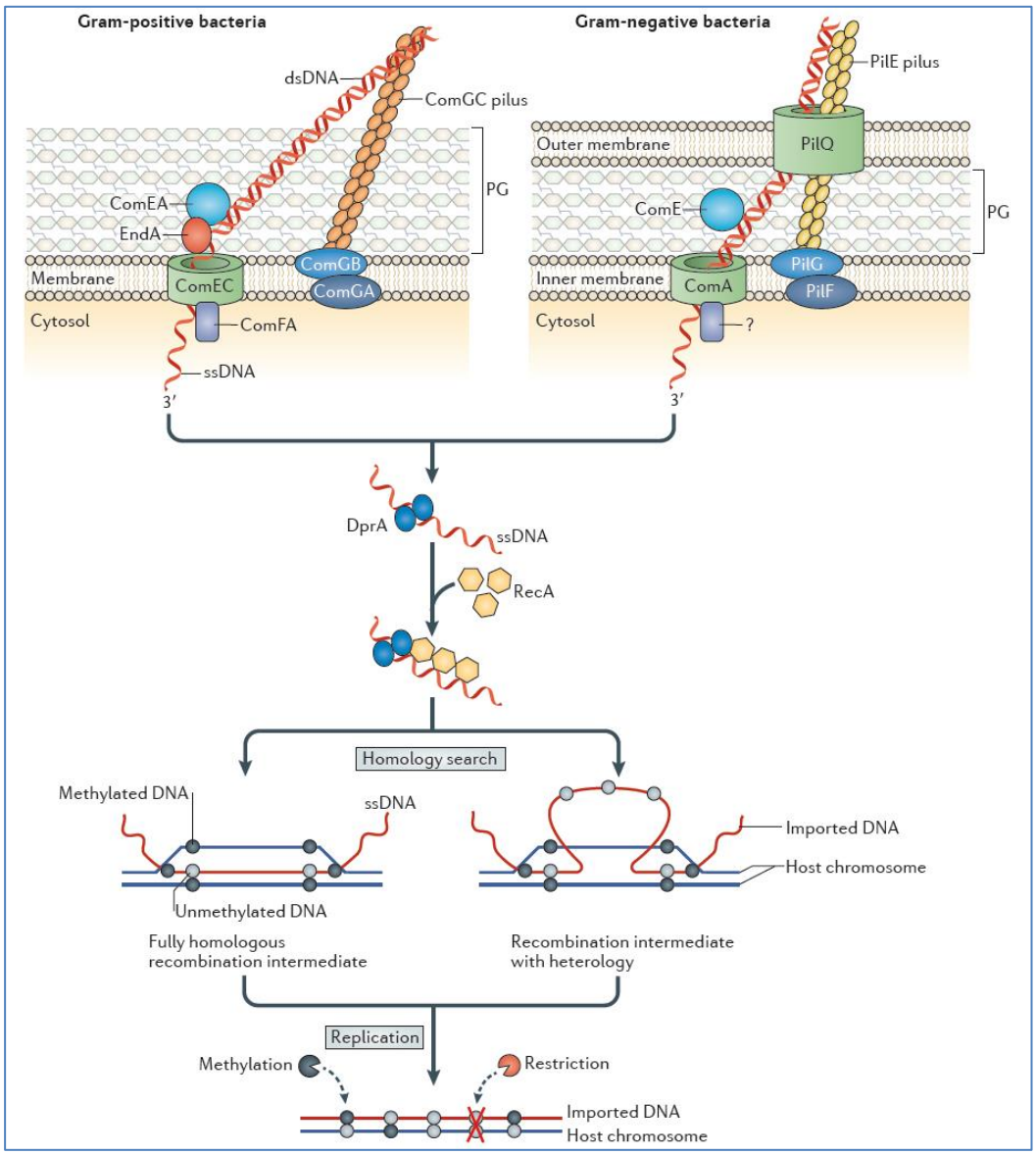


Figure 8: Key steps of the transformation process.

The DNA-uptake machinery generally comprises a **transformation pilus (Tfp)**, which consists mainly of **ComGC** subunits in Gram-positive bacteria and captures exogenous double-stranded DNA (dsDNA), the DNA receptor **ComEA** and the transmembrane pore **ComEC**. In *Streptococcus pneumoniae*, the **EndA nuclease** receives DNA from the DNA receptor ComEA and degrades one DNA strand, whereas unidentified nucleases (or strand-separating proteins) generate single-stranded DNA (ssDNA) for uptake in other species. In Firmicutes, ssDNA internalization through ComEC is presumably driven by the ATP-dependent **translocase ComFA**. In Gram-negative bacteria, such as *Neiserria gonorrhoeae*, the **PilQ** secretin channel enables the pilus (which is mainly composed of **PilE** subunits) to cross the outer membrane and dsDNA is transported across the outer membrane through PilQ. In both Gram-positive and Gram-negative cells, additional proteins are required for DNA uptake (for example, the ComGA and ComGB proteins of Firmicutes). A homologue of the ComFA translocase might be present in Gram-negative bacteria, but this is currently unclear. Internalized ssDNA is presumably bound by **DprA (**DNA processing protein A), which recruits the **recombinase RecA.** RecA polymerizes on ssDNA and promotes a homology search along chromosomal DNA, followed by strand exchange. The transformation heteroduplex that forms can be a **fully homologous** double-stranded recombination intermediate, or if the imported DNA contains **heterologous sequences** (such as a pathogenicity island) **flanked by homology**, a recombination intermediate with a single-stranded loop is formed. If heterologous donor DNA is unmethylated (light grey circles), this DNA remains fully unmethylated in the recipient chromosome after replication. The methylation and restriction activities of the restriction–modification (R–M) system compete (dashed arrows) for access to this sensitive DNA, and restriction can kill transformants and limit heterologous transformation. PG, peptidoglycan (data from Johnston et al.[3]).

Biological function of imported DNA

The evolutionary impetus for DNA import by bacteria may be for nutritional reasons, for genome maintenance and/or for creating genomic diversity.[1] It is clear that competence provides the machinery for genetic exchange and, thus is an essential factor for genome plasticity and genome repair.[84] Although it was proposed that the uptake of DNA as "food" provides nutritional benefits, which would generate sufficient selective advantage for the maintenance of the DNA uptake machinery in evolutionary terms[93], the arguments in support of this hypothesis are inconclusive: Digestion of internalized DNA into nucleotides has been documented in *H. influenzae*[107] and *S. pneumoniae*[108], however, the selective internalization of self-species DNA in *H. influenzae*[109] and the induction of DNA release only from other pneumococci and closely related streptococci via fratricide in *S. pneumoniae*[110-112] are inconsistent with the import of DNA for catabolism. Moreover, the intruding ssDNA molecule is actively protected from endogenous nucleases by single strand binding proteins like DprA[101], SsbB[113] and recA[101]. Direct evidence for the "DNA for genetic diversity"-hypothesis derives from analysis of the streptococcal DpnA/DpnII restriction modification system which shows that the uptake of heterologous DNA - like pathogenicity islands - is dependent on the methylation of ssDNA and promoted by DpnA.[114] As heterologous DNA is of no use for genome maintenance and repair imported DNA is thus primarily used for generating genetic diversity in *S. pneumoniae*.[114, 115]

Purposes of DNA transformation

DNA for genetic diversity: acquisition of potential useful genetic information (novel metabolic functions, virulence traits or antibiotic resistance)[1]

DNA repair: environmental DNA from closely related bacteria might serve as template for the repair of DNA damage.[1]

DNA as food: source of carbon, nitrogen, phosphorous[1]

Accidental by-product of bacterial adhesion and twitching motility[2]

The majority of naturally transformable bacterial strains have been detected in the oropharynx and the upper respiratory tract of mammals (e.g. *Actinobacillus* spp., *Aggregatibacter* spp., *Cardiobacterium* spp., *Haemophilus* spp., *Kingella* spp., *Moraxella* spp., *Neisseria* spp., *Porphyromonas* spp., *Streptococcus* spp., *Sutonella* spp., *Tanerella* spp.). Several strains used in the food industry for milk fermentation, youghurt and cheese production have been described to be capable for the uptake of exogenous DNA (e.g. *Lactobacillus spp.*).[116, 117] It is also noteworthy that bacteria living in extreme habitats (hot springs, vents, arctic ice/soil) are frequently found to be highly transformable with exogenous DNA (*Chlorobium* spp., *Chlorobaculum* spp., *Deinococcus* spp., *Methanothermobacter* spp., *Nostoc* spp., *Psychrobacter* spp., *Pyrococcus* spp., *Thermoanaerobacter* spp., *Thermococcus* spp., *Thermosynechococcus* spp., *Thermotoga* spp., *Thermus* spp., *Thiobacillus sp*.)(Table 1).

1.6.6 Induction of Competence

The term competence describes a physiological condition of bacteria which are primed for the uptake and chromosomal integration of exogenous DNA. Competence comprises the coordinated expression of a set of proteins necessary for the production of a multiprotein complex - the transformasome - traversing the bacterial cell wall and membrane(s) and providing the apparatus for extracellular DNA binding, transport and processing.[1] The involved proteins belong to the family of type II secetion system proteins and produce pilus structures similar to type IV (pseudo-) pili and are usually not permanently expressed. [74, 118] The process of competence induction is tightly regulated

usually comprising a short transient "competence window" during the growth of bacterial cells and is in general a response to species-specific environmental conditions triggering the activation of a master regulator of competence (e.g. altered sigma factor σ^X in *S. pneumoniae*, comK in *Bacillus subtilis*, or TfoX in *Vibrio cholerae*). However, the activity of these master regulators of competence is again regulated by a species-specific plethora of transcriptional (e.g. comCDE, ComRS cell-to-cell signaling pathways) or post-transcriptional (e.g. Clp proteolytic system) mechanisms.[74]

The major inducers of competence are environmental stimuli like availability of certain carbon sources, stress conditions imposed by starvation or exposure to genotoxic compounds or cell-to-cell signals sensing the global status of bacterial communities.[84]

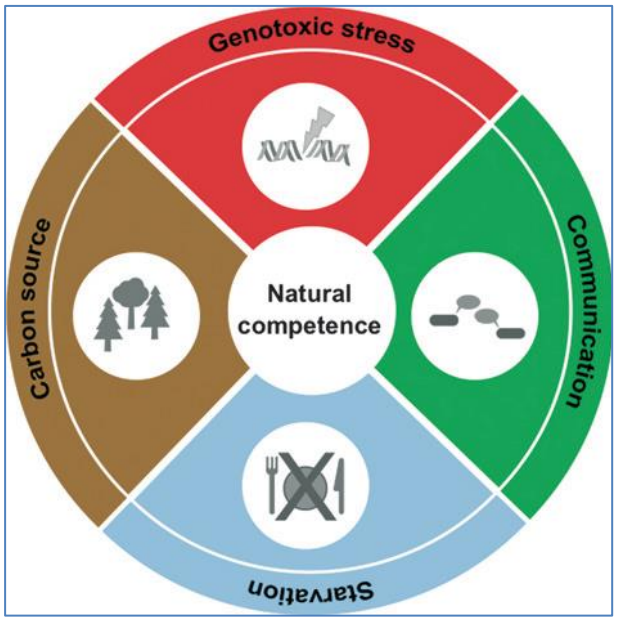


Figure 9: Environmental cues involved in competence induction and the fine-tuning of competence.

These signals include genotoxic stresses causing DNA damage, such as UV light or certain antibiotics (in red); bacterial cell-cell communication systems (e.g. quorum-sensing; in green); the starvation of preferred carbon sources, leading to the accumulation of the intracellular secondary messenger cAMP (carbon catabolite repression; in blue); and the presence or absence of certain carbon sources (such as the GlcNAc polymer chitin for *Vibrio cholerae*) (data from Seitz et al.[84]).

Competence is usually limited to a proportion of the total bacterial population.[3] Approximately 1% of the cells of a *Streptoccoccus mutans* population are actually primed for DNA uptake after induction of competence. The same is true for 10% of *Streptococcus thermophilus* or 15% of *Bacillus subtilis* cells in culture.[3] However, 100% of *Streptococcus pneumoniae* become competent after induction.[74] Competence is conserved in at least six different phyla and an old pathway in evolutionary terms.[119] Only for *Neisseria* spp. it is reported that a constitutively expressed competence status is reported.[8] *Helicobacter pylori* is also assumed to be continuously competent during its whole life cycle but the evidence is inconclusive.[120, 121]

Competence regulation

Induction and overall complexity of competence regulation is exemplarily demonstrated with the processes active in Gram positive streptococci.[74] Competence in *S. pneumoniae* is regulated in two phases:

During the early phase of competence induction regulation of the process is executed on three different layers via i) proximal transcription, ii) distal transcription and iii) postranslational control.

Early phase – proximal transcription control

The early phase of comptence induction is initiated by an intracellular increase of ComX (i.e. the altered sigma factor σ^X) which is triggered transcriptionally by competence pheromones i.e. auto-inducers responsible for cell-to-cell communication. This system is activated at certain cell densities (i.e. via quorum sensing), at certain conditions of the surrounding medium (i.e. diffusion sensing), or by the overall spatial organization of the cells in the habitat. This regulatory cascade forms an "alarmone" comparable to the SOS response system in *E. coli* which is peripherally activated by environmental stress signals like antibiotics, mutagens, acidic, oxidative or temperature stress or nutritional signals. The direct expression control of comX is mediated by two-component regulatory systems (TCS) like comCDE consisting of a pheromone precursor protein (ComC), a sensor (ComD) for the mature ComC protein (CSP) and a transcription factor (ComE) in the case of *S. pneumoniae*, *S. mitis* and *S. anguinosus*. In the case of *S. bovis, S. mutans*, and *S. pyogenes* the comRS system consisting of a pheromone precursor (ComS) and a cytoplasmic transcriptional activator (comR) execute the proximal – transcription – control of comX.

Early phase – distal transcription control

The expression of ComX is additionally regulated by fluctuations of environmental parameters sensed by different regulatory systems with connections to the competence regulons. A prominent example for this is the *blp*-cluster coding for bacteriocins in *S. pneumoniae* located in close proximity to the com regulon on the chromosome. Bacteriocin production has an impact on competence induction by co-activation of the com regulon.

Cell surface integrity sensors are interlinked with competence regulons because genome replication and transformation must be coordinated to ensure a correct chromosome rearrangement following homologous recombination with incoming DNA. The mediators of this kind of regulation of ComX expression are streptococcal two-component systems of the PhoB/OmpR family and serin/threonine kinases of the StkP/PknB type. These TCS respond typically to the following abiotic stimuli: pH, oxygen supply, salt, divalent cation concentrations, antibiotics and antimicrobial peptides targeting cell wall components or disrupting the proton motive force.

Agents inducing replication stress like fluoquinolones, mitomycin C, hydroxylurea and UV radiation are potent activators of ComX expression. Nitrogen sensing systems like CodY have been also shown to be involved in the regulation of ComX.

Early phase – post-translational control

Competence induction is additionally regulated over the stability of the master regulators of competence. In the presented case ComX may be actively stabilized by additional proteins or may be target to proteolytic degradation by the Clp proteolytic system of streptococci.

Late phase

During the late phase of competence induction ComX associates with the core of RNA polymerase II to target especially the CIN box promoter element regulating the expression of competence related proteins. In streptococci 14 essential late competence encode the transformasome. The com regulon contains additional genes coding for ssDNA methylation and lysis (fratricide) functions. The expression of these genes is simultaneously regulated and all genes are present in all streptococci implying that all streptococci may at least have the potential to become naturally transformable.[95, 122]

Competence regulons in Gram negative and Gram positive bacteria

It is intriguing that genes responsible for bacterial competence (com regulons) are present in all gamma-proteobacteria (including *Enterobacteriaceae* and in the usually as naturally non-transformable classified species *E. coli*)[99]. The signature of com genes is found in several bacterial phyla. However, it remains to be demonstrated that these genes are also functional in their carrier species.

Literature analysis revealed only a minute number of bacterial species to date actually tested for their competence to actively take up DNA and express a new phenotype. A recent publication in Nature Reviews Microbiology lists only 82 bacterial species to be naturally transformable.[3] A major obstacle in this respect is the identification of proper culture conditions to induce competence *in vitro*. It is reasonable to assume that for many bacterial species these proper conditions simply have not been found, yet.[95]

Biological functions of competence

Although competence is generally viewed as the condition of bacteria primed to take up exogenous DNA there are several competence dependent phenomena which do not involve uptake of DNA: Competence per se enhances survival of pneumococci under stress conditions and promotes resistance to UV radiation and acts as a substitute of the SOS repair system active in other bacteria but not present in pneumococci.[123] Also *Legionella pneumophila* lacks a SOS response but becomes competent upon DNA damage.[124] This observation is in support of the hypothesis that competence has evolved as a globas stress response in bacteria with no functional SOS response.[123] But not all transformable species lack an SOS response system. *Bacillus subtilis, Vibrio cholerae* and *S. thermophilus* have maintained a functional SOS and the competence regulon.[3]

It is interesting that a bulk of information is available about the mechanism of competence induction but there is substantially less data accessible dealing with the shut-down of competence although it has been known for a long period of time that in nearly all bacteria the competent state is a tightly regulated window in the life cycle of the cells and bacterial populations sometimes rather quickly lose the capability for DNA uptake during growth.[3] There are indications that DprA plays a decisive role by repressing transcription of central competence genes (like comX).

1.6.7 Naturally Transformable Bacterial Species

Approximately 80 species so far have been reported to be naturally transformable in the most recent review on bacterial transformation in Nature Reviews Microbiology in 2014.[3] Gram positive and Gram negative bacteria appear to be equally represented in this collection although a strong selection bias (according to the research interests of the involved scientists) has to be taken into account.[89] But also the number of reported archaeal bacterial species capable for natural transformation is increasing.[90-92]

The table below presents data on natural transformability, transformation frequencies, and conditions for induction of competence, the environmental habitat, and the status as pathogen for more than 130 bacterial species experimentally proven to be naturally compentent for the uptake of free DNA under distinct environmental conditions (Table 1). This is an increase in number of approximately 50 taxa since the last review on this topic as published in Nature Reviews Microbiology in 2014.[3] There are indications that many more bacterial species may become competent under certain environmental conditions because the genetic signature for functional competence genes is present in several bacterial phyla/classes (e.g. all γ-proteobacteria).[99] For many of these strains the conditions for the induction of competence simply have not been discovered in laboratory experiments, yet. The phylogenetic relationship of all bacterial genera shown to contain naturally transformable species is depicted in Figure 10.

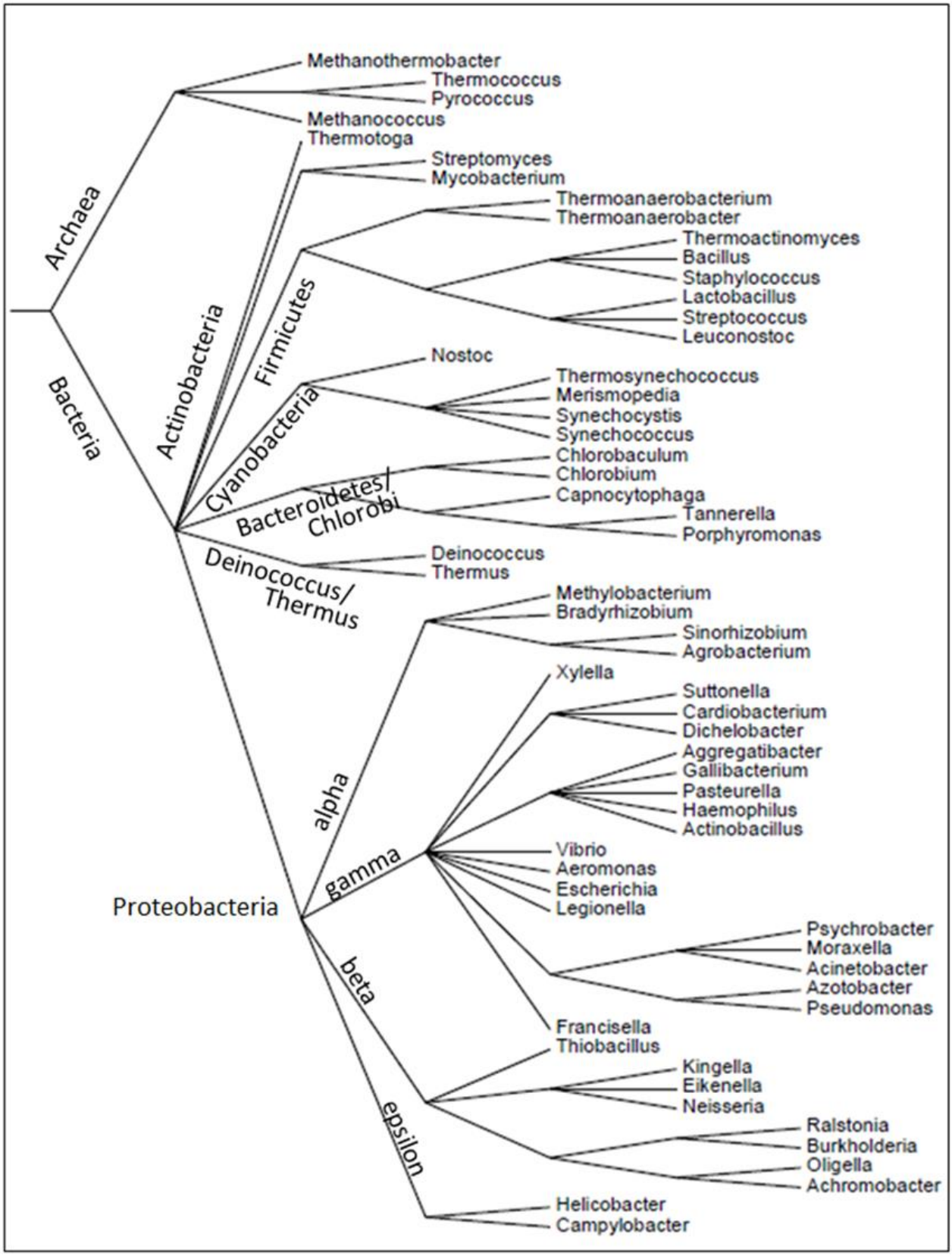


Figure 10: Phylogentic relationship of naturally transformable bacteria.

Table 1. Naturally transformable bacterial species.

[a)] Competence genes present, but no explicit experimental evidence available in support of natural transformability; 1st, 2nd primary and secondary literature references

	Species	Class	Environment	Competence	References	
					1st	2nd
1	*Achromobacter* spp.	β-Proteobacteria	heterotrophic (rare pathogen); aerobic;(sea) water; fish, poultry, foods	Plate transformation[125]; no additional information available[125]	92	8
2	*Acinetobacter baumannii*	γ-Proteobacteria	pathogen (human), multi resistant nosocomial; skin	DNA uptake during movement on wet surfaces, type 4 pilus dependent (pilT), complete set of competence genes present; comA; planktonic cells not naturally competent; no competence induction under *A. baylyi* conditions; transformation frequency: $3 \times 10^{-3} - 6 \times 10^{-8}$	126	
3	*Acinetobacter baylyi* *(old: Acinetobacter* sp. *ADP 1 (=BD413))*	γ-Proteobacteria	heterotrophic; soil	Induction: High nutrient levels[127] Early logarithmic phase and for several hours in the stationary phase, 25% of population (liquid culture) transformable, no sequence specificity, divalent cation-dependent[128], no competence induction in soil[127, 129]	92	8, 130
4	*Acinetobacter calcoaceticus*	γ-Proteobacteria	heterotrophic; soil	See *A.baylyi.* Early to late log-phase[8]; DNA uptake in natural environments (soil, sediment, groundwater)[131], 10 - 25% of population competent, complex and minimal medium[128] Transformation frequency: 7.0×10^{-3} [8]	92	8, 130
5	*Actinobacillus actinomycetemcomitans* *(old: Haemophilus actinomycetemcomitans)*	γ-Proteobacteria	pathogen (human, facultative) periodontis, endocarditis; commensal oral cavity	Some clinical isolates naturally competent Transformation frequency: $10^{-3} - 10^{-5}$) [132]	92	130
6	*Actinobacillus pleuropneumoniae*	γ-Proteobacteria	heterotrophic; pathogen (pig), respiratory tract	Induction: nutrient starvation (transfer to MIV medium) or entry into stationary phase; intact set of competence genes, competence activator: Sxy[133] (plate transformation) Transformation frequency low: $<10^{-8}$ (single strain: 10^{-4})	89	133, 134
7	*Actinobacillus suis*	γ-Proteobacteria	beta-haemolytic, opportunistic pathogen (animal)	Induction: nutrient starvation[94], CRP-Sxy dependent, purine nucleotide depletion; compare *H. influenzae*	3	94
8	*Actinobacillus ureae* *(old: Pasteurella ureae)*	γ-Proteobacteria	pathogen (human) sinusitis, bronchopneumonia; commensal nasopharynx	Competence genes present, comEA,EC,M, pilQ	92	130

	Species	Class	Environment	Competence	References	
					1st	2nd
9	*Aeromonas salmonicida* *(Proteus salmonicida)*	γ-Proteobacteria	non-pathogenic (human); fish furunculosis; river sediments; aquatic environments	Incubation: 20°C for 20 day, humic acids, tryptone; induction: nutrient limitation, no DNA sequence specificity, transformation frequency not determined	135	
10	*Aeromonas* spp.	γ-Proteobacteria	pathogen (human) and non-pathogenic environmental isolates; gastrointestinal and extra-intestinal infections; aquatic environments, ARG carrier	(Late) stationary phase, pH 5 - 8, divalent cations, dilute growth medium (20% nutrient broth), 30°C; 73% of the population transformable Transformation frequency: 10^{-3}	136	
11	*Agmenellum quadruplicitum* *(Synechococcus sp. PCC 7002)*	Cyanobacteria	photolithotrophic; marine	Exponential growth - decrease in stationary phase or nitrogen deprivation, increase by lowering growth temperature from 39°C to 30°C[137], transformation frequency: 4,3 x 10^{-4}; exclusion from light resulted in loss of competence[8]	92	8, 138
12	*Aggregatibacter aphrophilus* *(old: Haemophilus aphrophilus)*	γ-Proteobacteria	pathogen (human, facultative) oral commensal, gingival crevices, dental plaque, rare infections (endocarditis, brain abscesses)	Some clinical isolates naturally competent Transformation frequency: 10^{-3} – 10^{-4}) [132]	89	132
13	*Agrobacterium tumefaciens*	α-Proteobacteria	plant pathogen; soil	Soil (sterile + non-sterile) microcosm experiments, transformation frequency: 10^{-8} – 10^{-9} induction conditions unknown[129]	89	129
14	*Anacystis nidulans* *(new: Synechococcus elongatus strain PCC7942)*	Cyanobacteria	photolithotrophic; freshwater	Early to late log-phase, low iron concentrations stimulates competence; exclusion from light resulted in loss of competence Transformation frequency: 8.0 x 10^{-4}) [8]	92	8
15	*Azotobacter vinelandii*	γ-Proteobacteria	heterotrophic; soil	Early to late log-phase, up to 100% of the population competent; nutrient limitation (minimal medium): high transformation frequencies (10^{-2} – 10^{-3}), no transformation in complex media, high available iron concentrations: low competence, Ca^{2+} dependent; peak at pH 7, 26°C - 37°C[8]	92	8
16	*Bacillus amyloliquefaciens*	Bacilli	heterotrophic, soil	Transformation frequency: 10^{-4}; Ca^{2+} dependent, nutrient limitation (minimal glucose)[139]	89	139
17	*Bacillus licheniformis*	Bacilli	heterotrophic, low GC Gram+; soil	Transformation frequency: 1.2 x 10^{-2}	92	8

	Species	Class	Environment	Competence	References	
					1st	2nd
18	*Bacillus mojavensis*	Bacilli	heterotrophic, low GC Gram+; desert soil	See *B. licheniformis* resistance to genetic transformation between taxa	92	140, 141
19	*Bacillus stearothermophilus*	Bacilli	thermophilic, soil, hot springs, ocean sediments; food poisoning	Transition from log phase to stationary phase[8]	3	
20	*Bacillus subtilis*	Bacilli	Heterotrophic, low GC Gram+; soil	Transition from log phase to stationary phase[8] - restricted to stationary phase[3], nutritional exhaustion[95], 10 - 25% of population competent[8], complex media repress competence, transformation frequency: 3.5×10^{-2} [8]	92	8
21	*Bradyrhizobium japonicum*	β-Proteobacteria	heterotrophic; plant symbiont, soil	No details on competence development available	89	142
22	*Burkholderia pseudomallei*	β-Proteobacteria	pathogen (human); melioidosis, bioterrorism agent	Minimal medium, early logarithmic growth phase; compare to *Ralstonia solanceae*; transformation frequency: $10^{-7} – 10^{-9}$ [143]	144	
23	*Burkholderia thailandensis*	β-Proteobacteria	relatively non-pathogenic; soil	See *Burkholderia pseudomallei*	144	
24	*Campylobacter coli*	ε-Proteobacteria	pathogen (human) gastroenteritis; microaerophilic, gut, alternative transformation mechanism	Early log phase, preferential uptake of species-specific DNA, plate transformations (Mueller-Hinton broth)[145] Transformation frequency: 1.2×10^{-3}	92	3, 8
25	*Campylobacter jejuni*	ε-Proteobacteria	pathogen (human); gut, alternative transformation mechanism	Early log phase, preferential uptake of species-specific DNA[145],transformation frequency: 2.0×10^{-4} [8]	92	3, 8, 130
26	*Capnocytophaga canimorsus*	Flavobacteria	oral cavity (dog; non-pathogenic)	No data on competence accessible.	92	130
27	*Cardiobacterium hominis (CDC EF-4 B)*	γ-Proteobacteria	oropharynx (human; non-pathogenic)	No data on competence accessible.	92	130
28	*Chlorobium limicola*	Chlorobia	photolithotrophic, thermophile; hot springs	Transition from log phase to stationary phase Transformation frequency: 1.0×10^{-5} [8]	92	8

	Species	Class	Environment	Competence	References	
					1st	2nd
29	*Chlorobaculum tepidum*	Chlorobia	photolithotrophic, thermophile, green sulfur bacterium; hot springs	No data on competence accessible.	92	146
30	*Deinococcus radiodurans* *(Micrococcus radiodurans)*	Deinococci	heterotrophic, extremophile, ubiquitary (+ human gut)	Competence proceeds throughout exponential growth and declines during the stationary phase, transformation frequency: 2.1×10^{-2} [8]	92	8
31	*Dichelobacter nodosus*	γ-Proteobacteria	compare: Cardiobacterium	No data on competence accessible.	92	147
32	*Eikenella corrodens*	β-Proteobacteria	oropharynx, respiratory tract, commensal flora (rare infections)	No data on competence accessible.	92	130
33	*Escherichia coli*	γ-Proteobacteria	pathogenic strains; commensal gastrointestinal tract	No data on competence accessible.	148	
34	*Francisella* spp.[a)]	γ-Proteobacteria	pathogen (human)	Competence genes have been described. No experimental transformation data available.	149	
35	*Gallibacterium anatis* *(Pasteurella anatis)*	γ-Proteobacteria	pathogen (animal; poultry); commensal avian hosts	Competence induced by starvation. Tranformation frequency: >2 $\times 10^{-4}$.	150	
36	*Haemophilus influenzae*	γ-Proteobacteria	pathogen (human); oropharynx, respiratory tract	Entry into stationary phase[105] or nutrient starvation/limitation[151] modulated by nucleotide pools[94], CRP/Sxy-dependent[152] up to 100% of the population competent; Transformation frequency: 7.0×10^{-3} [8]	92	8, 130
37	*Haemophilus parainfluenzae*	γ-Proteobacteria	pathogen (human), rare infections	Transformation frequency: 8.6×10^{-3}	92	8, 130
38	*Haemophilus parasuis*	γ-Proteobacteria	pathogen (pig); commensal mucosa	No data on competence accessible.	92	130
39	*Helicobacter pylori*	ε-Proteobacteria	pathogen (human); stomach, alternative transformation mechanism	Constitutively competent (?)[3] Transformation frequency: 5.0×10^{-4}	92	3, 8, 130

	Species	Class	Environment	Competence	References	
					1st	2nd
40	*Kingella dentrificans*	β-Proteobacteria	compare *Neisseria*; commensal mucous membranes respiratory tract; opportunistic pathogen	No data on competence accessible.	92	130
41	*Kingella kingae*	β-Proteobacteria	compare Neisseria; commensal mucous membranes respiratory tract; opportunistic pathogen	No data on competence accessible.	92	130
42	*Lactobacillus delbrueckii subsp. lactis* *(old: Lactobacillus lactis)*	Bacilli	low GC Gram+, yoghourt, milk, anaerobic/aerotolerant; commensal	No data on competence accessible.	92	8
43	*Lactobacillus lactis*	Bacilli	heterotrophic, buttermilk, cheese; opportunistic pathogen: few reports; habitat: dairy	Transformation frequency: 2.3×10^{-5}	89	8, 116
44	*Lactobacillus plantarum*[a)]	Bacilli	fermented food products, decaying plant material; aerotolerant	Competence genes detected; no experimental evidence.	153	
45	*Lactobacillus sakei*	Bacilli	non-sporulating, meat-borne, anaerobic, heterofermentative	Competence genes and transcription factor detected; no experimental evidence.	154	
46	*Legionella pneumophila*	γ-Proteobacteria	pathogen (human); obligate aerobic, soil, water	Induction by aminoglycosides, fluoroquinolone antibiotics, mitomycin C, UV [124]	92	155
47	*Leuconostoc carnosum*	Bacilli	compare: lactic acid bacteria; milk, plants, anaerobic/aerotolerant; non-pathogenic,	Competence in early exponential phase during preparation for electrotransformation. Transformation frequency in absence of electrical field 3×10^{-6} and 19×10^{-6}	89	156
48	*Methanococcus voltae*	Methanococci	methane producer; anaerobic, GI tract animals, freshwater and marine sediments, sewage; non-pathogenic	Transformation frequency: 8.0×10^{-6} DNA packaged 4,4 kb	92	8
49	*Methanothermobacter thermoautotrophicus* *(Methanobacterium thermoautotrophicum)*	Methanobacteria	methane producer; sewage, thermophile (40°-70°C)	No data on competence accessible.	92	8

	Species	Class	Environment	Competence	References	
					1st	2nd
50	*Methylobacterium organophilum*	α-Proteobacteria	facultative methylotrophic + heterotrophic; compare Rhizobiales; soil, plant (+ surface), fuel/water interfaces -> biofilm	Transition from log phase to stationary phase Transformation frequency: 5.3×10^{-3}) [8]	92	8
51	*Moraxella catarrhalis* *(Branhamella catarrhalis)*	γ-Proteobacteria	pathogen (human); upper respiratory tract	Natural transformation during preparation for electrotransformation. Transformation frequency: 4.7×10^{-3}	92	8, 157
52	*Moraxella ovis* *(Neisseria ovis)*	γ-Proteobacteria	pathogen (cattle, sheep): eyes, keratoconjunctivitis	No data on competence accessible.	92	130
53	*Moraxella atlantae*	γ-Proteobacteria	CDC group M-3	No data on competence accessible.	92	130
54	*Moraxella bovis*	γ-Proteobacteria	pathogen (cattle): eyes, keratoconjunctivitis; strict aerobic, cornea conjunctiva; transmitted by flies	No data on competence accessible.	92	130
55	*Moraxella cuniculi*	γ-Proteobacteria	compare Neisseria	No data on competence accessible.	92	130
56	*Moraxella lacunata subsp. lacunata*	γ-Proteobacteria	inflamed + healthy conjunctiva (human), upper respiratory tract	No data on competence accessible.	92	130
57	*Moraxella lacunata subsp. liquefaciens*	γ-Proteobacteria	keratitis, endocarditis, sinusitis	No data on competence accessible.	92	130
58	*Moraxella nonliquefaciens*	γ-Proteobacteria	opportunistic pathogen (human), commensal of upper respiratory tract	No data on competence accessible.	92	130
59	*Moraxella osloensis*	γ-Proteobacteria	facultative pathogen (human)	Natural transformation on rich solid medium, after overnight growth on plates.	92	130, 158
60	*Moraxella, unnamed genetic entities*	γ-Proteobacteria	?	No data on competence accessible.	92	130

	Species	Class	Environment	Competence	References	
					1st	2nd
61	*Mycobacterium smegmatis*	Actinobacteria	heterotrophic, high GC gram+; commensal urogenitial tract, facultative pathogen (rare)	Competence proceeds throughout exponential growth and declines during the stationary phase, transformation frequency: 10^{-7} - 10^{-6} [8]	92	8
62	*Neisseria elongata subsp. elongata*	β-Proteobacteria	CDC group M-6; commensal oropharynx, opportunistic pathogen (endocarditis)	Constitutively competent[3]	92	130
63	*Neisseria elongata subsp. glycolytica*	β-Proteobacteria	transient colonizer of humans, opportunistic pathogen; wounds, endocarditis	Constitutively competent[3]	92	130
64	*Neisseria flava*	β-Proteobacteria	commensal (human), upper respiratory tract (opportunistic pathogen: endocarditis)	Constitutively competent[3]	92	130
65	*Neisseria flavescens*	β-Proteobacteria	commensal (human), upper respiratory tract (opportunistic pathogen: endocarditis)	Constitutively competent[3]	92	130
66	*Neisseria gonorrhoeae*	β-Proteobacteria	pathogen (human); urogenital tract	Constitutively competent[3], transformation frequency: 1.0×10^{-4}	92	8, 130
67	*Neisseria lactamica*	β-Proteobacteria	commensal (human); upper respiratory tract	Constitutively competent[3]	92	130
68	*Neisseria meningitidis*	β-Proteobacteria	pathogen (human); normal flora of nasopharynx (5-15%)	Constitutively competent[3], transformation frequency: 1.1×10^{-2}	92	8, 130
69	*Neisseria mucosa*	β-Proteobacteria	commensal (human), upper respiratory tract (opportunistic pathogen: endocarditis)	Constitutively competent[3]	92	130
70	*Neisseria perflava*	β-Proteobacteria	commensal (human); pharynx	Constitutively competent[3]	92	130
71	*Neisseria sicca*	β-Proteobacteria	commensal (human); oropharynx, opportunistic pathogen	Constitutively competent[3]	92	130
72	*Neisseria subflava*	β-Proteobacteria	commensal (human); upper respiratory tract, opportunistic pathogen (rare)	Constitutively competent[3]	92	130

	Species	Class	Environment	Competence	References	
					1st	2nd
73	*Neisseria weaveri*	β-Proteobacteria	CDC group M-6, commensal upper respiratory tract flora in dogs; dog bites	Constitutively competent[3]	92	130
74	*Nostoc muscorum* *(Desmonostoc muscorum)*	Cyanobacteria	photo-lithotrophic; terrestrial + freshwater environments; symbiosis with plants; hot	Transformation frequency: 1.2 x 10^{-3}	92	8
75	*Oligella urethralis*	γ-Proteobacteria	CDC group M-4, opportunistic pathogen: urosepsis	No data on competence accessible.	92	130
76	*Pasteurella dagmatis*	γ-Proteobacteria	zoonotic infections in humans, skin, soft tissue after animal bite	No data on competence accessible.	92	130
77	*Pasteurella pneumotropica*	γ-Proteobacteria	normal inhabitant of the oropharynx of mice, rats, cats, and dogs.	No data on competence accessible.	92	130
78	*Porphyromonas gingivalis*	Bacteroidia	pathogen (human), CFB group bacteria; gingivitis, periodontitis, oropharynx	Liquid growth until mid log phase. Natural transformation on solid medium. Transformation frequency: 10^{-4}	159	
79	*Pseudomonas ('Vibrio') WJT-1C*	γ-Proteobacteria	non-pathogenic; marine	Liquid growth until late log phase, natural transformation on filter. Transformation frequency: 3.4 x 10^{-7}	92	8, 160
80	*Pseudomonas alcaligenes*	γ-Proteobacteria	opportunistic pathogen (human); soil, water	Competent without pre-treatment. Transformable in solid and liquid culture. Maximum transformation frequency, in early stationary phase: 10^{-4}.	92	161
81	*Pseudomonas fluorescens*	γ-Proteobacteria	aerobic, psychrophilic, opportunistic pathogen after blood transfusions; soil, water,	Soil microcosm experiments Transformation frequency: 10^{-8}	89	129
82	*Pseudomonas mendocina*	γ-Proteobacteria	nosocomial infections (endocarditis); ubiquitous	Competent without pre-treatment. Transformable in solid and liquid culture. Maximum transformation frequency, in early stationary phase: 10^{-4}.	92	161
83	*Pseudomonas pseudoalcaligenes*	γ-Proteobacteria	non-pathogenic; environment	Competent without pre-treatment. Transformable in solid and liquid culture. Maximum transformation frequency, in early stationary phase: 10^{-4}.	92	161
84	*Pseudomonas spp.*	γ-Proteobacteria	ubiquitous	Liquid growth until late log phase, natural transformation on filter. Transformation frequency: 3.4 x 10^{-7}	89	162

	Species	Class	Environment	Competence	References	
					1st	2nd
85	*Pseudomonas stutzeri*	γ-Proteobacteria	heterotrophic; soil	Transition from log phase to stationary phase[8], small proportion of population competent, increasingly competent upon nutrient starvation; peak at pH 7, 20°C – 37°C; nutrient limited soil extract Transformation frequency: 7.0×10^{-5} [8]	92	8
86	*Psychrobacter immobilis*	γ-Proteobacteria	psychrophilic; antarctic soil, ice, sediment, food	No data on competence accessible.	92	130
87	*Pyrococcus furiosus*	Thermococci	marine hyper-thermophilic anaerobe	Transformable during early -, mid -, and late log phase, on solid medium	90	
88	*Ralstonia solanacearum*	β-Proteobacteria	non-pathogenic; plant, soil	Competent during exponential growth and not in response to any external factors. Transformation on solid medium. Transformation frequency: 4.06×10^{-7}	92	163
89	*Sinorhizobium meliloti*	α-Proteobacteria	heterotrophic; soil	Transformation frequency: 7.0×10^{-4}	92	8
90	*Staphylococcus aureus*	Bacilli	pathogen (human), low GC gram+; skin	Early to late log-phase, transformation frequency: 5.5×10^{-6} [8]	92	8, 88
91	*Streptococcus anginosus*	Bacilli	opportunistic pathogen (human); oropharynx	Induction: pheromone CSP (strain specific peptide; nanomolar)[74], comC, DE dependent[164]	89	165
92	*Streptococcus bovis JB1* *(Streptococcus equinus, S. gallolyticus)*	Bacilli	opportunistic pathogen (rare; endocarditis, colorectal cancer), low GC gram+; GIT cows, sheep (ruminants)	Pheromone induction; comX, RS dependent[166] comC, DE involvement[167] early exponential phase (2×10^{7} CFU/ml)[168]	92	168, 169
93	*Streptococcus constellatus*	Bacilli	purulent infections of the oral cavity; normal flora oropharynx;	Competence is induced by strain-specific peptide pheromones.	89	165
94	*Streptococcus crista* *(Streptococcus cristatus)*	Bacilli	commensal (human), low GC gram+; dental biofilm	No data on competence accessible.	92	170
95	*Streptococcus gordonii DL1 (Challis)* *(Streptococcus sanguis)*	Bacilli	human tooth enamel, low GC gram+; biofilm;	Pheromone induction	92	8, 171, 172

	Species	Class	Environment	Competence	References	
					1st	2nd
96	*Streptococcus infantarius*	Bacilli	dairy products, food, commensal human infants	Spontaneously late log phase (persistence: transiently, 30 min)[122] wide window of cell densities Pheromone induction; comX, RS dependent, high cell densities in CDM (10^9 CFU/ml), competence development in milk[173]	74	74, 173, 174
97	*Streptococcus infantis*	Bacilli	human tooth surface, low GC gram+; biofilm; oropharynx	No data on competence accessible.	92	175
98	*Streptococcus intermedius*	Bacilli	periodontitis; oropharynx	Competence is by strain-specific peptide pheromones.	89	165
99	*Streptococcus macedonicus*	Bacilli	food, fermented dairy products	Pheromone induction; comX, RS dependent[173]	173	176
100	*Streptococcus mitis*	Bacilli	odontogenic infections, mesophile; oropharynx	Induction: pheromone CSP (strain specific peptide; nanomolar)[74] comC, DE dependent[164]	89	8
101	*Streptococcus mutans*	Bacilli	pathogen (human), low GC Gram+; tooth decay; commensal oropharynx	Spontaneously late log phase (persistence: several hours) XIP: heptamer[177]; XIP responsiveness pH dependent[178], complex growth media refractory. Oral cavity: low cell density, neutral pH, low oligopeptide concentration[74] Pheromone induction; comX,RS dependent[179] Transformation frequency: $7.0 \times 10^{-4)}$ [8]	92	8
102	*Streptococcus oralis*	Bacilli	opportunistic pathogen; commensal oropharynx	The peak of competence was found at the early exponential phase of growth.	89	180
103	*Streptococcus pneumoniae*	Bacilli	pathogen (human), low GC Gram+; pneumonia, meningitis; nasopharynx of healthy carriers	Mid-log phase, independent from initial cell number. Rapid extinction of competence[74, 95] CSP autoinducer, alarmone; no nutritional signal reported.[3] Aminoglycosides, fluoroquinolone antibiotics[181] Early to late log-phase, up to 100% of the population competent; Ca^{2+} dependent, pH 7.3 to 8.0; transformation frequency: $2.9 \times 10^{-2)}$ [8]	92	8
104	*Streptococcus pyogenes*	Bacilli	pathogen (human); facultative anaerobic, beta hemolytic; skin, mucosa, oropharynx	Only transformable in biofilm[182] Pheromone induction; comX, RS dependent[166]	74, 182, 183	182

	Species	Class	Environment	Competence	References	
					1st	2nd
105	*Streptococcus salivarius*	Bacilli	oropharynx, upper respiratory tract of humans; commensal (opportunistic pathogen)	Pheromone induction; comX, RS dependent[166]	74	184
106	*Streptococcus sanguinis* *(Streptococcus sanguis)*	Bacilli	pathogen (human), facultative aerobic; commensal oropharynx	Transformation frequency: 2.0 x $10^{-2)}$ [8]	89	165, 185, 186
107	*Streptococcus suis*	Bacilli	pathogen (human, animal), meningitis, endocarditis	Pheromone induction: comX,RS dependent[119]	74	119
108	*Streptococcus thermophilus*	Bacilli	yoghourt, milk, facultative anaerobic; non-pathogenic	Spontaneously competent in complete defined medium (CDM) devoid of oligopeptides; XIP: 11 amino acids[187] Induction: casein-derived peptides[188] Early exponential phase[177] comX, RS dependent[166]	89	189
109	*Streptomyces spp.* *(S. virginiae, S. kasugaenis)*	Actinobacteria	heterotrophic, high GC Gram+; non-pathogenic (human); soil	Small proportion of population competent[8]	92	8
110	*Suttonella indologenes*	γ-Proteobacteria	compare Cardiobacteriaceae, opportunistic pathogen (endocarditis, rare); commensal	No data on competence accessible.	92	130
111	*Synechococcus elongatus*	Cyanobacteria	non-pathogenic, photoautotrophic; freshwater	Competence in overnight cultures. Transformation on solid medium. Tranformation frequency 1 x 10^{-7}	89	190, 191
112	*Synechocystis* sp. *strain 0L 50*	Cyanobacteria	photolithotrophic; freshwater	Transition from log phase to stationary phase, Transformation frequency: 2.0 x $10^{-4)}$ [8]	92	8
113	*Synechocystis* sp. *strain 6803*	Cyanobacteria	photolithotrophic; freshwater	Transition from log phase to stationary phase, Transformation frequency: 5.0 x $10^{-4)}$ [8]	92	8
114	*Tannerella forsythia*	Bacteroidia	pathogen (human); oropharynx, periodontitis	Competent under biofilm conditions.	192	
115	*Thermoactinomyces vulgaris*	Bacilli	heterotrophic, thermophile, low GC Gram+; pathogen by inhalation of spores (pulmonary	Transformation frequency: 2.7 x 10^{-3}	92	8

	Species	Class	Environment	Competence	References	
					1st	2nd
116	*Thermoanaerobacter* spp.	Clostridia	thermophile, obligate anaerobes	Competence during exponential growth without obvious induction event. Transformation in liquid, transformation frequency: 10^{-3}	193	
117	*Thermoanaerobacterium saccharolyticum*	Clostridia	thermophile, obligate anaerobes	Competence during exponential growth without obvious induction event. Transformation in liquid, transformation frequency: 10^{-4}	193	
118	*Thermococcus kodakaraensis KOD1*	Thermococci	hyperthermophile, sulfur-reducing	Competent in late exponential phase.	91	
119	*Thermosynechococcus elongatus BP-1*	Cyanobacteria	thermophile (55°C optimum); hot springs	Transformation rate: 10^{-3}/µg DNA[194]	89	194
120	*Thermotoga* spp.	Thermotogae	thermophile; strictly anaerobic	Competent in early exponential phase. Transformation in liquid. Transformation frequency: 10^{-7}	195	
121	*Thermus aquaticus*	Deinococci	heterotrophic, extreme thermophile (70°C optimum); hot springs	Transformation frequency: 6.4×10^{-4}	92	8
122	*Thermus caldophilus*	Deinococci	heterotrophic, extreme thermophile (75°C optimum); hot springs	Transformation frequency: 2.7×10^{-3}	92	8
123	*Thermus flavus*	Deinococci	heterotrophic; extreme thermophile (74°C optimum)	Transformation frequency: 8.8×10^{-3}	92	8
124	*Thermus thermophilus*	Deinococci	heterotrophic; extreme thermophile (65°C optimum); thermal vent/hot spring;	Transformation frequency: 1.0×10^{-2}	92	8
125	*Thiobacillus* sp. *strain Y*	β-Proteobacteria	chemolithotrophic; river water, sediments, hot acid springs, sewage, salt lakes	Transformation frequency: 1.7×10^{-3}	92	8
126	*Thiobacillus thioparus*	β-Proteobacteria	chemolithotrophic; river water, sediments, hot acid springs, sewage, salt lakes	Transformation frequency: 10^{-3}-10^{-2}	92	8
127	*Vibrio cholerae*	γ-Proteobacteria	pathogen (some strains; human), facultative anaerobic; brackish water and estuaries, often in	Induction: absence of glucose (carbon catabolite repression)[85] and presence of chitin[106]	89	106

	Species	Class	Environment	Competence	References	
					1st	2nd
128	*Vibrio fischeri*	γ-Proteobacteria	marine environment including chitin surfaces	Transformation occurs in the presence of chitin oligosaccharides. Transformation frequency: 10^{-7}	3	196
129	*Vibrio parahaemolyticus* *(Beneckea parahaemolytica)*	γ-Proteobacteria	heterotrophic, pathogen (human); water, aquatic animals	Transformation frequency: 1.9 x 10^{-9}	92	8
130	*Vibrio spp.*	γ-Proteobacteria	heterotrophic	Transition from log phase to stationary phase; drop after temperature shift from 33°C to 37°C, optimum: 15°C - 33°C Transformation frequency: 2 x $10^{-4)\ 8}$	89	162, 197
131	*Vibrio vulnificus*	γ-Proteobacteria	halophilic bacterium, opportunistic pathogen (human); estuarine waters, contaminates	Competent during growth on chitin in the form of crab shells.	84	198
132	*Xylella fastidiosa*	γ-Proteobacteria	plant pathogen; vector borne disease (*Homalodisca vitripennis)*	Competence during early exponential growth. Transformation in liquid. Transformation frequency: 10^{-6}	199	

1.6.8 Homologous Recombination

The exchange of DNA sequences between identical or substantially similar molecules is termed homologous recombination. Concerning transformation this process involves an RecA-mediated interaction between the bacterial host chromosome and an internalized single stranded DNA molecule of exogenous origin.[74] Bacteria were originally thought to reproduce mainly clonally and disseminate their genes predominantly vertically from a parent to the daughter cell by cell division. However, it is now clear that bacteria frequently exchange genes laterally, which is assumed to be the major driving force in bacterial evolution.[43, 73]

Extent and nature of recombination varies substantially among microbiological species and among lineages of the same species.[73] *Neisseria spp, Helicobacter spp., Flavobacterium spp.* and *Pelagibacter spp*. are reported to be highly recombinogenic, wheras *Escherichia coli* and *Salmonella enterica* are known to be rarely engaged in recombinogenic DNA exhanges.[200-202]

Non-effective recombination - when a phenotypic effect is not directly observable although DNA fragment exchange had occurred - is almost certainly very frequent in closely related species[73], because homologous recombination is crucial for DNA repair.[203] There is an interdependence among recombination and selection: recombination events are either quickly purged from population due to deleterious effects on bacterial fitness or widespread in population.[73] Homologous recombination has evolved in bacteria either for DNA repair, as by-product of DNA uptake for metabolisation, to remove deleterious mutations or to combine beneficial mutations.[204] A problem for the detection of recombination events is that recombinational exchanges of only a small number of nucleotides may be mistaken as point mutations.[47]

Bacterial speciation

A bacterial species is difficult to define because high rates of recombination result in the transfer of DNA between relatively distantly related bacteria – a process which would blur any species-specific boundaries if it occurred without any barriers.[205] The result would be a genetic continuum due to horizontal gene transfer and recombination. However speciation does occur and generates permanently distinct clusters of closely related bacteria.[206-208] This arises as a consequence of recombination failing more frequently between DNA sequences that are different than between those that are similar and not because of fundamental ecological constraints or geographic separation.[209, 210] The decline in the recombination rate as a function of genetic distance forms a soft barrier providing the primary basis for speciation.[205] This observation is also pivotal for the formation of mosaic genes: Homologous recombination is decreasing in a log-linear relationship with increasing sequence divergence among the involved DNA molecules and falls below the level of detection at a sequence divergence higher than 25 - 30% (Figure 11).[205]

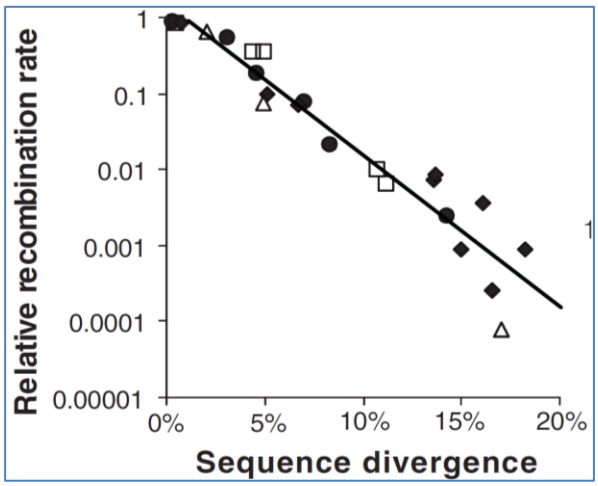


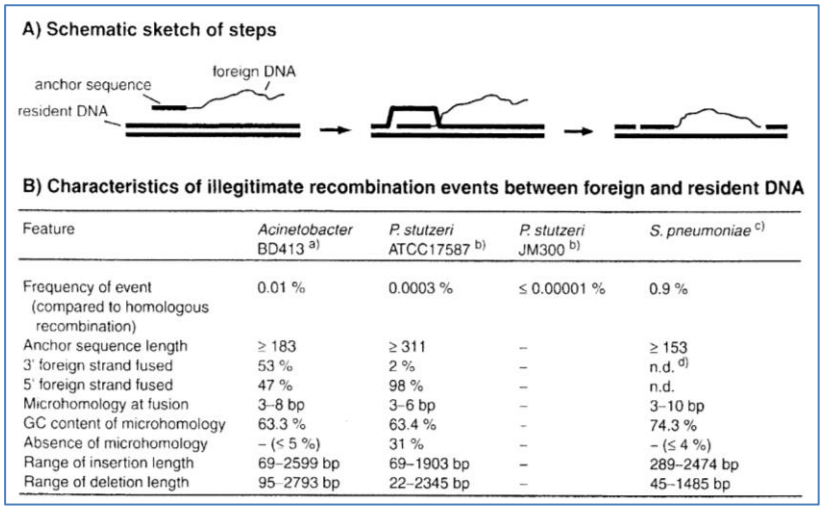
Figure 11: Recombination rates for a range of related donors.

The recombination rate is represented as a function of the proportion of sequence that is different (sequence divergence), for a variety of bacterial recipients.[205] The best fit log-linear curve is shown, with intercept at 0.8% and a slope of 19.8. Data are from [140, 211, 212].

Homology-directed illegitimate recombination

Non homologous sequences can be integrated into a bacterial chromosome via homology-directed/facilitated recombination.[56, 213] This process is thought to be of major importance concerning the formation of mosaic genes.[56]

The defining parameters as shown in Figure 12 indicate that there are only short stretches of homologous anchor sequences (153 bp for *S. pneumonaie* and 183 bp for *Acinetobacter baylyi*) required. The additionally necessary regions of microhomology between incoming DNA and chromosomal target are tiny (3 – 10 bp).



B) Characteristics of illegitimate recombination events between foreign and resident DNA

Feature	*Acinetobacter* BD413 [a]	*P. stutzeri* ATCC17587 [b]	*P. stutzeri* JM300 [b]	*S. pneumoniae* [c]
Frequency of event (compared to homologous recombination)	0.01 %	0.0003 %	≤ 0.00001 %	0.9 %
Anchor sequence length	≥ 183	≥ 311	–	≥ 153
3' foreign strand fused	53 %	2 %	–	n.d. [d]
5' foreign strand fused	47 %	98 %	–	n.d.
Microhomology at fusion	3–8 bp	3–6 bp	–	3–10 bp
GC content of microhomology	63.3 %	63.4 %	-	74.3 %
Absence of microhomology	– (< 5 %)	31 %	–	– (≤ 4 %)
Range of insertion length	69–2599 bp	69–1903 bp	–	289–2474 bp
Range of deletion length	95–2793 bp	22–2345 bp	–	45–1485 bp

Figure 12: Homology-directed/facilitated illegitimate recombination.

Foreign DNA integration during natural transformation by homology-facilitated illegitimate recombination (HFIR) of Gram-negative and Gram-positive bacteria is depicted. In (A) the basic mechanism is presented in a schematic sketch and in (B) molecular details identified in transformants generated by HFIR during natural transformation are summarized. a) de Vries and Wackernagel[213], b) Meier and Wackernagel[214], Prudhomme et al.,[56] d) n. d. = not determined (data from de Vries and Wackernagel[92]).

1.6.9 Hot Spots of Mosaic Gene Formation

Hot spots for the formation of mosaic genes are expected to co-localize with environments reported to sustain high levels of horizontal gene transfer and subsequent selection of transformants.[215] The most prominent examples are the rhizosphere, plant compartments, and the gastrointestinal tract of various organisms. DNA persistence was also found during food processing.[216] These environments display a variety of different microorganisms in close contact and nutrients available to support high bacterial cell densities. Transformation barriers are thought to be alleviated in these hot spots.[215] Important factors for each environment are DNA persistence and availability for transformation, the presence of competent bacteria and an appropriate selection pressure for transformants. It has to be stressed that there is no experimental evidence available to date dealing explicitly with hot spots for the formation of mosaic genes. However, the environments discussed in the following section are highly likely to support also the formation of mosaic genes because they have been shown to be matrices for horizontal gene transfer by natural transformation:

The gastrointestinal tract as hot spot for mosaic gene formation

Genome sequencing has identified traces of multiple transfer events in the history of species and suggests that horizontal gene transfer probably proceeds in any microbial habitat.[92] Concerning transformation events in the gastrointestinal tract (GIT), research has focused on possible transformation of bacteria with transgenes from dietary GM plants. The state of knowledge in this field has been reviewed by deVries and Wackernagel[92] and recently by Rizzi et al.[217] Both reviews indicate that such transfers are technically possible, but have not been demonstrated *in situ*, yet. To best knowledge there is also no direct evidence available, yet, for *in situ* transformation of gastrointestinal bacteria with DNA from other sources. In contrast, horizontal gene transfer via conjugation has already been documented in the GIT.[92, 218] In addition to the issue of transgenes in GM food, there is increasing interest in the possibility of gene transfer between dietary probiotic bacteria, which are selected for competitive colonization and pathogenic bacteria in the GIT.[219, 220]

The food matrix as element of the food/feed chain as hot spot for mosaic gene formation

Common food sources do not provide conditions which will lead a priori to a rapid degradation of extra-cellular DNA.[221] Apparently, a wide range of different kinds of foods (milk, soy drink, tomato-, orange-, carrot-, vegetable juice, canned cabbage, soy beans, shrimps etc.) supports transformation of *E. coli* indicating sufficient DNA stability in these environments.[222] Some food matrices even seem to provide protection against DNase activity and stability to free DNA (e.g. soymilk, tofu, fermented sausage).[223] Fresh plant material from leaves and grains usually contain, even after mechanical treatment, DNA fragments large enough to code for functional genes. Various canola substrates like whole seeds, cracked seeds, meal and diets could be shown to contain intact plant genes.[224] However, many fruits and vegetables provide highly acidic pH conditions. Thus, an increased DNA degradation activity by acid hydrolysis must be expected under such circumstances and disintegration of cellulose cell walls followed by the release of endogenous nucleases may lead to deleterious effects on the integrity of free DNA and to a reduction of transformability of potentially present competent bacteria.[225]

The rhizosphere / plant matrix as element of the food/feed chain as hot spot for mosaic gene formation

The rhizosphere and the rhizoplane of plants have been identified to support high rates of HGT, however data were preferentially obtained for conjugation.[215, 226] Due to agricultural tilling the soil is regularly mixed and humidity changes due to water channels and animal and plant induced processes generating habitats, in which various bacteria commingle. The rhizosphere is considered to offer optimal colonisation microhabitats compared to bulk soil.[163] Various bacterial species are present in biofilms or in micro-colonies supporting cell-to-cell contacts and gene transfer. Leaves of crop plants are usually inhabitated by approximately 10^4 bacteria per g wet weight. Their rhizosphere contains typically 10^8 bacteria per g soil.[227] Bacteria residing on or close to plants may be exposed to DNA from the host plant during plant decay and root decomposition, but also by mechanical disruption.[92]

The rhizosphere provides nutrients for bacterial growth and compounds that stimulate natural transformation. Root exsudates support increased metabolic activity and contain actively excreted plant DNA (Davison 1999; Nielsen et al. 2001).[228, 229] It has also been shown that organic compounds, such as organic acids, amino acids and sugars (naturally found in the rhizosphere), could stimulate the natural transformation of *Acinetobacter* sp. BD413.[227] A high phosphate salt level was found to allow detectable gene transfer. When using mixtures of organic compounds and high phosphate salts, based on estimated concentrations in the maize rhizosphere, the highest transformation frequencies were observed (4.4×10^{-6}).[227]

The phytosphere and plant internal regions like the intercellular space[199] can also constitute strong stimulants of transformation, likely due to their provision of favourable niches to bacteria.[230] As indicated before, the induction of competence of *Acinetobacter spp.* was observed *in planta*, when it was co-infecting tomato plants together with *Ralstonia solanacearum*.[231]

Additional environments prone for horizontal DNA fragment transfer by transformation

A substantial number of highly transformable bacterial species have been recovered from the mammalian oropharynx, the upper respiratory tract (*Streptococcus spp*.)[89] and from the urogenital tract (*Neisseria gonorrhoea*)[89] identifiying these habitats as highly prone for the formation of mosaic genes. The detection of transformable methanogenic bacteria (*Methanothermobacter spp.*)[8] underscores the importance of sewage and sewage treatment plants as hot spots of recombination, horizontal gene transfer and for the formation of mosaic genes.[92] It is important to note that several bacterial species living under harsh environmental conditions in hot springs (*Pyrococcus spp, Thermus spp.)*[8, 90] or in arctic ice or soil (*Psychrobacter spp.*)[130] have been identified to be naturally transformable rendering extreme habitats as supportive for DNA fragment transfer.[92] The application of transformable species like *Lactobacillus spp.*[89] in the food and dairy industry indicate the potential of milk, yoghurt and cheese for horizontal gene transfer.[92] There is evidence for natural transformation of *Bacillus subtilis* in foodstuffs.[232] Plant wounds are also reported to be supportive for natural transformation.[92]

1.7 Search Strategy

For the analysis of the recent literature concerning the research topic the bibliographic open access database PubMed was used.

PubMed is sponsored by the US government and hosted by the National Center for Biotechnology Information (NCBI). It provides access to peer-reviewed literature in the fields of biomedicine and health, covering portions of the life sciences, behavioral sciences, chemical sciences, and bioengineering, and additional relevant web sites and links to the other NCBI molecular biology resources (http://www.ncbi.nlm.nih.gov/pubmed).

The literature search was performed in September 2013 (as of 12/11/2013). Search terms were restricted to the field “Title/Abstract”. There was no restriction to special publication categories (e.g. reviews). The years 2000 up to 2013 were analyzed. Each year was separately checked. The following search strings were applied:

a) ((((((("mosaic gene"[Title/Abstract]) NOT virus) NOT viroid) AND ("2013"[Date - Publication] : "2013"[Date - Publication]))) NOT chromosome) NOT trisomy))

b) ((((((("mosaic genes"[Title/Abstract]) NOT virus) NOT viroid) AND ("2013"[Date - Publication] : "2013"[Date - Publication]))) NOT chromosome) NOT trisomy))

An additional search run was performed using the following string:

"Mosaic"[Title] NOT virus

References earlier than 2000 were added according to relevance. Papers dealing obviously with the wrong topic were excluded.

Several key papers dealing with the phenomenon of mosaic gene formation but not using the relevant key words in the title or abstract were added according to personal experience or by recommendation of the external project adviser. The ISI Web of Science was checked additionally for papers not retrieved by the PubMed analysis using the same search strategy.

This approach for the literature search retrieved 104 core papers dealing with mosaic gene formation in all its aspects. These core papers were analyzed in detail by the project team.

Additionally PubMed was screened for recent developments in the fields of “natural transformation” and “competence” with a focus on publications since 2004 using the following search strings:

a) ((natural transformation[Title/Abstract]) AND (competence[Title/Abstract]OR competent[Title/Abstract])) AND ("2004/01/01"[Date - Publication] : "3000"[Date - Publication])

b) (((natural competence[Title/Abstract]) AND DNA[Title/Abstract]) NOT natural transformation) AND ("2004/01/01"[Date - Publication] : "3000"[Date - Publication])

Scopus was checked for hits missing from the PubMed analysis using the following search strings:

(TITLE-ABS-KEY("natural transformation") AND (TITLE-ABS-KEY(competent) OR TITLE-ABS-KEY(competence))) AND PUBYEAR > 2003 and

(TITLE-ABS-KEY("natural competence") AND TITLE-ABS-KEY(dna) AND NOT TITLE-ABS-KEY("natural transformation")) AND PUBYEAR > 2003

The reason for starting this part of the literature analysis in 2004 was the seminal review on "Bacterial gene transfer by natural transformation in the environment" which was published by Lorenz and Wackernagel in 2004.[8] On sixty pages the knowledge on transformation and competence development available up to then was comprehensively presented and renders additional datamining before this date superfluously.

Up to date information on "recombination" was retrieved from PubMed for the years 2008 – 2013 (using the same strategy as with mosaic genes as explained above). All hits were combined with references retrieved from the core papers of relevance according to personal experience or due to recommendations by the external project adviser resulting in an additional 212 papers. Most of the data have been extracted and summarized in the relevant tables to be found in Annexes 1 and 2.

2 Analysis of Sequence Databases for Mosaic Genes

2.1 Summary

The present study analysed two transgenes – *aph(3')-IIa* and *epsps* - of bacterial origin that are present in many commercial GM plants for their potential to take part in partial horizontal gene transfers resulting in mosaic gene structures.

The antibiotic resistance gene *aph(3')-IIa* encodes an aminoglycoside phosphotransferase which inactivates the critically important aminoglycosides kanamycin and neomycin and serves as a marker in many GM plants. The recombination potential of *aph(3')-IIa* was analyzed in detail, as the evolution of antibiotic resistance genes is of high relevance for human health. For this purpose the GenBank database was thoroughly screened for *aph(3')-IIa* homologs. Three datasets comprising 25, 11 and 15 sequences suitable for recombination analysis were obtained. The retrieved sequences were aligned and checked for recombination signals with a battery of proficient and well-established recombination detection algorithms provided by the T-REX suite, the RDP4 software package and the online tool GARD. According to the outcome of our analyses *aph(3')-IIa* appears to be stable compared to the well-known mosaic nature of certain penicillin binding proteins and tetracycline resistance determinants. The available number of appropriate sequences and the overall variability in the accessible sequence datasets was low: 48 of a total of 51 homologs shared more than 99% sequence identity with *aph(3')-IIa*. However, a single recombination event between *aph(3')-IIa* homologs could be identified with high confidence.To our knowledge this is the first time to present statistically sound evidence for the involvement of *aph(3')-IIa* in the formation of mosaic genes.

It is important to note that the applied recombination detection algorithms are extremely sensitive to mis-alignments and to the overall composition of sequence datasets. The available datasets dealing with *aph(3')-IIa* sequence variability applicable for this study are scarce. Data from comprehensive studies dedicated on evaluating the variability of *aph(3')-IIa* similar sequences in natural environments are lacking. The acquisition of the respective sequence data is necessary and highly recommended to refine and validate the performed bioinformatic analyses and to determine the relevance of *aph(3')-IIa* in the formation of mosaic gene structures in natural environments, conclusively.

The second part of this work focused on 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and its genetic determinant, the *aro*A gene. EPSPS is an enzyme in the shikimate pathway of bacteria, plants and fungi and the cellular target of the herbicide glyphosate. Variability and recombination history were analyzed among genes homologous to CP4 *epsps*, the *aro*A gene of *Agrobacterium* sp. CP4 encoding a glyphosate tolerant EPSPS isoenzyme, which has been inserted as transgene into most glyphosate tolerant crops. BLAST search of GenBank recovered 144 different homologs sharing 70 - 87% sequence identity with CP4 *epsps*. Sequences with similarity to CP4 *epsps* higher than 87% were not detected, and all homologs were more similar to the original bacterial CP4 *epsps* sequence than to coCP4*epsps*, a codon optimized version for efficient expression in plants. All 144 CP4 *epsps* homologs were *aro*A genes from members of the order *Rhizobiales,* and sequence similarities reflected the phylogenetic relatedness of these organisms, suggesting primarily vertical inheritance.

Recombination analysis with the RDP4 software revealed partial recombination within the genera *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* and between members of the *Agrobacterium/Rhizobium* group. In conclusion the *aro*A gene appears to be susceptible to partial horizontal gene transfer. We recommend cultivation-independent surveys on the occurrence of *aro*A variants sufficiently similar for recombination with transgenic CP4 *epsps* fragments in habitats relevant to GMO production and consumption.

Finally, a general survey of literature and GenBank databases for already characterized mosaic genes was carried out and revealed a huge number of hits covering literally all functional categories of genes (metabolism, information storage, cellular processes, virulence, or resistance). Moreover, the phenomenon of partial horizontal gene transfer is not restricted to highly transformable bacterial species. These results support the notion that the formation of mosaic genes is a central and pivotal theme in bacterial evolution enabling prokaryotes a rapid adaptation to changing environmental conditions.

2.2 Background

Genetic recombination leading to mosaic patterns in antibiotic resistance genes results in therapy failure of infectious diseases in clinical settings.[15, 16] This is of particular concern because pathogens utilizing this mechanism gain the opportunity to rapidly respond to alternating antibiotic therapy. These pathogens represent agents causative for substantial adverse effects on animal and human health by increasing the morbidity and mortality rates due to infectious diseases and by amplifying the financial burden of public health systems.

The majority of the commercially available GMOs contain prokaryotic transgenes like antibiotic resistance markers or other metabolically advantageous genes of bacterial origin.[17] These transgenes will be part of the DNA released into the environment upon decay of GMOs.[18] Competent bacteria may acquire free DNA available in their environments via natural transformation.[19] The involved DNA can be of a size that includes intact genes, or if further fragmented, of sub-gene size. Irrespective of size, such DNA may transfer into competent bacteria and recombine with their sequence-similar counterparts in naturally occurring bacterial populations. If the recombining regions are short, the resulting gene acquires a mosaic structure (mosaic gene), in which sections of native gene sequences have been replaced by foreign sequence elements. Such mosaic genes may potentially generate a new phenotype that provides a benefit to the bacterial host under certain selective environmental conditions. Genetic recombination leading to mosaic patterns in antibiotic resistance genes provide bacterial pathogens with the opportunity to rapidly respond to alternating antibiotic selection pressure by generating new resistance determinants with alternative targets or expanded substrate specificities.

However, there is currently no experimental evidence available that can unambiguously support or disprove the hypothesis that antibiotic resistance marker genes like *aph(3')-IIa* or bacterial herbicide resistance genes like CP4 *epsps* can be involved in the formation of mosaic genes. This knowledge gap should be narrowed by a detailed analysis of the available data concerning sequence diversity of *aph(3')-IIa* on the one hand, and the prevalence and ecology of described mosaic genes in bacterial populations on the other hand.

2.2.1 Mosaic Genes

A mosaic gene is defined as "*an allele acquired through transformation or conjugation (e.g. from a different bacterium) and subsequent integration through intragenic recombination into the original host allele*" (Boc et al.).[13] The term "mosaic" originates from the pattern of interspersed blocks of sequences having different evolutionary histories but ending up combined in the resulting allele subsequent to recombination events. The recombined segments can be derived from other strains of the same species or from more distantly related bacterial or viral organisms.[32-34] A schematic representation of a typical mosaic gene structure is depicted in Figure 13. The evolutionary relationship of the presented mosaic leukotoxin A genes as shown in Figure 13 is exemplarily delineated in Figure 14.

The formation of mosaic genes was initially observed with penicillin binding proteins (pbp) in *Streptococcus pneumoniae* and was identified as a major threat for public health due to treatment failure of infectious diseases caused by this pathogen.[15, 16]

Additionally to pbp genes mosaic patterns have been detected in tetracycline resistance determinants (*tet*(M)/*tet*(O)/*tet*(W)/*tet*(32) genes), autolysin genes (*lyt*A), neuraminidase genes (*nan*A), surface proteins (*psp*A), immune response modulation M serotypes (emm gene family), streptokinase genes (*ska*), shikimate metabolism genes (*aro*E), glutamine synthetase genes (*glnA*) and a avariety of other genes in a huge number of different bacterial species.[32, 200, 233-237] Mosaic genes appear to be not restricted to highly transformable species because bacteria like *E. coli* and *Salmonella* spp. – usually reported as naturally non-transformable - have been observed to carry fragmented alleles.[7]

The development of mosaic genes is, thus, a common theme in bacterial evolution covering a large array of different gene families coding for various functions (e.g. antibiotic resistance, virulence, metabolic "housekeeping" genes).[13]

Bioinformatic analysis indicated that at least 10 – 20% of all bacterial protein coding genes may be classified as "mosaic genes".[7] However, this classification scheme relies on a functional annotation concept requiring the variable DNA region of the putative mosaic gene to contain a contiguous stretch of at least 210 basepairs coding for a minimum of 70 amino acids flanked by well conserved regions. It is obvious that incoming DNA does not adhere to functional protein domain boundaries. The actual number of bacterial genes affected by mosaic transformations must therefore be significantly higher than proposed by Zheng et. al.[7] Exchange of gene fragments is constantly occurring in bacterial populations but mostly invisible for external observers because exchange of identical sequence fragments is phenotypically silent. Even exchange between slightly different sequences may not always immediately lead to an altered phenotype but may become obvious only under modified environmental conditions.[16] There are indications that genes from all functional categories undergo DNA exchange.[47, 238] Therefore, it appears to be reasonable to assume that any bacterial gene in any bacterial host may be target of recombinational events resulting in mosaic gene structures.[13]

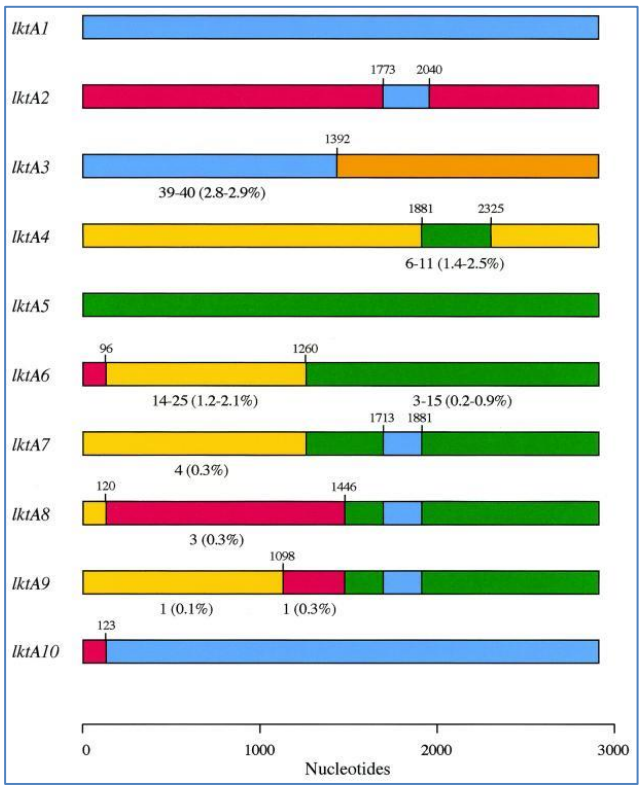


Figure 13: Mosaic structure of the leukotoxin A gene from *Pasteurella haemolytica*.

Schematic representation of the mosaic structures of alleles representative of the major allelic groups *lkt*A1 to *lkt*A10. The different colors indicate sequence identity and the likely origins of recombinant segments. The number of sites different from those in the corresponding region of the likely donor allele(s) (and the degree of divergence) is indicated below certain recombinant segments. All other segments exhibited 100% sequence identity with the corresponding regions of the donor alleles. Numbers above the proposed recombination sites indicate the position of the last nucleotide at the downstream end of the recombinant segment (data and figure from Davies et al.[36]).

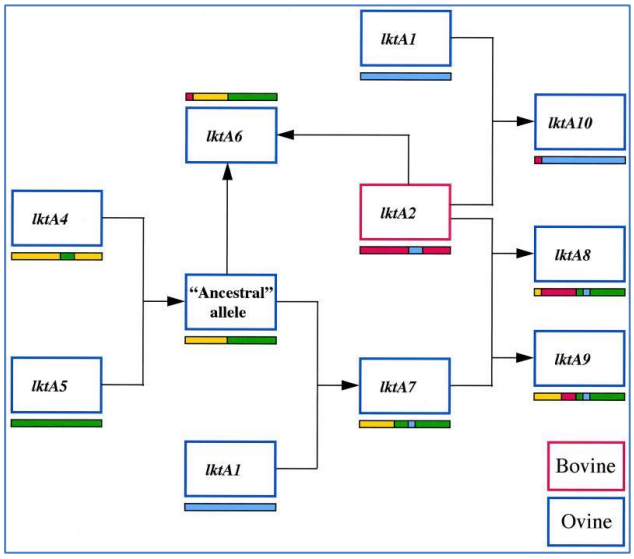


Figure 14: Phylogenetic relationship of leukotoxin A alleles from *Pasteurella haemolytica*.

Proposed sequence of recombination events in the evolution of *lkt*A leading to the formation of *lkt*A8 and *lkt*A10 type alleles in the ovine-specific lineages. The central role of the bovine *lkt*A2 allele in the evolution of ovine alleles *lkt*A6, *lkt*A8, *lkt*A9, and *lkt*A10 is displayed. The mosaic structures of the alleles are as shown in Figure 13 (data and figure from Davies et al.[36]).

2.2.2 Partial Horizontal Gene Transfers (PGT) and Recombination

The formation of mosaic genes is inherently connected with the uptake of free DNA from the environment via natural transformation and with the concept of partial gene transfer (PGT) in contrast to the traditional model of complete horizontal gene transfer (HGT) where the intact incoming gene either supplants the orthologous gene of the recipient genome, or is added to it.[35]

Horizontal gene transfer is mediated on the genome level by homologous recombination requiring usually a high level of similiarity between incoming and recipient DNA sequences which usually limits this process to closely related bacteria.[45] A log-linear decrease of the recombination frequency could be demonstrated with an increasing sequence divergence between donor and recipient DNA.[140, 205, 211, 212]

Mosaic gene formation is also mediated by homology-directed illegitimate recombination which facilitates the insertion of non-homologous DNA by homologous recombination (Figure 15).[16, 56] The creation of sequence variability via this mechanism is usually counterbalanced by the bacterial mismatch repair-system; however this system gets saturated if too many mismatches appear concomittantly.[239-241] An absolute requirement for homology-directed illegitimate recombination is the presence of an anchor-sequence in the donor DNA homologous to bacterial recipient chromosomal regions.[56] The minimum length of this anchor sequence was established to be 153 bp in *S. pneumoniae* and 183 bp in *A. baylyi*.[213, 242] Recently, evidence was provided that DNA fragments as short as 20 bp were taken up by bacteria and incorporated into the genome inducing variability in the chromosomal target sequences.[77] Homology-directed illegitimate recombination is frequently associated with microhomologies of 3 – 10 nucleotides at the strand terminus opposite the anchor sequence of the incoming DNA with the genomic target region.[56] The foreign DNA is usually transferred into to the recipient bacterium via natural genetic transformation; integration of the single stranded DNA into the genome is mediated via recA dependent recombination.[118]

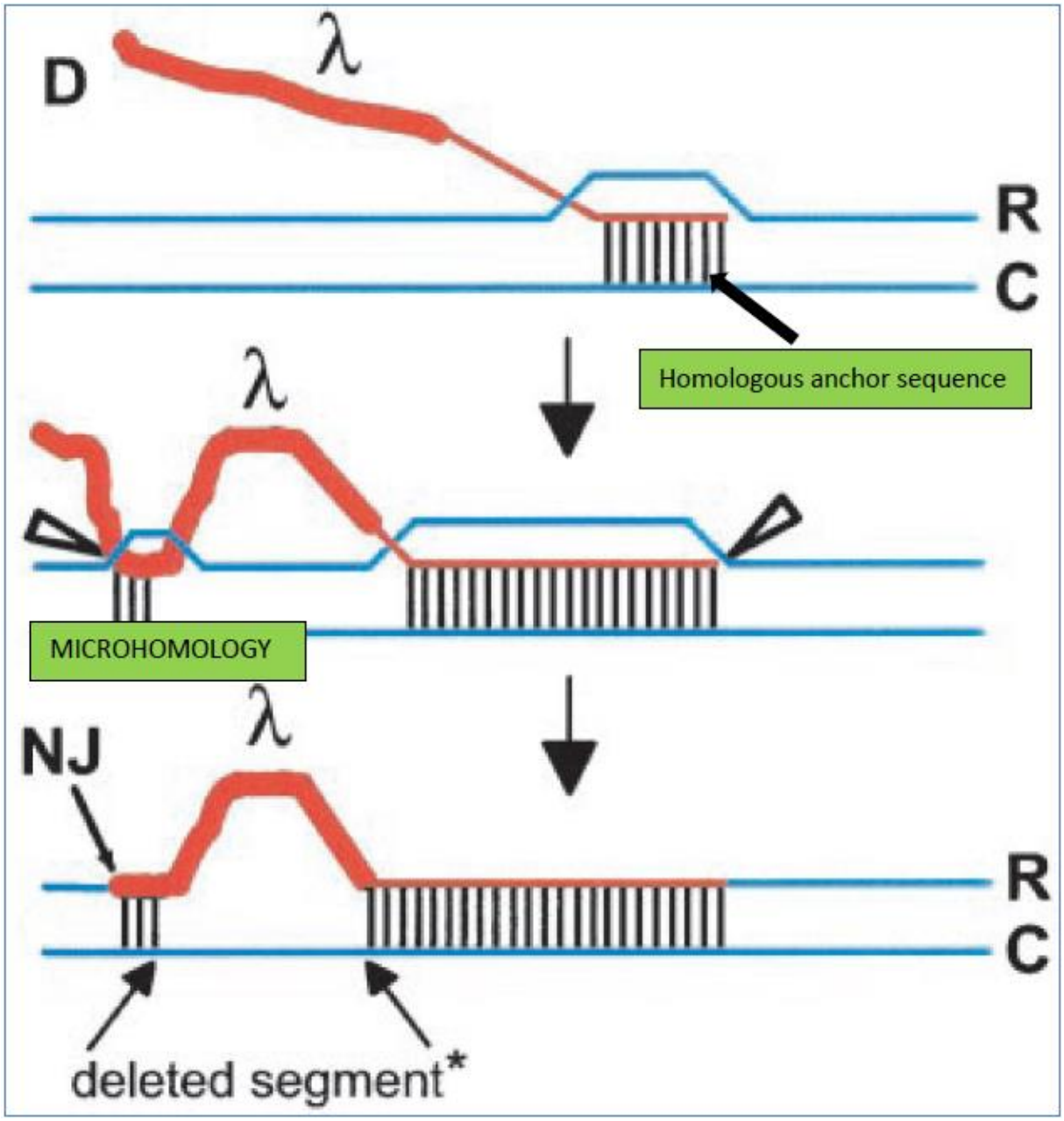


Figure 15: Homology-directed illegitimate recombination and microhomologies.

The role of novel joints during transformation of *S. pneumoniae*: RecA-driven invasion of the homologous donor strand (D, in red) into the recipient duplex (in blue) produces a donor-recipient heteroduplex that favors transient pairing between heterologous donor DNA (λ or any DNA; thick red line) and the complementary strand (C) of the recipient chromosome at microhomologies (three vertical bars). Concomitant incision of the heterologous strand and the displaced chromosomal recipient strand (R) by a putative resolvase (open triangle) followed by ligation would create a NJ. DNA replication would then generate a wild-type and a mutant chromosome. No strand polarity is indicated as the polarity of heteroduplex extension for D-loop joint molecules is unclear. *: deleted in the R strand; NJs: novel joints = microhomology (modified from Prudhomme et al.[56]).

2.2.3 Bioinformatic Tools for Recombination Detection in Aligned Sequence Datasets

The most commonly used bioinformatic tools to date to detect partial horizontal gene transfers leading to mosaic gene structures are the partial horizontal gene transfer module of T-REX and the RDP4 software package (for details see below).

T-REX/partial horizontal gene transfer module

The partial horizontal gene transfer detection algorithm developed by Boc et al. [13] requires as input a multiple sequence alignment (MSA) containing different variants of the gene of interest. In addition it requires a reference tree describing the overall phylogenetic relationships between the organisms carrying the different gene variants. Recombination in the gene of interest is detected by comparing the reference tree and multiple "partial gene trees" constructed from different regions of the MSA in a sliding window procedure as outlined in Figure 16. The reliability of the detected partial gene transfers is assessed in a bootstrapping procedure. The partial HGT detection module of T-REX can be executed online (http://www.trex.uqam.ca) and is considered as one of the most up-to-date and proficient analysis tools in intragenic recombination analysis. For more details on the versatility of the T-REX suite see Boc et al. and/or the T-REX server homepage.[51]

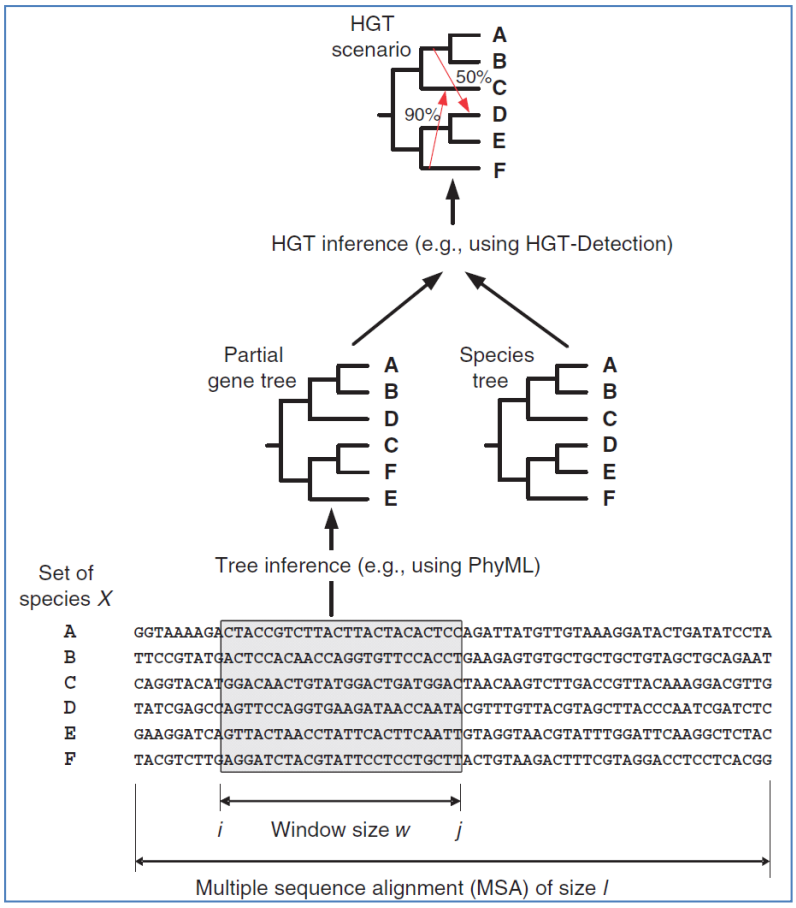


Figure 16: Inferring partial gene transfers from multiple sequence alignments using the T-REX software package.
Data from Boc et al.[13]

RDP4 (Recombination Detection Program)[243]

RDP4 is an updated and streamlined version of the software package RDP3 created by Martin et al.[243] It incorporates a series of different recombination detection methods (Table 2). In addition to the original RDP method of Martin and Rybicki[244] RDP4 includes BootScan[245], MaxChi[14], Chimaera[246], GeneConv[247], SiScan[248], LARD[249], 3Seq[250] and several other modules.[52] For details on the respective algorithms see the specified references. All of these methods require a MSA of the gene of interest as input, and most of them rely on comparing the sequences in triplets in sliding window procedures. As an example, Figure 17 shows the principle of the original RDP method. Each individual method includes a procedure to calculate the statistical support of detected recombination events as a p-value (Table 2). Bonferroni correction is applied to correct repetitive comparisons of multiple windows in multiple sequence triplets. The major asset of RDP4 is, that all these methods can be applied simultaneously to target various aspects of information in an MSA and to confirm detected recombination events by comparing the results obtained with different algorithms.

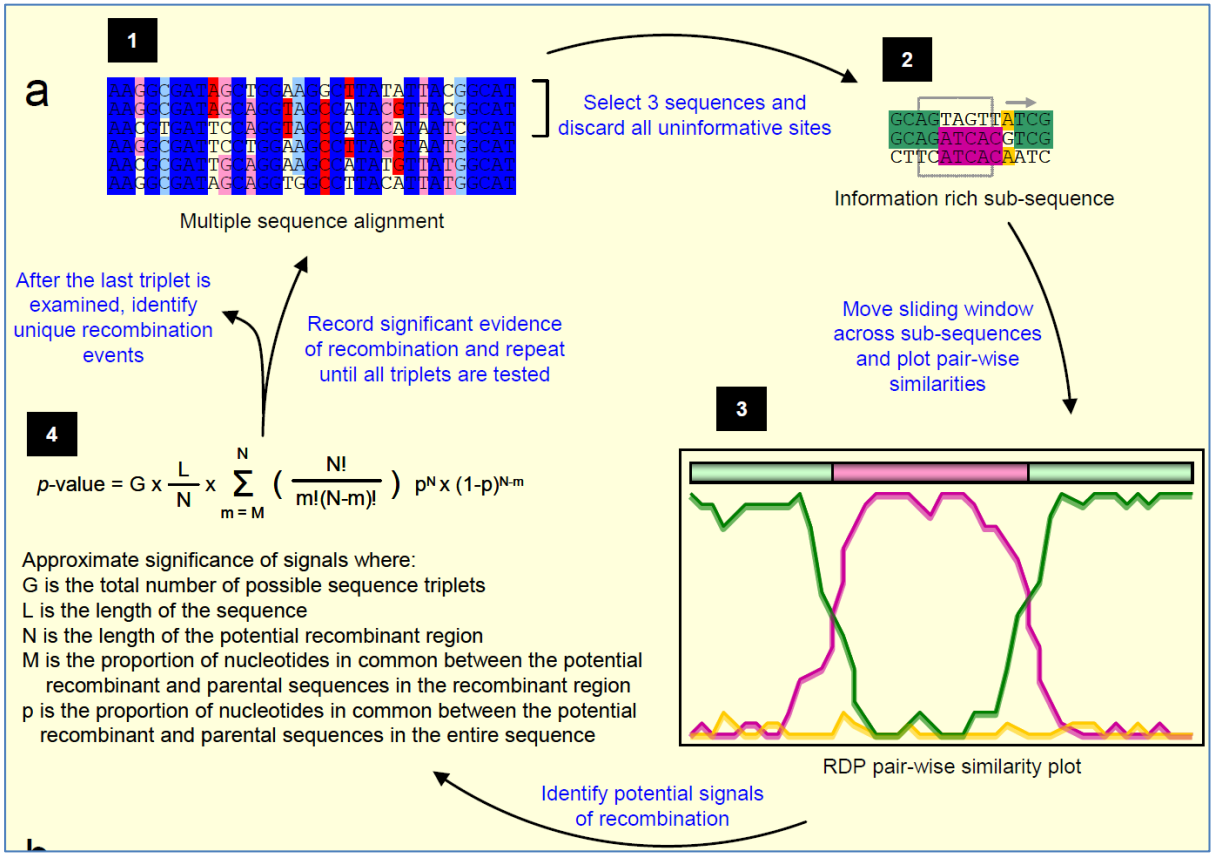


Figure 17: Iterative process for the identification of recombination events as employed by RDP4.

1… Multiple sequene alignment (= sequence dataset); usually prepared with programs based upon ClustalW
2… Sampling of sequence triplets for analysis. Removal of uninformative sites
3… Moving sliding window across subsequences and plot pair-wise similarities
4… Identification of potential recombination signals and plausibility check
Data from Martin et al.[52]

All functions are easily accessible via a userfriendly graphical interface and streamlined for analysis flow automation. Large datasets in the range of 1000 x 1kb or 20 x 2 megabasepairs are analysed within 48 h on a conventional desktop PC.

RDP4 automatically checks the retrieved recombination signals for whether they are misalignment artifacts. In this context RDP4 is capable to automatically detect crude misalignments in the imported datasets and implement counter measures to overcome this shortcoming.

Table 2. Recombination detection and analysis methods of RDP4 used in this study.

Method	Implementation	Identifies recombinants	Estimates breakpoints	Estimates regions	p-value calculations
Original RDP method	RDP4	+	+	+	Binomial distribution
GENECONV	RDP4 + GENECONV	+	+	+	Blast-Like Karlin-Altschul + Permutation
BOOTSCAN	RDP4 + PHYLIP	+	+	+	Bootstrapping + binomial distribution + χ^2
Maximum χ^2	RDP4	+	+	+/-	X^2 + Permutation
CHIMAERA	RDP4	+	+	+/-	X^2 + Permutation
Sister Scan	RDP4	+	+	+	Permutation and Z-Test
3SEQ	RDP4	+	+	+	Exact test

Data from Martin et al.[52]

2.3 Aims

1. GenBank is to be screened for the presence of cryptic mosaic *aph(3')-IIa* similar sequences already deposited but not identified as of "mosaic" origin by the original submitter of the sequence. The retrieved datasets are to be analysed by a battery of algorithms programmed to identify recombination events in aligned sequences. The applied procedures comprise the online tool "T-REX" and the standard software package for recombination detection "RDP4". This approach should provide information about the potential of this antibiotic resistance marker gene to take part in genetic recombination processes in naturally occuring bacterial populations.

2. In the same way, GenBank is to be screened for cryptic mosaic sequences homologous to CP4 *epsps*. In this second bioinformatic analysis only those software tools that performed best during analysis of *aph(3')-IIa* homologs are to be employed.

3. Literature and GenBank databases are to be screened for already deposited and correctly identified mosaic genes with a focus on mosaic antibiotic resistane genes. This approach should provide an overview about the extent of the mosaic gene formation issue in naturally occurring bacterial populations.

2.4 *In silico* Detection of Partial Horizontal Gene Transfers Among *Aph(3')-IIa* Similar Gene Sequences

2.4.1 Introduction

Analysis of a potential contribution of *aph(3')-IIa* to the variability of the antibiotic resistance kinome is of relevance because this resistance determinant is one of the most prevalent antibiotic resistance genes used in genetic engineering and plant gene technology.[251, 252] This resistance element is shed into the environment by artificial means via laboratory and hospital discharges and through deliberate release and cultivation of transgenic organisms.[253] The aminoglycoside phosphotransferase enzyme APH(3')-IIa inactivates the critically important aminoglycoside antibiotics neomycin and kanamycin as well as paromomycin, butirosin, gentamicin B, and ribostamycin.[254, 255] The clinically critically important aminoglycoside amikacin was shown to be phosphorylated to some extent only under *in vitro* conditions.[256] Recombination of plant-derived *aph(3')-IIa* fragments - which may carry mutations or suffer from lesions when present as free DNA in the environment - with endogenous *aph(3')-IIa* homologs present in recipient bacterial populations may lead to the formation of phosphotransferases with a different antibiotic inactivation spectrum. This may give rise to the development of mutant antibiotic resistance determinants which may eventually spread to human or animal pathogens compromising antibiotic therapy of infections, increasing morbidity and mortality and putting additional stress on public health systems.

2.4.2 Materials and Methods

Collection of sequence data

Published *aph(3')-IIa* like sequences were collected from GenBank (www.ncbi.nlm.nih.gov/genbank/; last access: September 22th, 2014). The *aph(3')-IIa* gene from the *E. coli* transposon Tn5 (Accession number V00618, positions 151-945), also known as neomycin phosphotransferase gene (*nptII*), was used as reference sequence. This reference sequence of 795 nt length was searched against the bacterial non-redundant nucleotide collection (http://www.ncbi.nlm.nih.gov/nuccore/) and the database of reference genomic sequences (http://www.ncbi.nlm.nih.gov/refseq/). The discontiguous megablast algorithm was used with default settings (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Vectors, artificial sequences and models were excluded from the search to avoid spamming of the results with 100% identity hits originating from artificial cloning vectors.

Sequence alignments

Gene sequences producing BLAST matches of 60% - 99.9% identity over a region of more than 700 nucleotides were selected for further analysis and retrieved from GenBank. Multiple sequence alignments were prepared using the ClustalW algorithm implemented in Bioedit (http://www.mbio.ncsu.edu/bioedit/bioedit.html) to assemble appropriate sequence datasets for *in silico* recombination analysis.[257] The sequence difference count matrix option of Bioedit was used to determine pairwise nucleotide differences as a measure for the extent of divergence of the retrieved sequences. Careful sequence selection, trimming and alignment is pivotal to retrieve optimal results from bioinformatic recombination analyses.[52]

Phylogenetic tree construction

Phylogenetic trees representing the taxonomic affiliation and interrelatedness of the organisms carrying the retrieved *aph(3')-IIa* homologs were constructed using the CommonTree tool available in the NCBI taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). The Newick formatted tree created with CommonTree was visualized using the online tree viewer Phylodendron (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html).

Detection of recombination events in aligned sequence datasets

Recombination events in multiple sequence alignments were searched using the Recombination Detection Program Beta 4.36 package (RDP4). Seven of the recombination signal detection algorithms available as modules in RDP4 were employed: RDP[52], BootScan[245], MaxChi[14], Chimaera[246], GeneConv[247], SiScan[248], and 3Seq.[250] In the general settings for the RDP4 recombination detection procedure, the highest acceptable p-value was set to 0.05, the Bonferroni method was selected to correct for multiple comparisons and the entire process was run in permutational mode with 100 permutations. For the remaining parameters in the general RDP4 options defaults were retained. These defaults involved running PhylPro and LARD as secondary detection methods. Default settings were also retained for the options in the individual detection modules, except for MaxChi, where the specific window size was set to “variable”. These settings and analysis modules were chosen in accordance to common practice in literature.[258-265]

Recombination breakpoints in multiple sequence alignments were confirmed with the genetic algorithm for recombination detection (GARD)[266] available at the datamonkey server (http://www.datamonkey.org/).

2.4.3 Results

Collection of aph(3')-IIa homolog sequences by BLAST search of the GenBank database

To enable sequence comparison-based search for recombination events in *aph(3')-IIa* homologs, the GenBank database was blasted for sequences similar to the *aph(3')-IIa* gene from the transposon Tn5 of *E. coli*. In order to obtain a maximum of information about published genes related to *aph(3')-IIa*, the discontiguous megablast algorithm that detects matches in both long and short sequence spans was applied. In addition to the non-redundant nucleotide collection, we also searched the database of reference genomes in GenBank. Table 3 summarizes the 94 highest scoring hits from both searches, and Figure 18 shows the regions of *aph(3')-IIa* matched by these hits.

Table 3. The highest scoring BLAST hits of the *aph(3')-IIa* gene from *E. coli* transposon Tn*5*.

Hit Accession	Hit description (simplified)	BLAST match length (nt)	Identity in blast match (%)	Gene length (nt)	Total Identity (%)[a]	Short name (short name of representative)[b]
V00618	*Escherichia* Tn*5* neomycin phosphotransferase (*npt*II)	795	100	795	100	ECaph3_IIa
KC853434	*Escherichia coli* ACN001 plasmid pACN001-A	795	100	795	100	(ECaph3_IIa)
U32991	*Escherichia coli* mini-Tn5 kanamycin transposon	795	100	795	100	(ECaph3_IIa)
X64335	*Escherichia coli* plasmid pMM234 DNA	795	100	795	100	(ECaph3_IIa)
AB255435	*Escherichia coli* plasmid pO86A1 DNA	795	100	795	100	(ECaph3_IIa)
L11017	*Escherichia coli* Tn*5*Tac1	795	100	795	100	(ECaph3_IIa)
U00004	*Escherichia coli* transposon Tn*5*	795	100	795	100	(ECaph3_IIa)
KJ747960	*Enterococcus faecalis* strain 3EH plasmid pCQ-3EH	795	100	795	100	(ECaph3_IIa)
CP000744	*Pseudomonas aeruginosa* PA7	795	100	795	100	(ECaph3_IIa)
AB366441	*Salmonella enterica enterica* sv. Dublin pMAK2 DNA	795	100	795	100	(ECaph3_IIa)
JN983042	*Salmonella enterica enterica* sv. Heidelberg pSH111_227	795	100	795	100	(ECaph3_IIa)
HF570109	*Shigella sonnei* plasmid pDPT3	795	100	795	100	(ECaph3_IIa)
JX469830	Uncultured bacterium plasmid pG527	795	100	795	100	(ECaph3_IIa)
NZ_JH724146	*Bacteroides dorei* CL02T12C06 supercont1.15	795	100	795	100	(ECaph3_IIa)
GQ463143	*Vibrio cholerae* Mex1 integrating conj.elem. ICEVchmex1	793	100	795	100	(ECaph3_IIa)
AB702969	*Escherichia coli* pCss165Kan strain: 4266 delta cssB::Km	795	99.9	795	99.8	Escheric03
AF244993	*Vibrio cholerae* aph3' gene	795	99.7	795	99.7	Vibrioch01
X57709	*E. coli* Transposon Tn5 DNA for aphA-2 gene	795	99.7	795	99.7	Escheric02
NZ_DS995603	*Clostridium nexile* DSM 1787 Scfld7	795	99.7	795	99.7	Clostrid01
NZ_KB849231	*Acinetobacter johnsonii* CIP 64.6 acLZl-supercont1.2	795	99.6	795	99.6	Acinetob01
CP001096	*Rhodopseudomonas palustris* TIE-1	795	99	795	98.9	Rhodopse01
KF767856	*S.enterica enterica* sv. Typhimurium MRS_10/765 *npt*II	754	100	754	-	(ECaph3_IIa)
DQ449896	Uncultured bacterium clone K040 *npt*II -like gene	726	99.7	731	-	UncultK040
DQ449901	Uncultured bacterium clone K047 *npt*IIgene	731	99.5	728	-	UncultK047
DQ449899	Uncultured bacterium clone K048 *npt*II gene	731	99.5	728	-	UncultK048
DQ449898	Uncultured bacterium clone K001 *npt*II gene	731	99.5	728	-	UncultK001

Hit Accession	Hit description (simplified)	BLAST match length (nt)	Identity in blast match (%)	Gene length (nt)	Total Identity (%)[a]	Short name (short name of representative)[b]
JQ664666	*Riemerella anatipestifer* strain GN19 aph gene	717	100	717	-	RiemerGN19
JQ664661	*Riemerella anatipestifer* strain GN12 aph gene	717	100	717	-	(RiemerGN19)
JQ664660	*Riemerella anatipestifer* strain GN10 aph gene	717	100	717	-	(RiemerGN19)
JQ664653	*Riemerella anatipestifer* strain FN3 aph gene	717	100	717	-	(RiemerGN19)
JQ664647	*Riemerella anatipestifer* strain 3 aph gene	717	100	717	-	(RiemerGN19)
JQ664646	*Riemerella anatipestifer* strain 1-5 aph gene	717	100	717	-	(RiemerGN19)
DQ449903	Uncultured bacterium clone K002 *npt*II gene	731	99.3	728	-	UncultK002
DQ449900	Uncultured bacterium clone K003 *npt*II gene	731	99.3	728	-	UncultK003
DQ449897	Uncultured bacterium clone K036 *npt*II-like gene	731	99.3	729	-	UncultK036
JQ664680	*Riemerella anatipestifer* strain X21-3N aph gene	717	99.9	717	-	RiemerX213
JQ664673	*Riemerella anatipestifer* strain LQ30 aph gene	717	99.9	717	-	RiemerLQ30
JQ664672	*Riemerella anatipestifer* strain LQ26 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664670	*Riemerella anatipestifer* strain GN52 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664668	*Riemerella anatipestifer* strain GN26 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664665	*Riemerella anatipestifer* strain GN18 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664664	*Riemerella anatipestifer* strain GN16 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664662	*Riemerella anatipestifer* strain GN13 aph gene	717	99.9	717	-	RiemerGN13
JQ664658	*Riemerella anatipestifer* strain GN5 aph gene	717	99.9	717	-	(RiemerLQ30)
JQ664676	*Riemerella anatipestifer* strain LY37 aph gene	717	99.7	717	-	RiemerLY37
JQ664671	*Riemerella anatipestifer* strain JN2N aph gene	717	99.7	717	-	RiemerJN2N
JQ664663	*Riemerella anatipestifer* strain GN15 aph gene	717	99.7	717	-	RiemerGN15
JQ664659	*Riemerella anatipestifer* strain GN9 aph gene	717	99.7	717	-	RiemerGN09
JQ664657	*Riemerella anatipestifer* strain GN3 aph gene	717	99.7	717	-	RiemerGN03
JQ664655	*Riemerella anatipestifer* strain GN1 aph gene	717	99.7	717	-	RiemerGN01
JQ664649	*Riemerella anatipestifer* strain 8 aph gene	717	99.7	717	-	Riemer008
DQ449902	Uncultured bacterium clone K056 *npt*II gene	731	99.0	728	-	UncultK056
JQ664678	*Riemerella anatipestifer* strain W9 aph gene	717	99.6	717	-	RiemerW009
JQ664677	*Riemerella anatipestifer* strain T2 aph gene	717	99.6	717	-	RiemerT002
JQ664675	*Riemerella anatipestifer* strain LY18 aph gene	717	99.6	717	-	RiemerLY18
JQ664674	*Riemerella anatipestifer* strain LY6 aph gene	717	99.6	717	-	RiemerLY06
JQ664667	*Riemerella anatipestifer* strain GN22 aph gene	717	99.6	717	-	RiemerGN22
JQ664651	*Riemerella anatipestifer* strain 256 aph gene	717	99.6	717	-	Riemer0256
JQ664650	*Riemerella anatipestifer* strain 9 aph gene	717	99.6	717	-	Riemer0009
DQ449895	Uncultured bacterium clone K049 *npt*II gene	723	99.3	729	-	UncultK049
DQ449894	Uncultured bacterium clone K009 *npt*II-like gene	727	99.0	728	-	UncultK009
JQ664681	*Riemerella anatipestifer* strain X23-4N aph gene	717	99.4	717	-	RiemerX234
JQ664679	*Riemerella anatipestifer* strain X21-1N aph gene	717	99.4	717	-	RiemerX211
JQ664656	*Riemerella anatipestifer* strain GN2 aph gene	717	99.4	717	-	RiemerGN02
JQ664654	*Riemerella anatipestifer* strain FX2 aph gene	717	99.4	717	-	RiemerFX02
JQ664652	*Riemerella anatipestifer* strain C6 aph gene	717	99.4	717	-	RiemerC006
JQ664648	*Riemerella anatipestifer* strain 5 aph gene	717	99.4	717	-	Riemer0005
JQ664669	*Riemerella anatipestifer* strain GN27 aph gene	717	99.3	717	-	RiemerGN27
DQ449904	Uncultured bacterium clone K025 *npt*II-like gene	722	99.0	730	-	UncultK025

Hit Accession	Hit description (simplified)	BLAST match length (nt)	Identity in blast match (%)	Gene length (nt)	Total Identity (%)[a]	Short name (short name of representative)[b]
KC543497	*Pseudomonas aeruginosa* plasmid pOZ176	795	95	795	95.2	Pseudomo02
NZ_KI519248	*Pseudomonas aeruginosa* U2504 adgfx-supercont1.7	795	92	795	91.9	(Pseudomo14)
NZ_KI519246	*Pseudomonas aeruginosa* U2504 adgfx-supercont1.5	795	92	795	91.9	(Pseudomo14)
NZ_KI519240	*Pseudomonas aeruginosa* U2504 adgfx-supercont1.1	795	92	795	91.9	Pseudomo14
EF067857	*Escherichia coli* plasmid E99 aph	618	100	618	-	(ECaph3_IIa)
FN826652	Uncultured bacterium partial 16S rRNA gene US18.18	330	99.7	379	-	-
JQ937279	Uncultured bacterium aphA2 gene	347	97	347	-	-
CP008824	*Enterobacter cloacae* ECNIH2 plasmid pKEC-39c	750	74	795	72.3	Citrobac01
CP008790	*Klebsiella oxytoca* KONIH1 plasmid pKOX-86d	750	74	795	72.3	(Citrobac01)
CP007732	*Klebsiella pneumoniae* pneumoniae KPNIH27 pKEC-dc3	750	74	795	72.3	(Citrobac01)
CP007558	*Citrobacter freundii* CFNIH1 plasmid pKEC-a3c	750	74	795	72.3	Citrobac01
NZ_GG698326	*Staphylococcus aureus* aureus TCH130 SCAFFOLD169	235	100	235	-	(ECaph3_IIa)
HG938371	*Burkholderia cenocepacia* H111 chromosome 2	728	69	795	65.9	Burkhold03
AM747721	*Burkholderia cenocepacia* J2315 chromosome 2	730	69	795	65.5	Burkhold01
V00615[c]	Transposon Tn*5* left end	151	99.3	151	-	-
CP007509	*Pseudomonas stutzeri* strain 19SMN4	627	69	795	63.8	Pseudomo13
NZ_JH636049	*Saccharomonospora xinjiangensis* XJ-54 Sacxiscaffold_2	737	67	792	64.5	Saccharo01
CP000152	*Burkholderia* sp. 383	704	67	795	63.6	Burkhold02
GU721005	Uncult. bacterium plasmid clone mllc.F06 aph-like gene	111	98	197	-	-
CP002585	*Pseudomonas brassicacearum* subsp. brassicacearum NFM421	265	75	795	49.3	-
NZ_CM001512	*Pseudomonas fluorescens* Q8r1-96 chromosome	265	75	795	49.1	-
CP007236	*Ensifer adhaerens* OV14 chromosome 1 sequence	293	70	795	58.6	-
AY882987	*Sinorhizobium fredii* strain HH303-like gene	334	69	795	58.1	-
NZ_GG698343[c]	*Staphylococcus aureus* subsp. aureus TCH130 SCAFFOLD186	66	100	66	-	-
CP001111	*Stenotrophomonas maltophilia* R551-3	233	72	804	51	-

[a]Total identities with ECaph3_IIa are only given for complete full length genes, not for partial sequences.

[b]Short names were given to sequences selected for further analysis. From groups of 100% identical sequences one representative was chosen for further analysis. Short names of representatives are indicated between parentheses are given for all group members.

[c]Hits of short fragments, not shown in figure 18.

Search was performed against the non-redundant nucleotide collection and the database of genomic reference sequences.

Perfect 100%-matches were detected in various transposon sequences from *Escherichia coli,* in genomic sequences of *Pseudomonas aeruginosa, Vibrio cholerae* and *Bacteroidetes dorei* strains and in plasmids of *Escherichia coli, Enterococcus faecalis, Salmonella enterica* and *Shigella sonnei* strains. In addition perfect matches with several incomplete gene fragments, including one from a *Staphylococcus aureus* genome draft were observed. Figure 18 shows all perfect matches in red. Figure 19 gives an overview over the diversity of bacterial species that have been observed to carry the *aph(3')-IIa* gene.

19 full length homologs with high similiarity to *aph(3')-IIa* were detected in various bacterial genomes or plasmids. In Figure 18 these BLAST matches are marked in blue color. More than 99% identity was found in genomic sequences of *Escherichia coli, Vibrio cholerae*, *Clostridium nexile*, *Acinetobacter johnsonii* and *Rhodopseudomonas palustris.* 90 - 95% identity was found in plasmid and genomic sequences from *Pseudomonas aeruginosa*. 67 - 74% sequence similarity was found in plasmid sequences from *Enterobacter cloacae*, *Citrobacter freundii, Klebsiella pneumoniae* and *Klebsiella oxytoca* and in genomic sequences of *Saccharomonospora xinjiangensis, Pseudomonas stutzeri and* three *Burkholderia* strains*.*

Although the 67 - 74% similarity BLAST matches did not cover the 5' end of the *aph(3')-IIa* gene (Figure 18), they were located in open reading frames (ORFs) of 792 - 795 nt length with overall identities to *aph(3')-IIa* of 64 - 72% (Table 3).

The high scoring BLAST matches included two sets of partial sequences originating from bacterial population surveys specifically targeting *aph(3')-IIa* diversity. In the first survey *aph(3')-IIa* -like sequences have been amplified and sequenced from isolates of the avian pathogen *Riemerella anatipestifer*, that had been collected from diseased ducks in China between 2005 and 2010.[267] The second survey was a cultivation independent monitoring of *aph(3')-IIa* homologs in a Canadian river, which have been cloned and sequenced directly from water samples collected monthly in the years 2003 and 2004.[268] The sequences of both sets showed over 99% sequence identity with *aph(3')-IIa* (V00618), including several perfect matches. However, as these sequences had been produced by PCR amplification with primers binding within the *aph(3')-IIa* gene, sequence information was missing at their ends. In Figure 18 matches with the *Riemerella* dataset are shown in yellow, matches with the river dataset are shown in green.

Moreover, sequence identity of 97% and 98% with *aph(3')-IIa* was observed with two short sequence entries from PCR-based studies on antibiotic resistance genes in water (JQ937279) and activated sludge (GU721005). These short matches are presented in pink (Figure 18).

Finally, the discontiguous megablast search revealed numerous matches in genomic sequences of various bacterial species spanning 50 - 370 nucleotides of the *aph(3')IIa* gene between positions 360 and 720. The highest scoring matches of this type are shown in grey at the bottom of Figure 18. The region between positions 360 and 720 of the *aph(3')-IIa* gene contains two functional domains, known as motif1 and motif2, that are conserved across different clades of the aph gene family.[254] The full length of the ORFs producing these short matches varied between 783 and 813nt, and their nucleotide sequence identity with *aph(3')-IIa* over the entire stretch varied between 44% and 59%. These genes were considered more closely related to other aph genes than to *aph(3')-IIa.* They actually included hits identified as *aph(3')-IIc* gene of *Stenotrophomonas maltophilia* (HQ424460) and as *aph(3')-IIb* (X90856) gene of *Pseudomonas aeruginosa* (data not shown). The last hit presented in Table 3 and Figure 18 is an open reading frame of a *Stenotrophomonas maltophilia* strain (CP001111) with 97% sequence identity to *aph(3')-IIc*.

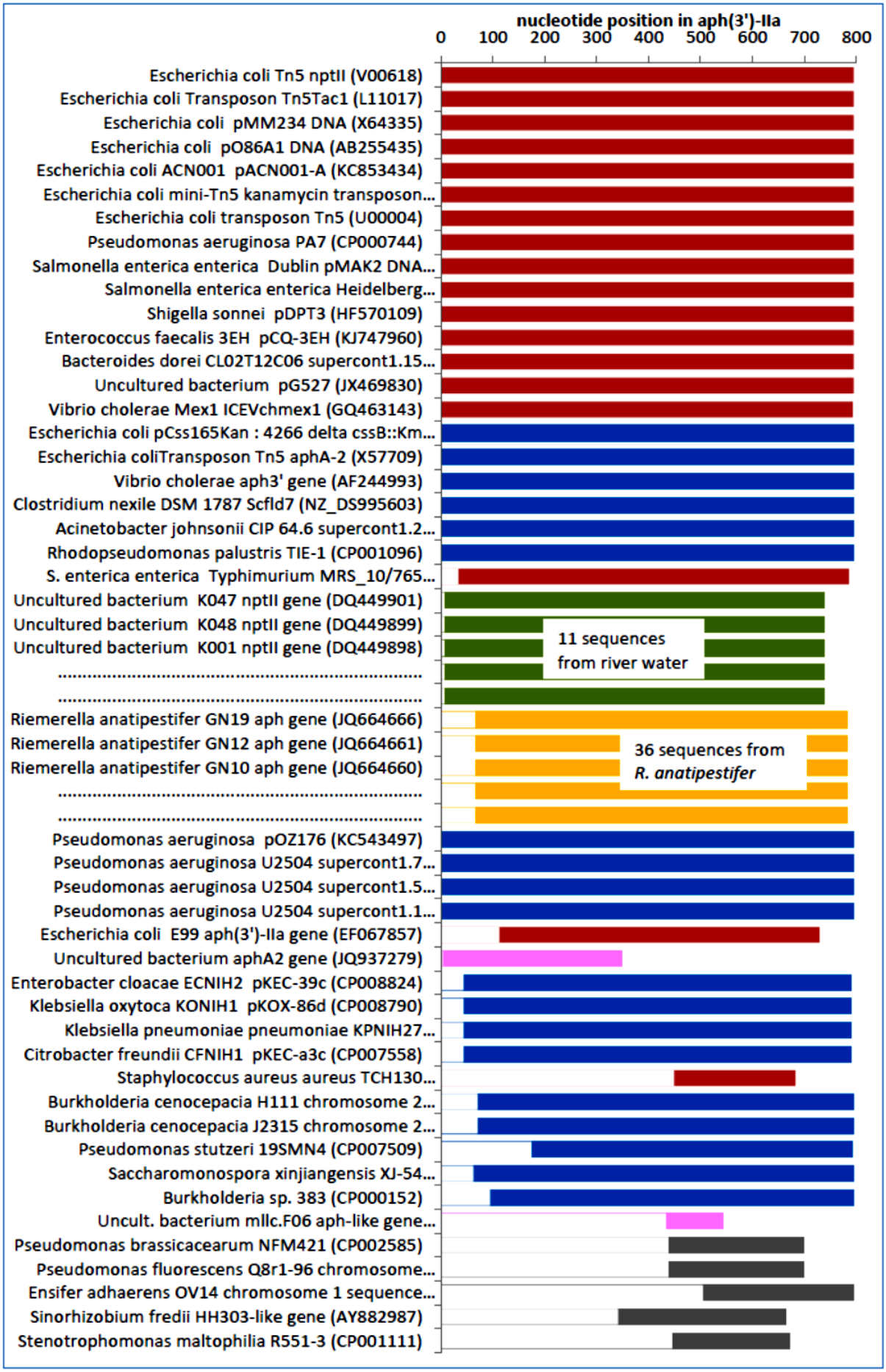


Figure 18: The highest scoring BLAST matches of the *aph(3')-IIa* gene with GeneBank entries.

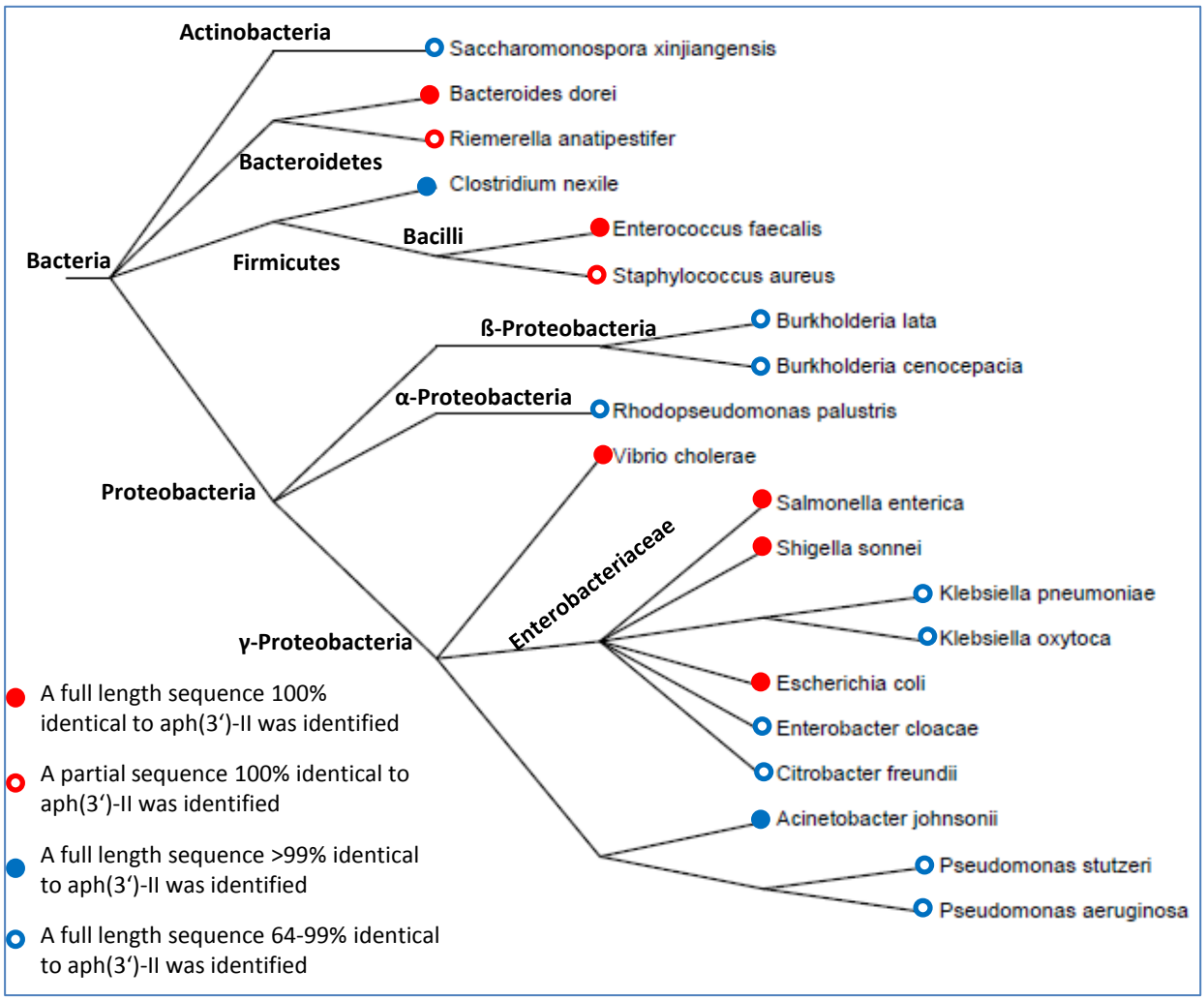


Figure 19: Phylogenetic tree of bacterial species carrying *aph(3')-IIa* homologs.

Selection of sequence sets for recombination event analysis

From the bulk of sequences collected from GenBank (Table 3), three sequence datasets were selected for recombination analysis:

i) Dataset 1: the 36 partial sequences from the *Riemerella anatipestifer* isolate collection, representing the intra-species variation of *aph(3')-IIa* homologs in a pathogen population of a single host species; depicted in yellow in Figure 18.

ii) Dataset 2: the 11 partial sequences from river water, representing the variation of *aph(3')-IIa* homologs occurring in any bacterial species in a defined natural aquatic environment; depicted in green in Figure 18.

iii) Dataset 3: the original *aph(3')-IIa* gene (V00618) plus the 19 homologs from various bacterial genomes and plasmids representing the entire variation of *aph(3')-IIa* genes known to date (i.e. as officially desposited in GenBank: shown in blue (Figure 18)).

Hits showing 100% nucleotide identity with *aph(3')-IIa* were ignored for analysis (i.e. not included into the datasets), because homologous recombination between identical sequences cannot be detected by bioinformatic sequence comparison. Divergent sequences, showing less than 60%

sequence identity to *aph(3')-IIa* were also excluded, because *in silico* recombination detection methods rely on accurate sequence alignments, which can hardly be achieved between very dissimilar sequences. Suboptimal sequence alignments may introduce artificially polymorphic sites into the alignments and lead to the detection of false positive recombination signals by the applied software algorithms. For this reason it is not recommended that datasets analyzed for recombination contain sequences that share less than 60% nucleotide sequence identity.[52] Moreover experimental evidence suggests that homologous recombination in bacteria requires at least 70% sequence similarity of the involved DNA fragments.[57, 205, 212]

For convenience, the sequences selected for recombination analysis were given short, 10-character names (Table 3). Importantly, the original *aph(3')-IIa* gene from the *E. coli* Tn5 transposon (V00618) used as "reference" gene, is referred to as "ECaph3_IIa" in Figures and Tables.

Tools for analysis of selected sequence datasets

We decided to use the well-established RDP4 software package that implements several recombination event detection programs, which rely solely on sequence comparison between the genes of interest. For additional validation of recombination breakpoints detected with RDP4, the genetic algorithm for recombination detection (GARD) was selected.[266]

RDP4 was preferred over the partial horizontal gene transfer detection tool of the T-REX online suite[13], which was initially intended to be used, because of certain shortcomings of this module: the T-REX partial HGT detection tool requires a reference tree describing the overall phylogenetic relationships between the organisms carrying the gene of interest. The analysis routines of T-REX rely on comparisons between "partial gene trees" constructed from different regions of the gene of interest and this phylogenetic reference tree. Such reference trees were unavailable for the uncharacterized *Riemerella anatipestifer* isolates (dataset 1) and for the uncultured river bacteria (dataset 2). For dataset 3 comprising *aph(3')-IIa* homologs from well described bacteria, a phylogenetic tree could be constructed based on the classification available from the NCBI taxonomy browser Figure 19. However, this information was insufficient for the T-REX based recombination analysis.

Sequence variation and recombination analysis in aph(3')-IIa homologs from Riemerella anatipestifer isolates (dataset 1)

Of the 36 sequences from *Riemerella anatipestifer* isolates, 25 were unique. One unique representative was selected from each group of identical sequences (Table 3). The most frequent sequence type (RiemerGN19) was identical with the *aph(3')-IIa* reference gene from the *E. coli* transposon Tn*5* (V00618). The sequences contained parts of the PCR primers that had been used by the survey authors.[267] After removal of the uninformative primer regions, a 686 nt gene segment, spanning *aph(3')-IIa* between position 85 and 770 remained for recombination analysis. ClustalW aligned the 25 sequences without introduction of gaps. The sequence variation in the alignment was low. In total there were 45 polymorphic sites in the sequence alignment. Table 4 shows the number of different nucleotides between each pair of sequences in the dataset. The highest pairwise nucleotide differences were 8 - 9 nucleotides.

Table 4. Pairwise nucleotide differences in dataset 1 (25 unique partial *aph(3')-IIa* homologs from *Riemerella anatipestifer* isolates).

	RiemerGN19	RiemerX213	RiemerLQ30	RiemerGN13	RiemerLY37	RiemerJN2N	RiemerGN15	RiemerGN09	RiemerGN03	RiemerGN01	Riemer0008	RiemerW009	RiemerT002	RiemerLY18	RiemerLY06	RiemerGN22	Riemer0256	Riemer0009	RiemerX234	RiemerX211	RiemerGN02	RiemerFX02	RiemerC006	Riemer0005	RiemerGN27
RiemerGN19	ID	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	5
RiemerX213	1	ID	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	6
RiemerLQ30	1	2	ID	2	3	3	1	1	1	1	1	2	2	4	2	4	4	2	5	5	3	5	3	5	4
RiemerGN13	1	2	2	ID	3	3	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	6
RiemerLY37	2	3	3	3	ID	4	4	4	4	4	4	5	5	5	5	5	1	5	6	6	6	6	6	6	7
RiemerJN2N	2	3	3	3	4	ID	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	6	7
RiemerGN15	2	3	1	3	4	4	ID	2	2	2	2	3	3	5	3	3	5	3	6	6	4	6	4	6	5
RiemerGN09	2	3	1	3	4	4	2	ID	2	2	2	3	3	5	3	5	5	3	6	6	4	6	4	6	5
RiemerGN03	2	3	1	3	4	4	2	2	ID	2	2	3	3	5	3	5	5	3	6	6	4	6	4	6	5
RiemerGN01	2	3	1	3	4	4	2	2	2	ID	2	3	3	5	3	5	5	3	6	4	4	6	4	6	5
Riemer0008	2	3	1	3	4	4	2	2	2	2	ID	3	3	5	3	5	5	3	6	6	4	6	4	6	5
RiemerW009	3	4	2	4	5	5	3	3	3	3	3	ID	4	6	4	6	6	4	7	7	5	7	5	7	6
RiemerT002	3	4	2	4	5	5	3	3	3	3	3	4	ID	6	4	6	6	4	7	7	5	7	5	7	6
RiemerLY18	3	4	4	4	5	5	5	5	5	5	5	6	6	ID	6	6	6	6	1	7	7	7	7	7	8
RiemerLY06	3	4	2	4	5	5	3	3	3	3	3	4	4	6	ID	6	6	4	7	7	5	7	5	7	6
RiemerGN22	3	4	4	4	5	5	3	5	5	5	5	6	6	6	6	ID	6	6	7	7	7	7	7	7	8
Riemer0256	3	4	4	4	1	5	5	5	5	5	5	6	6	6	6	6	ID	6	7	7	7	7	7	7	8
Riemer0009	3	4	2	4	5	5	3	3	3	3	3	4	4	6	4	6	6	ID	7	7	5	5	5	7	6
RiemerX234	4	5	5	5	6	6	6	6	6	6	6	7	7	1	7	7	7	7	ID	8	8	8	8	8	9
RiemerX211	4	5	5	5	6	6	6	6	6	4	6	7	7	7	7	7	7	7	8	ID	8	8	8	8	9
RiemerGN02	4	5	3	5	6	6	4	4	4	4	4	5	5	7	5	7	7	5	8	8	ID	8	6	8	7
RiemerFX02	4	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7	7	5	8	8	8	ID	8	8	9
RiemerC006	4	5	3	5	6	6	4	4	4	4	4	5	5	7	5	7	7	5	8	8	6	8	ID	8	3
Riemer0005	4	5	5	5	6	6	6	6	6	6	6	7	7	7	7	7	7	7	8	8	8	8	8	ID	9
RiemerGN27	5	6	4	6	7	7	5	5	5	5	5	6	6	8	6	8	8	6	9	9	7	9	3	9	ID

The four divergent sequences selected for subset recombination analysis are printed bold.

To estimate, whether this level of polymorphism was sufficient for recombination detection, we used the following formula provided by the developer of RDP4: *min* = (2 ln4X) / L; where *min* is the pair-wise nucleotide difference threshold below which no recombination between a pair of sequences can be reliably detected, X is the number of sequences in the dataset and L is the length of the alignment.[52] Apart from not being informative for recombination detection, groups of sequences having distances below this threshold reduce detection efficiency in the whole dataset. This is because multiple testing and the severeness of Bonferroni correction increase exponentially with the number of sequences. Therefore the developers of RDP4 recommend to retain only one representative from such groups of similar sequences.[52] Given a sequence length 686 nt RDP4 would require 13 nucleotides difference to detect recombination in a pair of sequences within a dataset of 25 sequences (Table 5). Pairwise differences of eight nucleotides difference would only be sufficient to detect recombination in a dataset of four sequences. It is important to note that these genetic distance thresholds are no strict exclusion criteria but rather a guiding principle to retrieve optimal recombination analysis results with RDP4.[52]

Table 5. Calculation of the optimal range of pairwise nucleotide differences in datasets for RDP4 analysis.

# of sequences (X)	Length (L) = 686 nt		Length (L) = 688 nt		Length(L) = 795 nt	
	min*	max*	min*	max*	min*	max*
3	7	206	7	206	6	239
4	8	206	8	206	7	239
5	9	206	9	206	8	239
6	9	206	9	206	8	239
7	10	206	10	206	8	239
8	10	206	10	206	9	239
9	10	206	10	206	9	239
10	11	206	11	206	9	239
11	11	206	11	206	10	239
12	11	206	11	206	10	239
13	12	206	11	206	10	239
14	12	206	12	206	10	239
15	12	206	12	206	10	239
16	12	206	12	206	10	239
17	12	206	12	206	11	239
18	12	206	12	206	11	239
19	13	206	13	206	11	239
20	13	206	13	206	11	239
21	13	206	13	206	11	239
22	13	206	13	206	11	239
23	13	206	13	206	11	239
24	13	206	13	206	11	239
25	13	206	13	206	12	239

The table displays the recommended limits of sequence divergence for retrieving optimal results for recombination analysis with RDP4. Sequences outside the cut off range should be excluded from the analysed datasets. Cut off values are dependent on the absolute number of distinct sequences in a given dataset (see first column: X).
* Minimum and maxiumum numbers of pairwise nucleotide differences were calculated as follows:
min = (2 x ln 4X) / length ; max = length x 0.3.[52]

Therefore, RDP analysis was first carried out with the entire dataset 1 containing 25 sequences, and then repeated with a subset comprising four of the most divergent sequences separated by at least eight different nucleotides and, thus, meeting the recommendations of the RDP4 developers (Table 4). However, RDP4 detected recombination neither in the 25 sequence dataset, nor in the four sequence set.

Sequence variation and recombination analysis in aph(3')-IIa homologs from river water (dataset 2)

All of the eleven *aph(3')-IIa* sequences from river water were unique. Sequence UncultUK40 was identical with the original *aph(3')-IIa* gene. After removal of PCR primers, a 688 nt gene segment, spanning *aph(3')-IIa* between position 27 and 714 remained for recombination analysis. ClustalW was able to align the eleven sequences without gaps. The number of pairwise nucleotide differences between the eleven sequences is shown in Table 6. The three most divergent sequences were separated by eight different nucleotides. Given a sequence length of 688 nt, the calculated pair-wise difference threshold was eleven nucleotides for a dataset of eleven sequences. Eight nucleotides

difference were only sufficient for pairwise recombination detection in a dataset of four sequences or less (Table 5).

Table 6. Pairwise nucleotide differences in dataset 2 (11 partial *aph(3')-IIa* homologs from river water samples).

	UncultK040	UncultK047	UncultK048	UncultK001	UncultK036	UncultK002	UncultK003	UncultK056	UncultK049	UncultK009	UncultK025
UncultK040	ID	1	1	1	3	2	2	4	1	4	4
UncultK047	1	ID	2	2	4	3	3	5	2	5	5
UncultK048	1	2	ID	2	4	3	3	5	2	5	5
UncultK001	1	2	2	ID	4	3	3	5	2	5	5
UncultK036	3	4	4	4	ID	5	5	7	4	7	7
UncultK002	2	3	3	3	5	ID	4	6	3	6	6
UncultK003	2	3	3	3	5	4	ID	6	3	6	6
UncultK056	4	5	5	5	7	6	6	ID	5	8	8
UncultK049	1	2	2	2	4	3	3	5	ID	5	5
UncultK009	4	5	5	5	7	6	6	8	5	ID	8
UncultK025	4	5	5	5	7	6	6	8	5	8	ID

The three most divergent sequences selected for subset recombination analysis are printed bold.

Therefore the same strategy as for dataset 1 was applied. First, the entire dataset 2 containing all eleven sequences was analyzed. In a second step a subset consisting of the three most divergent sequences and meeting the criteria recommended in (Table 5) was separately checked for recombination signals. As most of the methods implemented in RDP4 involve analysis of sequence triplets, alignments of three sequences represent the smallest possible dataset that can be analyzed. Again recombination was neither detected in the alignment consisting of eleven sequences, nor in the alignment with three sequences.

Sequence variation and recombination analysis in full length aph(3')-IIa homologs from various bacterial genomes and plasmids (dataset 3)

The final analysis included the orginal *aph(3')-IIa* gene and 19 full length homologs producing high scoring matches during the BLAST search (Figure 18). Among these 20 sequences, 15 were unique. The *aph(3')-IIa* homologs detected in plasmids of *Enterobacter cloacae*, *Klebsiella oxytoca, Klebsiella pneumoniae* and *Citrobacter freundii* (Table 3) were identical, and the sequence from *Citrobacter* was retained as representative for further analysis (Citrobac01). Moreover the three sequences from *Pseudomonas aeruginosa* U2504 were identical, and one was retained as representative (Pseudomo14). The original *aph(3')-IIa* gene (short name: ECaph3_IIa) was representative for 18 sequences producing perfect BLAST matches (Table 3). The 15 unique sequences comprised of 795 nucleotides, except for Sacharo01, which was three nucleotides (i.e. a coding triplet for one amino acid) shorter. ClustalW aligned the 15 sequences with introduction of a single three nucleotide gap in Sacharo01, accounting for the missing single amino acid. Table 7 shows the pairwise nucleotide distances between the 15 sequences. A cluster of six sequences including *aph(3')-IIa* (ECaph3_IIa) was separated from one another by only 1-10 nucleotides, whereas higher variation occurred among the remaining nine sequences. According to the estimations in, a pairwise distance of at least nine nucleotides is required to detect pairwise recombination in a 15 sequence set of 795 nt length. On the other hand sequence identities of less than 70%, i.e. pairwise nucleotide differences above 239 in

a 795 nt alignment (Table 5) may increase the risk of detecting false positives.[52] As these conditions were not met by all sequences in dataset 3, we first analyzed the entire 15 sequence set, and subsequently focused on a subset of five sequences, meeting the requirements of RDP4 recombination detection detailed in (Table 5).

Table 7. Pairwise nucleotide differences in dataset 3 (15 full length *aph(3')-IIa* homologs from various bacterial species).

	ECaph3_IIa	Escheric03	Escheric02	Vibrioch01	Clostrid01	Acinetob01	Rhodopse01	Pseudomo02	Pseudomo14	Citrobac01	Burkhold03	Burkhold01	Saccharo01	Pseudomo13	Burkhold02
ECaph3_IIa	ID	1	2	2	2	3	8	38	64	220	271	274	282	287	289
Escheric03	1	ID	3	3	3	4	9	39	63	219	271	274	282	286	289
Escheric02	2	3	ID	4	4	5	10	40	66	222	272	275	283	288	289
Vibrioch01	2	3	4	ID	4	3	8	40	66	222	271	274	283	287	290
Clostrid01	2	3	4	4	ID	3	8	40	66	222	273	276	282	289	288
Acinetob01	3	4	5	3	3	ID	5	41	67	223	273	276	282	289	291
Rhodopse01	8	9	10	8	8	5	ID	44	70	226	274	277	282	292	290
Pseudomo02	38	39	40	40	40	41	44	ID	45	226	276	279	277	288	286
Pseudomo14	64	63	66	66	66	67	70	45	ID	222	274	277	273	281	283
Citrobac01	220	219	222	222	222	223	226	226	222	ID	260	261	266	257	262
Burkhold03	271	271	272	271	273	273	274	276	274	260	ID	4	297	300	139
Burkhold01	274	274	275	274	276	276	277	279	277	261	4	ID	299	299	139
Saccharo01	282	282	283	283	282	282	282	277	273	266	297	299	ID	315	298
Pseudomo13	287	286	288	287	289	289	292	288	281	257	300	299	315	ID	324
Burkhold02	289	289	289	290	288	291	290	286	283	262	139	139	298	324	ID

The five sequences selected for subset analysis are printed bold.

For the 15 sequence dataset, seven recombination detection methods in RDP4 identified a recombination event among Pseudomo02, Pseudomo14 and a sequence out of the cluster around *aph(3')-IIa* (Table 8). The results suggested that Pseudomo02 was a mosaic of Pseudomo14 and *aph(3')-IIa* or a sequence highly similar to *aph(3')-IIa* Figure 20 . The seven methods congruently identified the exchange of a fragment in the region between alignment positions 100 and 500. However, the exact positions of the recombination breakpoints varied between the programs which is not an uncommon result. Figure 20 visualizes the recombination event and highlights the recombination breakpoints at positions 224 and 484, which were proposed congruently by three different methods (Table 8).

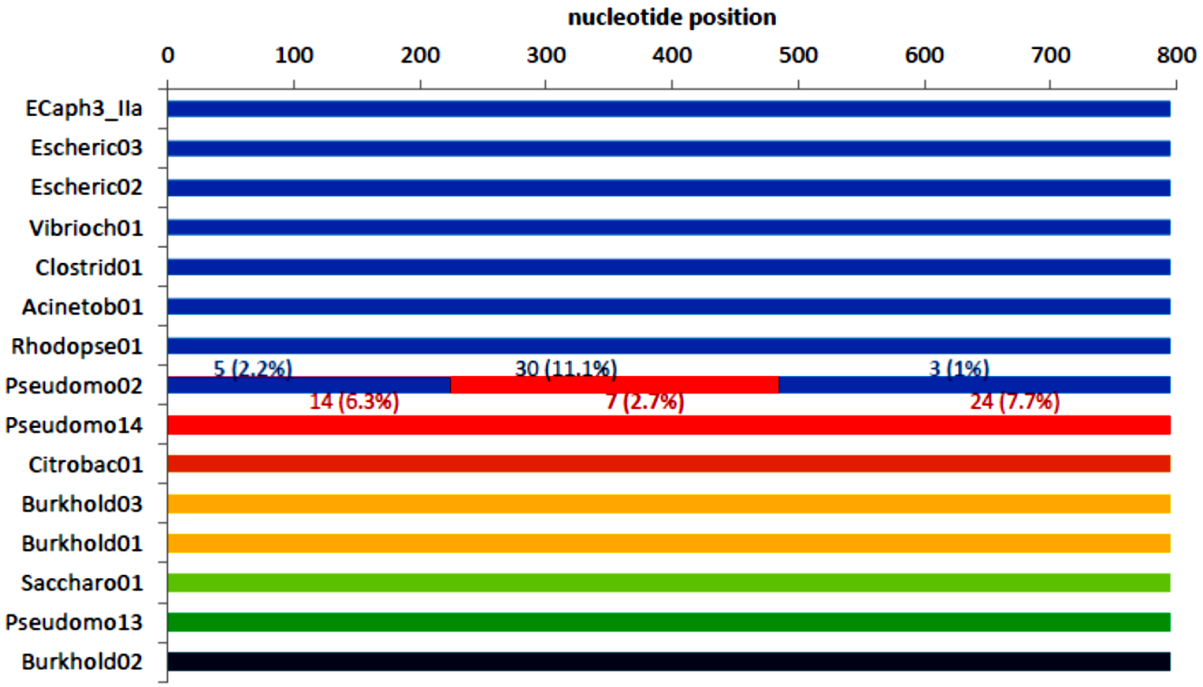


Figure 20: Detection of a recombination event in dataset 3 (15 full length *aph(3')-IIa* homologs) by RDP4.

The different colors indicate sequence identity and the likely origins of the segments in the recombinant sequence. For each segment of the recombinant, the number of sites different from those in the corresponding regions of the proposed parents is indicated: differences to ECaph3_IIa are indicated in blue above the segment, differences to Pseudomo14 are given in red below the segment.

Analysis of the five sequence dataset including only sequences with the recommended level of pairwise nucleotide differences confirmed the results obtained with the 15 sequence dataset (Table 8). The statistical support of the recombination event was higher than in the 15 sequence set, while the identification of the hybrid among the three sequences involved in this event was less conclusive. GenConv, MaxChi and 3Seq identified the first recombination breakpoint for the five sequence subset at a different position compared to the 15 sequence dataset. This was due to the fact that in the five sequence subset ECaph3_IIa had been selected to represent the cluster of six similar sequences (Table 7) while Escheric02 had produced the strongest recombination signals in the 15 sequence set.

Table 8. Detection of a recombination event in dataset 3 (15 full length *aph(3')-IIa* homologs from various bacterial species) and in a subset of 5 sequences with RDP4.

15 Sequence Dataset		5 Sequence subset	
Sequences	**Role in the recombination event**	**Sequences**	**Role in the recombination event**
ECaph3_IIa	Major Parent	ECaph3_IIa	Major Parent
Escheric03	Major Parent	Rhodopse01	Major Parent
Escheric02	Major Parent	Pseudomo02	Recombinant
Vibrioch01	Major Parent	Pseudomo14	Minor Parent
Clostrid01	Major Parent	Citrobac01	-
Acinetob01	Major Parent		
Rhodopse01	Major Parent		
Pseudomo02	Recombinant		
Pseudomo14	Minor Parent		
Citrobac01	-		
Burkhold03	-		
Burkhold01	-		
Saccharo01	-		
Pseudomo13	-		
Burkhold02	-		

Detection method	**Breakpoint positions**		**p-Value**	**Breakpoint positions**		**p-Value**
	begin	**end**		**begin**	**end**	
RDP	224	456	1.34E-02	224	456	2.64E-04
GENECONV	245	434	1.94E-02	275	434	4.26E-04
Bootscan	224	484	6.69E-03	224	484	1.22E-04
Maxchi	96*	552*	7.92E-06	185*	552*	1.18E-09
Chimaera	99*	434*	2.71E-03	114*	485*	8.92E-09
SiSscan	224	484	3.49E-07	224	484	1.84E-08
PhylPro			NS			NS
LARD			NS			NS
3Seq	98	484*	2.99E-08	214	484	2.46E-09

* The actual breakpoint position is undetermined (it was most likely overprinted by a subsequent recombination event).
Minor Parent = Parent contributing the smaller fraction of sequence.
Major Parent = Parent contributing the larger fraction of sequence.
NS = No significant p-value was recorded for this recombination event using this method.

For further confirmation the 15 sequence set and the five sequence set were analyzed with the genetic algorithm for recombination detection (GARD). GARD analysis of the 15 sequence set detected a single significant recombination breakpoint signal at position 198 (Table 9). Upon analysis of the five sequence subset, GARD produced several statistically non significant breakpoint signals, including one at position 482.

Table 9. Confirmation of the recombination event in dataset 3 (15 full length *aph(3')-IIa* homologs from various bacterial species) and in a subset of 5 sequences with GARD.

	Breakpoints	LHS p-value[a]	RHS p-value[b]	significance[c]
15 sequence dataset				
	198	8.80E-03	2.58E-02	**
5 sequence subset	32	1.50E-01	1.00E+00	N.S.
	348	2.28E-02	7.74E-01	N.S.
	482	2.80E-01	2.10E-02	N.S.

[a] LHS p-value that the partition left of this breakpoint has a topology different from that inferred from the partition on the right
[b] RHS p-value that the partition right of this breakpoint has a topology different from that inferred from the partition on the left
[c] only breakpoints with both p values < 0.05 are considered significant
** significant, N.S. not significant

2.4.4 Discussion

Sequence analysis of antibiotic resistance genes such as the penicillin binding proteins (pbp) or the tetracycline resistance genes has revealed that some of their variability and rapid evolution is due to intragenic recombination, also referred to as partial horizontal gene transfer.[15, 234, 237, 269-273] The aim of this work was to elucidate, whether intragenic recombination also occurs in *aph(3')-IIa*-type aminoglycoside resistance genes. To this end we have searched GenBank for sequence variation in *aph(3')-IIa* homologs, and we have applied specialized sequence analysis tools to investigate, whether this variation has likely arisen from intragenic recombination.

GenBank search revealed that *aph(3')-IIa* homologs identical to the Tn*5* reference sequence occur in members of several bacterial phyla and are located on both chromosomes and plasmids (Figure 19). This indicates that complete *aph(3')-IIa* gene sequences have been frequently transferred via horizontal gene transfer.

The large majority of the *aph(3')-IIa* homologs documented in GenBank were 99 - 100% identical to the *aph(3')-IIa* gene of the *E. coli* transposon Tn5, which had been used as reference sequence in our GenBank search. Only 13 sequences with identities to *aph(3')-IIa* in the range of 64 – 99% were retrieved from the database. Three of these 13 sequences were identical genes from the same genomic sequence (*Pseudomonas aeruginosa* U2504) and should therefore be regarded as a single database entry (Table 3). Thus, the hitherto described sequence diversity of known (i.e. already deposited in GenBank) *aph(3')-IIa* homologs is low compared to that of mosaic penicillin resistance determinants, and other genes for which partial horizontal gene transfer has been described.[15, 32, 56, 270, 274-276]

It should be noted, that the single available study directly targeting the diversity of *aph(3')-IIa* homologs in river water only recovered sequences that were more than 99% identical with the *aph(3')-IIa* reference sequence.[268] The observation of low *aph(3')-IIa* gene diversity in environmental samples is in accordance with our own observations on agricultural soil. Recently, we have pyro-sequenced thousands of *aph(3')-IIa* gene fragments from soil DNA, most of which were perfectly identical to the *aph(3')-IIa* reference gene (Wögerbauer et al., 2014; unpublished results).

A straightforward strategy to detect intragenic recombination is to analyze the gene of interest in bacterial populations that share a common habitat and therefore have had the spatial and temporal opportunity to exchange DNA *in vivo*. This type of approach has revealed the emergence of mosaic pbp2 genes in clinical isolates of *Neisseria gonorrhoeae*[15], or the partial pbp2B gene transfer between *Streptococcus pneumoniae* and its commensal *Streptococcus mitis and Streptococcus oralis.*[57] Therefore we set a focus on data from the two currently available population surveys on *aph(3')-IIa* diversity: The first sequence set originated from *Riemerella anatipestifer* isolates, collected from diseased ducks in China over a period of five years. The second set originated from PCR-based monitoring of a Canadian river environment over two years.[267, 268] DNA sequence variation occurred in both datasets and *in silico* translation revealed that some nucleotide polymorphisms were reflected in respective amino-acid polymorphisms (data not shown). However, this variation was too low to enable computational tracking of possible recombination events.

Recombination events in the evolution of a gene of interest can be also assessed by sequence comparison without the prerequisite that the source organisms have been isolated from a common habitat. For example Oggioni et al. have discovered mosaic patterns in tetracycline resistance genes by comparing previously published sequences of *Enterococcus faecalis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Ureaplasma urealyticum* and *Neisseria.*[270] Similarly Boc et al. have detected numerous recombination events in the evolution of the rubisco gene *rbc*L by comparison of amino acid sequences from various photosynthetic bacteria and algae.[13] In line with this concept we have compiled a dataset comprising the *aph(3')-IIa* reference sequence from the *E. coli* transposon Tn*5* and 14 unique full length *aph(3')-IIa* homologs showing 64 - 99% sequence similarity to *aph(3')-IIa*. The sequences originated from human pathogenic bacteria and from two biotechnologically relevant organisms, *Rhodopseudomonas palustris* and *Saccharomonosprora xinjangiensis*. Most of them have been recovered in genome sequencing projects without specific focus on aminoglycoside resistance.

Seven sequence comparison algorithms of the RDP4 suite found evidence for a recombination event that has occurred in the evolution of the *aph(3')-IIa* homologs in this dataset. The different methods agreed on the exchanged gene region and on the recombination partners involved in this event, but proposed different alignment positions as recombination breakpoints. This reflects the different aspects of information that these methods target in a sequence alignment.[52]

The risk of producing false positives, i.e. of mistaking mutation for recombination events, is inherent to any *in silico* recombination detection strategy.[13, 52] Therefore it is current practice to study recombination with several methods, including phylogeny-based and substitution distribution-based algorithms.[13, 277] The described recombination event in the *aph(3')-IIa* gene dataset is supported by three phylogeny-based methods (RDP, BootScan, SiScan), four substitution distribution-based methods (MaxChi, GeneConv, Chimera, 3Seq) and to some extent also by the phylogeny-based genetic algorithm (GARD). In bacterial sequence typing, a major application area of the RDP4 software, many authors have convened to disregard a recombination event as bioinformatic artifact, if it is detected by at least three methods with a Bonferroni-corrected p value <0,05.[263-265] These requirements absolutely apply to the recombination event described here.

One common source of false positives is high sequence variation, leading to misalignments and introducing an excess of variable sites into the alignment. The set of 15 *aph(3')-IIa* genes comprised sequences sharing less than 70% sequence identity and thus exceeded the maximum variation level recommended for RDP4 analysis.[52] However, we were able to demonstrate, that the detected recombination event was not an artifact caused by the presence of these deviant sequences in the

dataset: After elimination of sequences below 70% similarity the same event was detected with greater statistical significance. Statistical significance increased with the reduction of the dataset, because less severe correction for multiple comparisons is required.[52] In contrast, GARD could gather more information from the complete unique 15 sequence set than from the subset.

At a first glance, the presented results suggest, that recombination has occurred between *Pseudomonas aeruginosa* and one of the organisms in Table 3 harboring a homolog identical or highly similar to *aph(3')-IIa*. As *Pseudomonas aeruginosa* and most of the other identified carriers of *aph(3')-IIa* homologs are human pathogens, the recombination may have occurred in a clinical environment. However, computational analysis reveals recombination events that have occurred rather between *ancestors* than between the analyzed sequences themselves.[52] Given the obvious transferability of the entire *aph(3')-IIa* sequence (Figure 19), it is conceivable that the homologs sequenced in *Pseudomonas aeruginosa* have undergone whole gene horizontal transfer as well. Therefore it appears pointless to speculate about the organisms and environmental factors involved in the actual recombination event, at present.
Comparing the recombination history of *aph(3')-IIa* homologs (Figure 20) to that of notorious mosaic genes, such as *lkt*A (Figure 14), the recombination frequency among *aph(3')-IIa* homologs appears low. However, it remains to be clarified, whether *aph(3')-IIa* is actually less prone to genetic mosaicism. This may be the case, if there is strong selective pressure in the environment on a functional "wild type" gene and variation in the encoded aminoglycoside phosphotransferase enzyme does not confer much of a selective advantage. On the other hand, the scarceness of detected recombination events may also be due to limitations of the currently available sequence data. The number of sequences in GenBank that were informative for this recombination study was small. Most of the analyzed *aph(3')-IIa* homologs shared either more than 99% or less than 70% sequence identity and were therefore either too closely or too distantly related for recombination detection.[243]

The continuously growing database of bacterial genome sequences will likely reveal novel *aph(3')-IIa* homologs and improve our understanding of the variability and evolution of *aph(3')-IIa* genes. Presuming the discovery of additional variants sharing 80% - 98% sequence identity, a more precise picture on the recombination history of *aph(3')-IIa* homologs may be inferable in the near future.

2.4.5 Conclusions

From the present study we conclude that the variability of the currently known *aph(3')-IIa* homologs is low.

Furthermore we conclude, that there is evidence for recombination to have occurred in the evolution of recent *aph(3')-IIa* homologs. Therefore we consider *aph(3')-IIa* and its homologs susceptible to intragenic recombination.

We state a lack of studies targeting the diversity of *aph(3')-IIa* like genes in different natural environments and linking *aph(3')-IIa* gene sequence variations to antibiotic resistance patterns. This situation precludes conclusions on the phenotypical relevance of intragenic recombination in *aph(3')-IIa* homologs. Research in this field is urgently needed to improve the risk assessment of intragenic recombination regarding aminoglycoside antibiotic resistance development.

2.5 *In silico* Detection of Partial Horizontal Gene Transfers Among CP4 *Epsps* Similar Gene Sequences

2.5.1 Introduction

5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) is an enzyme in the shikimate pathway of bacteria, plants and fungi, where aromatic amino acids are formed (http://www.ebi.ac.uk/interpro/entry/IPR006264). The gene encoding EPSPS is called *aro*A. In plants *aro*A is located in the chloroplast genome.[278]

EPSPS is the cellular target of the herbicide glyphosate. With respect to glyphosate, EPSPS enzymes can be categorized in two classes. Class I EPSPS enzymes are inhibited by glyphosate, which competes for binding sites with the natural substrate phosphoenolpyruvate (PEP). In contrast, class II EPSPS enzymes are naturally tolerant to glyphosate. Class I EPSPS enzymes are present in wild type plants and in certain bacteria, including *Escherichia coli*. Class II EPSPS enzymes have been identified in various bacterial strains, many of which have been isolated from glyphosate contaminated environments.[279]

The most prominent class II *aro*A gene is that of *Agrobacterium* sp. CP4 ("CP4 *epsps*"), which is the primary transgene in glyphosate resistant plants.[279] Usually in glyphosate-tolerant transgenic crops the nucleotide sequence of CP4 *epsps* has been codon optimized for efficient expression in the plant. The codon optimized transgene ("coCP4*epsps*") has only 74% nucleotide sequence identity with the original CP4 *epsps* gene, while the protein sequences encoded by CP4 *epsps* and coCP4*epsps* are identical with the exception of one amino acid.

When fragments of CP4 *epsps* or coCP4*epsps* are released from transgenic plant tissues, they may be taken up by naturally transformable bacteria.[92] Subsequently intragenic recombination may lead to the emergence of novel mosaic variants of EPSPS. Sequence variation can alter affinity for glyphosate and PEP and thus increase or decrease glyphosate tolerance.[279]

Here we have analyzed the variability and recombination history of bacterial CP4 *epsps* homologs in order to understand their potential for mosaic gene formation.

2.5.2 Materials and Methods

Our research strategy was the same as for *aph(3')-IIa* homologs (section 2.4). Briefly, the non-redundant nucleotide collection and the database of bacterial reference genomes of GenBank were BLAST-searched for sequences homologous to CP4 *epsps* and coCP4*epsps*. All intact bacterial genes with more than 70% nucleotide sequence identity to CP4 *epsps* or coCP4*epsps* were collected from the database for recombination analysis with the RDP4 software.

2.5.3 Results

BLAST search recovered 531 complete bacterial genes with more than 70% identity to CP4 *epsps*, 144 of which were unique (Table 10). All of these CP4 *epsps* homologs originated from members of the order *Rhizobiales* (Figure 21) and all were annotated as *aro*A genes. The closest identified relative of CP4 *epsps* was the *aro*A gene of *Shinella* sp. DD12 (NZ_AYLZ02000022) with 87% sequence identity.

There was only one additional *aro*A gene from the genus *Agrobacterium* (CP000628). It showed 78% sequence identity to CP4 *epsps*.

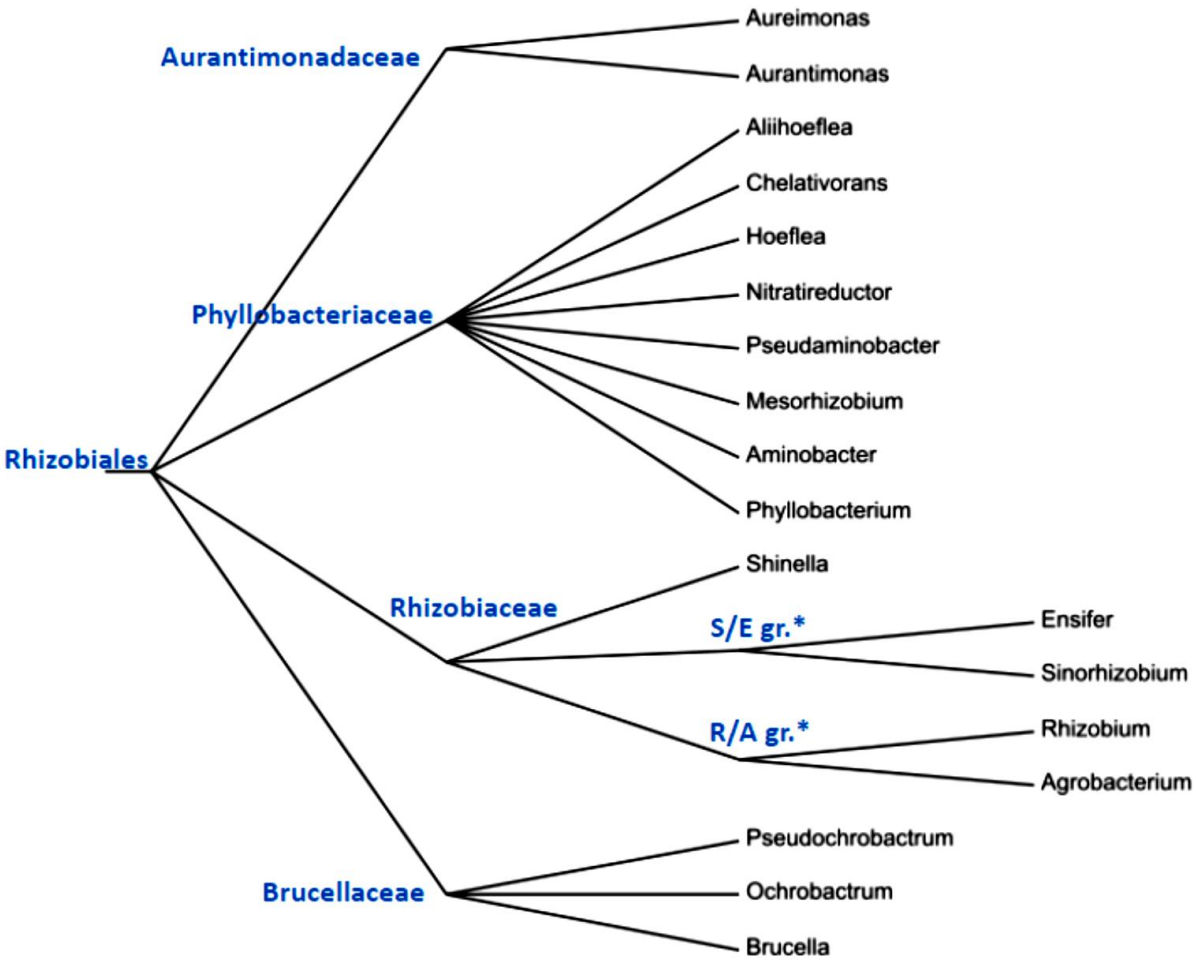


Figure 21: Phylogenetic tree of bacterial genera with *aro*A genes (>70% sequence identity to CP4 *epsps*).

* Subgroups of the family *Rhizobiaceae*: **S/E gr.**-*Sinorhizobium/Ensifer* group; **R/A-gr.** – *Rhizobium/Agrobacterium* group.

For the codon optimized transgene sequence coCP4*epsps* BLAST search yielded 137 hits above the sequence identity threshold of 70%. Twenty-five of these sequences were unique and all of them were more similar to the original CP4 *epsps* sequence than to coCP4*epsps* (Table 10). The closest identified relative of coCP4*epsps* was the *aro*A gene of *Sinorhizobium meliloti* WSM1022 (NZ_AZNW01000043) with 73.9% similarity to co4*epsps* and 83.4% similarity to CP4 *epsps*.

The 144 unique homologs of CP4 *epsps* were collected from GenBank (Table 10), aligned and compared to each other. The heatmap in Figure 22 visualizes the pairwise similarities among the 144 sequences. Sequence similarity was generally much higher within genera (typically 85%-99.9%) than between sequences from members of different genera (typically 70-80%). While GenBank did not harbor any cluster of sequences closely related to CP4 *epsps*, groups of highly similar sequences were available from strains of *Sinorhizobium*, *Rhizobium*, *Orchrobactrum*, *Brucella* and *Mesorhizobium*.

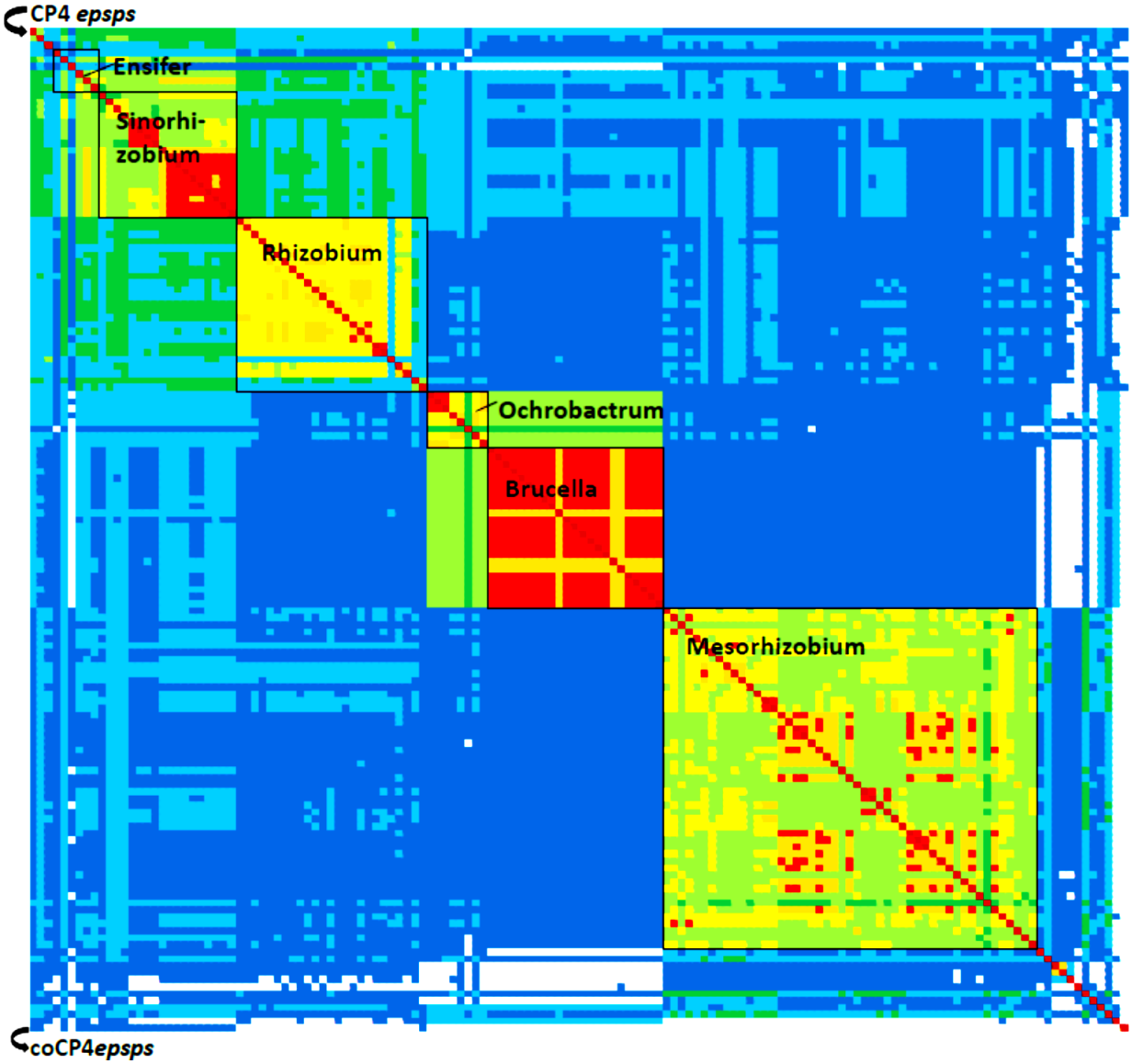


Figure 22: Heatmap showing pairwise sequence identities among the 144 *aro*A gene sequences.
Sequences are listed in Table 10 and plotted top-down and left-to-right in the same order as in Table 9. Pairwise sequence identities are color coded: ●100%, ●>99%, ●95-99%, ●90-95%, ●85-90%, ●80-85%, ●75-80%, ●70 - 75%, blank: <70%.

In a first step, recombination analysis was carried out separately for each of the genera *Ensifer*, *Sinorhizobium*, *Rhizobium*, *Ochrobactrum*, *Brucella* and *Mesorhizobium* (Table 11). Separate alignments were prepared for each genus, and the variable 3'-ends were trimmed back to a position, where all sequences overlapped unambiguously (Table 10). The recombination detection program RDP4 was employed as described in section 2.4. In addition to the settings detailed there, the option "automask for optimal recombination detection" was set to select one representative from each group of highly similar sequences. This step reduces the number of sequence comparisons and the severity of multiple testing correction.[52]

No recombination signals were detected within the genus *Ensifer* where intra-genus sequence variability was particularly high and within the genus *Brucella*, where intra-genus variability was particularly low (Table 11). Moreover no recombination signals were detected among the *Ochrobactrum* sequences. Recombination was detected, however, consistently by several recombination detection methods within the genera *Sinorhizobium* and *Rhizobium* (Table 11).

Among *aro*A gene sequences from the genus *Mesorhizobium*, recombination signals were reported by individual methods, but without significant statistical support. We suspected that this was due to the large size of the *Mesorhizobium* sequence set and to inefficient clustering by the "automask" function. We observed that "automask" had only grouped sequences sharing about 99% sequence identity, reducing the *Mesorhizobium* set from 49 to 37 sequences. Therefore we employed the nearest neighbor clustering method implemented in "mothur"[280] to group *Mesorhizobium* sequences at a similarity level of 92%. Nine different clusters were obtained and when nine sequences, representing these clusters were analyzed together in RDP4, significant evidence for one recombination event was detected by three methods (Table 11).

In a second step, we studied possible traces of inter-genus recombination events in the evolution of *aroA* genes. A data subset representing the entire diversity of CP4 *epsps* homologs was compiled by nearest neighbor-clustering the 144 sequences at a similarity level of 90%. Twenty nine clusters were obtained, seven of which consisted of single sequences sharing less than 70% identity with most of the others. The seven divergent clusters were excluded, and one representative sequence was retained from each of the remaining 22 clusters. RDP4 analysis of the 22 representative sequences revealed evidence for one recombination event within the Rhizobium/Agrobacterium group (Table 11). The *aro*A genes of *Agrobacterium radiobacter* K84 and *Rhizobium leucaenae* USDA9039 were identified as hybrids that had descended from the same recombinant ancestor (Figure 23). However, only one parental sequence, *Rhizobium* sp. NT-26, could be identified in the dataset. The same fragment exchange between *Rhizobium leucaenae* USDA9039 and *Rhizobium* sp. NT-26 had also been detected upon analysis of the *Rhizobium* dataset (Table 11, details not shown).

The recombination events detected by RDP4 were re-assessed using the Genetic Algorithm for Recombination Detection tool (GARD).[266] Several recombination breakpoints detected by RDP4 were confirmed by GARD (Table 11).

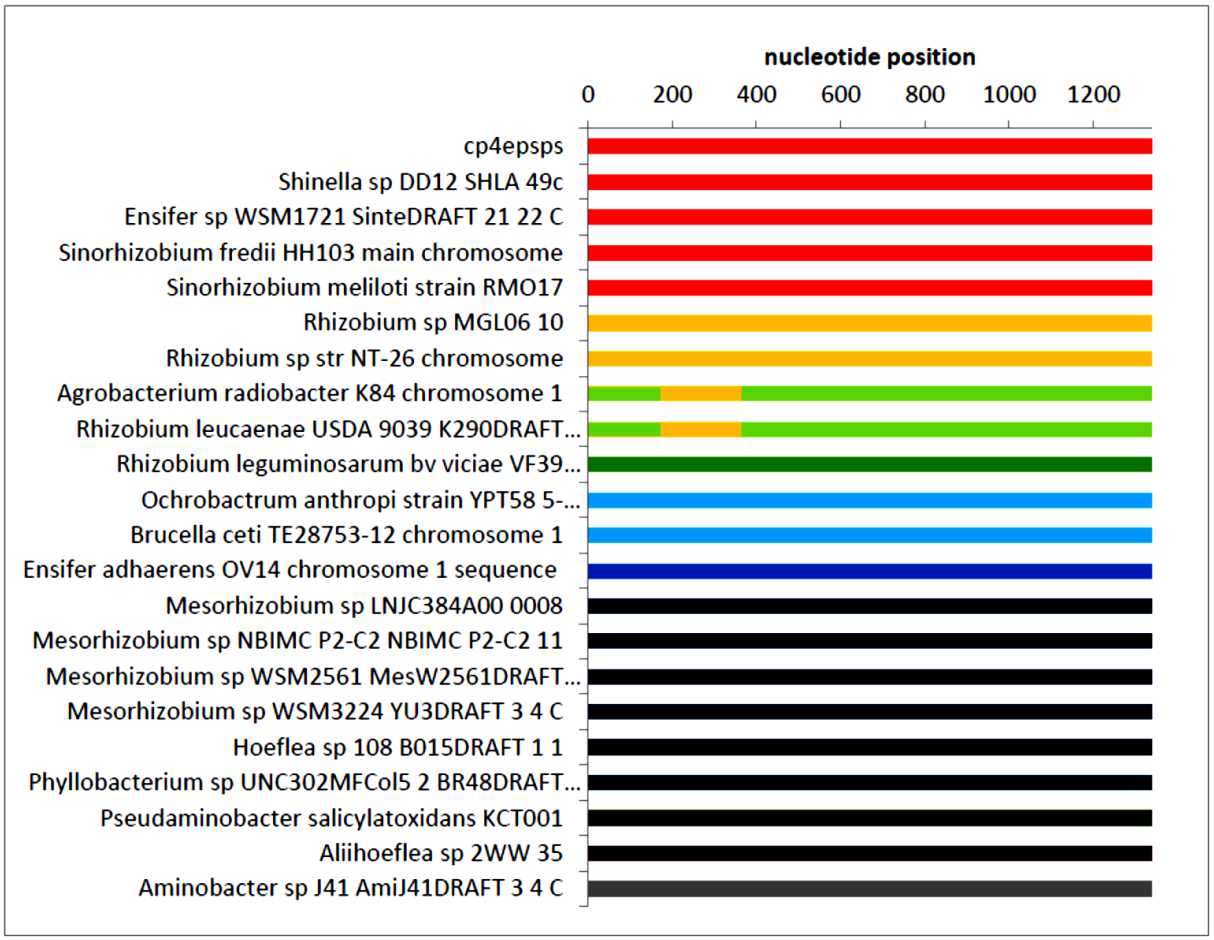


Figure 23: Detection of a recombination event among 22 *aroA* genes from different rhizobacterial genera.

The different colors indicate sequence similarity and the likely origins of the segments in the recombinant sequences.

Table 10. Sequences in GenBank with more than 70% sequence identity to CP4 *epsps**

Sequence description	**Accession number**	**gene length (nt)**	**identity with CP4 *epsps***	**identity with coCP4*epsps*[b]**
CP4 epsps (sequence from patent# 5633435)[a]	-	1368	100%	**73%**
Shinella sp. *DD12 SHLA 49c*	NZ_AYLZ02000022	1347	87%	**72%**
Agrobacterium radiobacter K84 chromosome 1	CP000628	1365	78%	68%
Ensifer adhaerens OV14 chromosome 1 sequence	CP007236	1344	75%	66%
Ensifer sp. *BR816 1*	NZ_KB905370	1347	83%	**71%**
Ensifer sp. *JGI 0001011-A08 H202DRAFT 2522129329 1 C*	NZ_AXWX01000001	1347	74%	66%
Ensifer sp. *TW10 SinTW10DRAFT 0 1 C*	NZ_AZNX01000001	1368	84%	**73%**
Ensifer sp. *USDA 6670 K291DRAFT 00013 13 C*	NZ_ATWE01000013	1368	83%	**74%**
Ensifer sp. *WSM1721 SinteDRAFT 21 22 C*	NZ_AZUW01000025	1368	84%	**73%**
Sinorhizobium arboris LMG 14919 SinarDRAFT 1 7 C	NZ_ATYB01000014	1371	82%	**73%**
Sinorhizobium fredii HH103 main chromosome	HE616890	1353	83%	**71%**
Sinorhizobium fredii NGR234	CP001389	1347	83%	**71%**
Sinorhizobium fredii USDA 257	CP003563	1347	83%	**71%**
Sinorhizobium medicae DI28 YU5DRAFT 7 8 C	NZ_ATTL01000008	1371	81%	**72%**
Sinorhizobium medicae WSM1369 A3C5DRAFT 0 1 C	NZ_AQUS01000307	1371	81%	**72%**
Sinorhizobium medicae WSM244 A3C7DRAFT 9 10 C	NZ_ATTR01000010	1371	81%	**72%**
Sinorhizobium medicae WSM419	CP000738	1371	81%	**72%**
Sinorhizobium meliloti 4H41 B075DRAFT 0 1 C	NZ_AQWP01000001	1368	84%	**73%**
Sinorhizobium meliloti AK11 393	NZ_AKZV01000393	1368	84%	**74%**
Sinorhizobium meliloti AK83 chromosome 1	CP002781	1368	84%	**74%**
Sinorhizobium meliloti BL225C	CP002740	1368	84%	**74%**
Sinorhizobium meliloti C0431A 379	NZ_AKZT01000379	1372	83%	**73%**
Sinorhizobium meliloti C0438LL 145	NZ_AKZS01000145	1370	83%	**73%**
Sinorhizobium meliloti CCNWSX0020 00079	NZ_AGVV01000079	1368	84%	**74%**
Sinorhizobium meliloti GVPV12 A3C9DRAFT 9 10 C	NZ_ATZC01000010	1368	84%	**74%**
Sinorhizobium meliloti strain RMO17	CP009144	1368	84%	**74%**
Sinorhizobium meliloti WSM1022 SINM1022DRAFT 38 39 C	NZ_AZNW01000043	1368	84%	**74%**
*Rhizobium gallicum bv gallicum R602*sp. *B028DRAFT 4 5*	NZ_KB890253	1359	78%	68%
Rhizobium leguminosarum bv phaseoli 4292 Rleg18DRAFT 1 1	NZ_KB905373	1359	78%	69%
Rhizobium leguminosarum bv phaseoli FA23	NZ_ATTN01000013	1359	78%	68%
Rhizobium leguminosarum bv trifolii CC278f Rleg6DRAFT RLE 4	NZ_KI912100	1359	78%	69%
Rhizobium leguminosarum bv trifolii SRDI565 RLEG16DRAFT	NZ_AQUD01000001	1359	78%	68%
Rhizobium leguminosarum bv trifolii SRDI943	NZ_AQUN01000001	1359	78%	68%
Rhizobium leguminosarum bv trifolii TA1 Rleg7DRAFT RLF 4	NZ_KE387142	1359	78%	68%
Rhizobium leguminosarum bv trifolii WSM1325	CP001622	1359	79%	68%
Rhizobium leguminosarum bv trifolii WSM1689	CP007045	1359	79%	69%
Rhizobium leguminosarum bv viciae 248 RLEG17DRAFT 1 7 C	NZ_ARRT01000006	1359	78%	68%
Rhizobium leguminosarum bv viciae chromosome	AM236080	1359	79%	69%
Rhizobium leguminosarum bv viciae GB30 A3A3DRAFT 27 28 C	NZ_ATTP01000028	1359	78%	68%
Rhizobium leguminosarum bv viciae TOM RLEG15DRAFT 1 6 C	NZ_AQUC01000006	1359	79%	69%
Rhizobium leguminosarum bv viciae UPM1131 A19QDRAFT 6 7	NZ_KE386608	1359	79%	69%

Sequence description	**Accession number**	**gene length (nt)**	**identity with CP4 *epsps***	**identity with coCP4*epsps*[b]**
Rhizobium leguminosarum bv viciae UPM1137 YSWDRAFT 0 1 C	NZ_ATYN01000001	1359	78%	68%
Rhizobium leguminosarum bv viciae Vc2 B052DRAFT 27 28	NZ_KB890835	1359	78%	68%
Rhizobium leguminosarum bv viciae VF39 YUSDRAFT 15 16 C	NZ_ATYQ01000016	1359	78%	68%
Rhizobium leguminosarum bv viciae Vh3 B054DRAFT 18 19	NZ_KB890961	1359	78%	68%
Rhizobium leguminosarum bv viciae WSM1455 Rleg5 2	NZ_JH719384	1359	79%	69%
Rhizobium leguminosarum bv viciae WSM1481	NZ_AQUM01000001	1359	79%	69%
Rhizobium leucaenae USDA 9039 K290DRAFT 00024 24 C	NZ_AUFB01000024	1368	78%	69%
Rhizobium sp. *JGI 0001003-A11 G071DRAFT 2519946120 6 C*	NZ_AXYS01000006	1359	78%	68%
Rhizobium sp. *JGI 0001019-L19 G075DRAFT 2523242459 39 C*	NZ_AXYP01000039	1359	78%	68%
Rhizobium sp. *MGL06 10*	NZ_JMQK01000009	1368	81%	70%
Rhizobium sp. *str NT-26 chromosome*	FO082820	1368	79%	**72%**
Ochrobactrum anthropi ATCC 49188 chromosome 1	CP000758	1353	77%	67%
Ochrobactrum anthropi CTS-325 1	NZ_JH605226	1353	77%	67%
Ochrobactrum anthropi strain YPT58 5-enolpyruvylshikimate	GU992200	1353	77%	67%
Ochrobactrum intermedium 229E 163	NZ_ASXJ01000080	1355	77%	67%
Ochrobactrum intermedium 2745-2 21	NZ_JFHY01000021	1353	78%	67%
Ochrobactrum rhizosphaerae SJY1 037	NZ_AZRT01000037	1353	74%	67%
Ochrobactrum sp. *EGD-AQ16 5*	NZ_AWEU01000045	1350	77%	67%
Ochrobactrum sp. *UNC390CL2Tsu3S39 BS36DRAFT 00001 1 C*	NZ_JOOD01000008	1353	76%	67%
Brucella abortus 07-0994-2411 adggd-supercont1 19	NZ_AXND01000019	1353	76%	67%
Brucella abortus 80/101 acsaX-supercont1 7	NZ_KB850375	1353	76%	67%
Brucella abortus biovar 1 str 9-941 chromosome I	AE017223	1353	76%	67%
Brucella abortus CNGB 759 acscu-supercont1 2	NZ_AQMB01000002	1353	76%	67%
Brucella abortus F6/05-2 acsaW-supercont1 10	NZ_AQMH01000010	1353	76%	67%
Brucella abortus strain BER chromosome 1	CP007682	1353	76%	67%
Brucella abortus strain ZW053 chromosome 1 sequence	CP009098	1353	76%	67%
Brucella canis ATCC 23365 chromosome I	CP000872	1353	76%	67%
Brucella ceti TE28753-12 chromosome 1	CP006898	1353	76%	67%
Brucella inopinata BO1 chromosome I VBI00041 1	NZ_ADEZ01000001	1353	76%	67%
Brucella melitensis biovar Abortus 5-enolpyruvyl shikimate	AF326475	1353	75%	67%
Brucella melitensis bv 1 str 16M chromosome 1	CP007763	1353	76%	67%
Brucella melitensis bv 2 str 63/9 chromosome 1	CP007789	1353	76%	67%
Brucella melitensis UK19/04 acqZI-supercont1 2	NZ_AQNN01000002	1353	76%	67%
Brucella microti CCM 4915 chromosome 1	CP001578	1353	76%	67%
Brucella ovis ATCC 25840 chromosome I	CP000708	1353	76%	67%
Brucella sp. *83/13 supercont1 1*	NZ_DS999649	1353	76%	67%
Brucella sp. *BO2 VBI00229 62*	NZ_ADFA01000134	1350	76%	67%
Brucella sp. *NVSL 07-0026 supercont1 15*	NZ_GG770510	1353	76%	67%
Brucella sp. *UK1/97 acqZp-supercont1 39*	NZ_KB850735	1353	76%	67%
Brucella suis bv 2 strain Bs143CITA chromosome I	CP007695	1353	76%	67%
Brucella suis F8/06-3 acxwN-supercont1 1	NZ_KB850910	1353	76%	67%
Brucella suis strain ZW046 chromosome 1	CP009096	1353	76%	67%
Mesorhizobium australicum WSM2073	CP003358	1359	75%	64%

Sequence description	**Accession number**	**gene length (nt)**	**identity with CP4 *epsps***	**identity with coCP4*epsps*[b]**
Mesorhizobium ciceri biovar biserrulae WSM1271	CP002447	1359	75%	65%
Mesorhizobium ciceri ca181 chromosome chr MciceriCa181 1	NZ_CM002796	1359	73%	63%
Mesorhizobium ciceri CMG6 MescicDRAFT 1 2 C	NZ_AWZS01000002	1359	75%	65%
Mesorhizobium ciceri WSM4083 MESCI2DRAFT 0 1	NZ_KI912610	1359	75%	65%
Mesorhizobium huakuii 7653R genome	CP006581	1359	75%	65%
Mesorhizobium loti CJ3sym A3A9DRAFT 25 26 C	NZ_AXAL01000027	1359	73%	64%
Mesorhizobium loti CJ3sym A3A9DRAFT 25 26 C.falseName	NZ_AUEL01000005	1359	73%	64%
Mesorhizobium loti MAFF303099 DNA	BA000012	1359	75%	65%
Mesorhizobium loti NZP2037 Meslo3DRAFT 1 1	NZ_KB913026	1359	75%	65%
Mesorhizobium loti R7A MesloDRAFT 1 1	NZ_KI632510	1359	75%	65%
Mesorhizobium loti R88b Meslo2DRAFT 1 1	NZ_KI912159	1352	73%	65%
Mesorhizobium loti USDA 3471 A3AUDRAFT 7 8 C	NZ_AXAE01000013	1359	75%	65%
Mesorhizobium opportunistum WSM2075	CP002279	1359	74%	65%
Mesorhizobium opportunistum WSM2075 chromosome	NC_015675	1357	74%	65%
Mesorhizobium sp. *L103C105A0 0009*	NZ_AYXF01000009	1357	73%	64%
Mesorhizobium sp. *L103C119B0 0016*	NZ_AYXE01000016	1359	73%	64%
Mesorhizobium sp. *L103C120A0 0011*	NZ_AYXD01000011	1357	73%	64%
Mesorhizobium sp. *L103C131B0 0009*	NZ_AYXC01000009	1357	73%	64%
Mesorhizobium sp. *L103C565B0 0002*	NZ_AYXB01000002	1357	73%	64%
Mesorhizobium sp. *L2C054A000 0010*	NZ_AYXA01000010	1359	74%	64%
Mesorhizobium sp. *L2C066B000 0015*	NZ_AYWZ01000015	1352	74%	64%
Mesorhizobium sp. *L2C067A000 0010*	NZ_AYWY01000010	1357	74%	64%
Mesorhizobium sp. *L2C084A000 0007*	NZ_AYWX01000007	1354	74%	64%
Mesorhizobium sp. *L2C089B000 0011*	NZ_AYWV01000011	1357	73%	64%
Mesorhizobium sp. *L48C026A00 0001*	NZ_AYWU01000001	1359	74%	64%
Mesorhizobium sp. *LNHC209A00 0002*	NZ_AYWT01000002	1359	75%	64%
Mesorhizobium sp. *LNHC220B00 0002*	NZ_AYWS01000002	1359	74%	64%
Mesorhizobium sp. *LNHC221B00 0001*	NZ_AYWR01000001	1352	74%	64%
Mesorhizobium sp. *LNHC229A00 0006*	NZ_AYWQ01000006	1359	75%	64%
Mesorhizobium sp. *LNHC232B00 0020*	NZ_AYWP01000020	1359	74%	64%
Mesorhizobium sp. *LNHC252B00 0023*	NZ_AYWO01000023	1359	74%	64%
Mesorhizobium sp. *LNJC384A00 0008*	NZ_AYWK01000008	1359	74%	64%
Mesorhizobium sp. *LNJC386A00 0005*	NZ_AYWJ01000005	1352	74%	64%
Mesorhizobium sp. *LNJC391B00 0001*	NZ_AYWI01000001	1352	73%	64%
Mesorhizobium sp. *LNJC399B00 0008*	NZ_AYWE01000008	1359	74%	64%
Mesorhizobium sp. *LNJC403B00 0001*	NZ_AYWD01000001	1357	73%	64%
Mesorhizobium sp. *LSHC412B00 0004*	NZ_AYWB01000004	1359	73%	64%
Mesorhizobium sp. *LSHC420B00 0003*	NZ_AYVY01000003	1359	74%	64%
Mesorhizobium sp. *LSHC422A00 0004*	NZ_AYVX01000004	1357	73%	64%
Mesorhizobium sp. *LSJC268A00 0005*	NZ_AYVO01000005	1359	74%	64%
Mesorhizobium sp. *LSJC280B00 0007*	NZ_AYVL01000007	1357	73%	64%
Mesorhizobium sp. *NBIMC P2-C2 NBIMC P2-C2 11*	NZ_AVBO01000011	1353	76%	65%
Mesorhizobium sp. *URHB0007 N550DRAFT 00001 1 C*	NZ_JIAO01000011	1359	73%	64%

Sequence description	**Accession number**	**gene length (nt)**	**identity with CP4 *epsps***	**identity with coCP4*epsps*[b]**
Mesorhizobium sp. *URHC0008 N549DRAFT 00001 1 C*	NZ_JIAP01000001	1359	75%	64%
Mesorhizobium sp. *WSM1293 MesloDRAFT 4 5*	NZ_KI911320	1359	74%	65%
Mesorhizobium sp. *WSM2561 MesW2561DRAFT 1 2 C*	NZ_AZUX01000002	1359	74%	64%
Mesorhizobium sp. *WSM3224 YU3DRAFT 3 4 C*	NZ_ATYO01000004	1359	74%	65%
Mesorhizobium sp. *WSM3626 Mesw3626DRAFT 6 7 C*	NZ_AZUY01000007	1352	73%	64%
Aliihoeflea sp. *2WW 35*	NZ_AYOD01000035	1341	72%	64%
Aminobacter sp. *J41 AmiJ41DRAFT 3 4 C*	NZ_JAGL01000004	1341	76%	65%
Aurantimonas coralicida DSM 14790 H536DRAFT 00014 14 C	NZ_ATXK01000014	1365	71%	62%
Aurantimonas manganoxydans SI85-9A1	NZ_CH672387	1365	71%	62%
Aureimonas ureilytica DSM 18598 B146DRAFT 0 1 C	NZ_ARQE01000001	1356	71%	60%
Chelativorans sp. *BNC1*	CP000390	1350	70%	64%
Hoeflea sp. *108 B015DRAFT 1 1*	NZ_KB890024	1353	74%	63%
Nitratireductor aquibiodomus NL21 = JCM 21793	NZ_BAMP01000071	1338	70%	63%
Phyllobacterium sp. *UNC302MFCol5*	NZ_JMLL01000010	1356	76%	67%
Pseudaminobacter salicylatoxidans KCT001	NZ_CAIU01000021	1359	75%	65%
Pseudochrobactrum sp. *AO18b 151*	NZ_ANNO01000149	1353	71%	65%
coCP4epsps (sequence 95 from patent EP2726493)[b]	JC491270	1368	73%	100%

*The non-redundant nucleotide collection of GenBank and the database of reference genomes were searched in October 2014 using discontiguous megablast. From groups of identical sequences only one representative has been listed. In total 531 sequences were obtained, 144 of which were unique and are listed in this table.

[a]CP4 *epsps* was used as query sequence for the BLAST searches.

[b]Sequences showing identities with coCP4 *epsps* greater than 70% are printed bold.

Table 11. Recombination detection in *aro*A genes in individual genera and in the whole dataset using RDP4

	Number of sequences[a]						p-Values for recombination events obtained with different methods						
Dataset	**GenBank entries**	**Unique Sequences**	**Automasked dataset**	**Clusters**	**Alignment[b] length (nt)**	**Exchanged region[c]**	**RDP**	**GENECONV**	**Bootscan**	**Maxchi**	**Chimaera**	**SiSscan**	**3Seq**
Ensifer	7	6	6	-	1347	-	-	-	-	-	-	-	-
Sinorhizobium	48	18	8	-	1340	454 - 1004	4,6E-02	-	-	3,1E-03	4,6E-03	7,7E-08	-
Rhizobium	30	25	23	-	1353	*18 - 1240	8,7E-05	3,2E-04	7,1E-06	2,1E-04	4,5E-04	-	-
						*203 – 406	-	-	2,7E-02	5,5E-03	-	5,8E-04	3,5E-04
						260 - 958	-	-	3,2E-02	1,2E-03	-	3,3E-07	1,9E-04
Ochrobactrum	10	8	8	-	1353	-	-	-	-	-	-	-	-
Brucella	346	23	5		1333	-	-	-	-	-	-	-	-
Mesorhizobium	76	49	37	9	1352	476 - 739	2,6E-02	-	-	4,2E-03	2,3E-04	-	-
						-							
All[d]	531	144	-	22	1340	172 - 363	8,7E-06	1,1E-03	2,3E-02	1,0E-04	2,1E-02	2,4E-04	-

[a] Since recombination detection with RDP4 is inefficient in redundant datasets, representative sequences were selected for recombination analysis as follows: The total of sequences retrieved from GenBank was de-replicated to retain only unique sequences. Highly similar unique sequences were grouped using the "automask" option of RDP4. In large and variable datasets "automask" was inefficient and nearest neighbor clustering was employed to group similar sequences; the *Mesorhizobium* dataset was clustered at a similarity level of 92%; the complete set of all 144 CP4 *epsps* homologs was clustered at a similarity level of 90%.

[b] The variable 3'-ends of the analyzed *aro*A genes were trimmed in each dataset back to a point, where all sequences overlapped. Therefore the analyzed datasets differed slightly in alignment length.

[c] Recombination breakpoints confirmed by GARD are underlined.

[d] "All" refers to the total of sequences in GenBank sharing more than 70% sequence identity with CP4 *epsps* (Table 10).

***** The asterisk signalizes that the actual breakpoint position is undetermined (it was most likely overprinted by a subsequent recombination event).

2.5.4 Discussion

Analysis of 144 bacterial *aro*A genes from GenBank revealed a pattern of sequence variation that reflected the overall phylogenetic relatedness of the source bacteria. All sequences sharing more than 70% sequence identity with CP4 *epsps* originated from members of the order *Rhizobiales*, and most of these sequences also shared 70% identity among each other. Moreover sequence similarity was higher among *aro*A variants from the same genus than between variants from different genera. This suggests, that compared to *aph(3')-IIa* homologs, *aro*A genes are rarely subjected to complete horizontal gene transfer and have mainly evolved via vertical inheritance and speciation (compare Figure 22).

However, we found evidence for partial horizontal gene transfer to have occurred within the genera *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* and between members of the *Agrobacterium/Rhizobium* group. From these observations we conclude, that the *aro*A gene is susceptible to intragenic recombination. On the other hand it has to be noted, that the statistical support of the recombination signals and the consensus among the detection methods was lower in the *aroA* sequence sets than in the *aph(3')-IIa* dataset presented in section 2.4.

The sequence information available in GenBank did not enable direct analysis of the recombination history of CP4 *epsps*, the *aro*A gene of *Agrobacterium* sp. CP4, which has been modified and inserted into transgenic plants. BLAST search yielded only one bacterial sequence sharing more than 85% sequence identity with CP4 *epsps*, and according to our experience sequence similarity above 85% is required for efficient recombination detection with RDP4. The lack of closely related sequences was unexpected, since GenBank contains genome sequences of numerous *Agrobacterium* strains. It was beyond of the scope of this study to elucidate, why only one *Agrobacterium* genome produced a significant match with CP4 *epsps*. The available *Agrobacterium* genome sequences may have been either incomplete or may carry very divergent *aro*A genes, such as class I EPSPS determinants.

Glyphosate tolerant transgenic crops usually carry coCP4*epsps*, a version of CP4 *epsps* adapted to the codon usage in green plants that shares only 74% nucleotide sequence identity with the original CP4 *epsps* gene. Homologous recombination at a sequence similarity level of 74% is very unlikely.[57, 205, 212] In theory, it is conceivable, that an *aro*A gene of an unrelated bacterium is by chance more similar to coCP4*epsps* than the "mother gene" CP4 *epsps*. Such genes would be prone to fragment exchange with the modified transgene. However BLAST search of GenBank's two largest bacterial sequence collections did not reveal any example of accidental sequence similarity to coCP4*epsps*.

The analyzed 144 *aro*A genes originated without exception from culturable bacterial isolates. Environmental sequences with more than 70% similarity to CP4 *epsps* are currently not available. While analysis of sequences from culturable organisms can reveal recombination in the evolutionary history of genes, it cannot provide information on the likelihood of recombination events in natural environments. The likelihood of recombination depends on numerous factors including the co-occurrence of sufficiently related gene variants (see section 1.5). As the majority of bacteria is not culturable[281], the *aro*A sequence collection from culturable strains cannot reflect the *aro*A sequence variation present in soil, in the human intestine or in any other environment.

2.5.5 Conclusions

In conclusion we recommend cultivation-independent analysis of *aro*A gene sequences in agricultural soils, human and animal feces and other habitats in contact with glyphosate resistant crops. This would be a straight forward way to assess, whether *aro*A genes sufficiently similar for fragment exchange with CP4 *epsps* or coCP4*epsps* are present in these environments. Furthermore, frequent *aro*A variants could be expressed to analyze the corresponding enzyme variants for glyphosate and phosphoenolpyruvate affinity as described by Wang et al.[279] Such enzyme activity assays would help to estimate the relevance of *aro*A sequence variation and recombination to glyphosate resistance in bacteria.

2.6 Screening of the GenBank Database for the Presence of Mosaic Genes

To obtain an overview over the extent of the mosaic genes issue a screening of the nucleotide database at NCBI (GenBank) was performed. However, it became obvious that a straight-forward approach by directly searching for the term “mosaic” or “mosaic gene” in the nucleotide database is unrewarding for two reasons:

i) GenBank datasets are lacking a special qualifier which would allow an easy identification of the sequences as of mosaic nature.

ii) GenBank entries are primarily supplied by the original submitter of the sequence and are therefore frequently affected by mis-classifications or incomplete annotations because original submitters are often no experts on the nomenclature of the gene or protein of interest. For example it is impossible to obtain a suitable dataset from deposited *aph(3’)-IIa* sequences by using this label in GenBank as a search term because the respective entries are usually classified as “aminoglycoside phosphotransferase” only or have a completely wrong designation at all.

Currently there are only two possibilities to retrieve comprehensive sequence information for mosaic genes:

i) obtain the accession numbers of previously described mosaic genes by a preceeding literature survey or

ii) use a reference sequence of the gene of interest as query sequence in a BLAST search for more dissimilar sequences (e.g. megablast) and analyse the retrieved sequences for potential recombination signals with bioinformatic tools like T-REX or RDP4 to establish their possible cryptic mosaic nature.

For the present study the first approach was chosen using the relevant literature references from 2000 to 2013 (expanded by additional references added due to personal experience of the authors and by recommendations of the external scientific advisor) as basis for screening of GenBank for the corresponding mosaic gene sequences.

The results as depicted in Table 12 confirm the notion that a broad variety of genes coding for virtually any possible cellular function may be target for recombinational rearrangements leading to the formation of mosaic structures. Mosaic genes are not only encountered in highly transformable bacterial species but can also be found in bacterial strains thought to be more or less refractory to DNA uptake from external sources in natural environments (e.g. *E. coli*, *Salmonella* etc...). The collected data confirm mosaic genes to be a general theme in bacterial evolution not restricted to certain types of genes or to certain bacterial hosts. Additional information including GenBank accession numbers can be found in Annex 1 (Table 28)

Table 12. Mosaic genes: functions and transformability of carrier species

Data are presented as available in the scientific literature (see Annex 1).

1) Transformability is shown according to references[89, 92, 130, 282]: +….naturally transformable; - ….naturally not transformable, + / -….some species are transformable some not; ?.... no information available

Gene	Function	Organism	Naturally transformable[1)]	Reference
adk	adenylate kinase housekeeping gene	*Neisseria gonorrhoeae* *N. meningitidis* *N. lactamica* *N. polysaccharea* *N. cinerea, N. mucosa* *N. pharyngis var. flava* *N. flavescens* *N. animalis*	+	283
*aro*E	shikimate dehydrogenase housekeeping gene	*Neisseria gonorrhoeae* *N. meningitidis*	+	200
*atp*D	ATP synthase housekeeping gene	*Rhizobium*	+	284
blp	bacteriocin	*Streptococcus pneumoniae*	+	285
		Enterococcus faecalis	-	
BoNT C/D	toxin	*Clostridium botulinum*	-	276, 286-292
*cag*A	pathogenicity factor virulence	*Helicobacter pylori*	+	293
*com*C *com*CD	competence	*Streptococcus pneumoniae* *S. mitis* *Streptococcus* spp.	+	274
*chv*I *chv*G	non coding	*Rhizobium meliloti* *Rhizobium* sp. *strain NGR234* *Rhizobium leguminosarum* *Agrobacterium rhizogenes*	+	294
cyl	cytolysin	*Enterococcus faecalis*	-	285, 295
		S. pneumoniae	+	
dca / *ppt*A	competence-associated protein / pilin phosphorylcholine transferase	*Neisseria*	+	296
dhps	housekeeping gene	*Neisseria meningitidis*	+	297
*dot*A	virulence	*Legionella pneumophila*	+	298
*dpn*I, dpnII	restriction/modification	*Streptococcus pneumoniae*	+	299
emm	immune response modulation	*Streptococus pyogenes*	?	57, 300
emm-like	immune response modulation	*S. pyogenes*	?	301
*fha*B	hemagglutinin	*Neisseria meningitidis*	+	296
*frp*B / *fet*A	iron-regulated outer membrane protein	*Neisseria*	+	296

Gene	Function	Organism	Naturally transformable[1)]	Reference
genome	non protein coding	*Helicobacter pylori*	+	302
genome	general	*bacteria*	+ / -	239
*gln*A	glutamine synthetase housekeeeping gene	*Neisseria meningitidis*	+	200
gnd	6-phosphogluconate dehydrogenase housekeeping gene	*E. coli*	-	303
*hmb*R	hemoglobin receptor	*Neisseria*	+	296
*hpu*A	hemoglobin-haptoglobin utilization protein	*Neisseria*	+	296
IgA protease *por*A	protease	*Neisseria* spp. *Streptococcus pneumoniae*	+	269
Intergenic region	undefined ORFs	*Escherichia coli*	-	304
*lgt*A	acto-N-neotetraose biosynthesis glycosyl transferase	*Neisseria*	+	296
*lgt*C	lipopolysaccharide biosynthesis protein	*Neisseria*	+	296
*lgt*D	lipopolysaccharide biosynthesis protein	*Neisseria*	+	296
*lgt*G	lipopolysaccharide glycosyl transferase	*Neisseria*	+	296
*lkt*A	leukotoxin virulence factor	*Mannheimia (Pasteurella) haemolytica*	?	36
*lkt*CABD operon	leukotoxin virulence factor	*Mannheimia (Pasteurella) haemolytica* *Mannheimia glucosida* *Pasteurella trehalosi*	?	305
LOS	lipooligo-saccharide biosynthesis region	*Campylobacter jejunii*	+	306
*lyt*A	autolysin	*Streptococcus mitis* *S. pneumoniae*	+	57
maf	adhesin	*N. lactamica*	+	296
*maf*B	adhesin	*Neisseria meningitidis*	+	296
*mef*E	resistance	*Streptococcus pneumoniae*	+	307
*mur*M	cell wall	*Streptococcus pneumoniae*	+	308
*gyr*A, *gyr*B	replication	*Mycobacterium tuberculosis*	-	309
*nad*A	putative adhesin / invasin	*Neisseria*	+	296
*nan*A	neuraminidase virulence	*Streptococcus pneumoniae*	+	57, 310
*nan*B	neuraminidase virulence	*Streptococcus pneumoniae*	+	311
*nan*C	neuraminidase virulence	*Streptococcus pneumoniae*	+	312
NIME	neisserial intergenic mosaic elements	*Neisseria*	+	296
NTNH	non-toxic-non-haemagglutinins virulence	*Clostridium botulinum*	-	313
opa	membrane protein	*Neisseria*	+	314

Gene	Function	Organism	Naturally transformable[1)]	Reference
opc	class 5 outer membrane protein	*Neisseria*	+	296
*omp*A	outer membrane protein	*Mannheimia (Pasteurella) haemolytica* *Mannheimia glucosida* *Pasteurella trehalosi*	?	315
*omp*A	outer membrane protein	*Haemophilus parasuis*	+	316
operon	gene regulation	*several*	+ / -	317
*opr*F	membrane protein	*Pseudomonas*	+	318
*par*C, *par*E, *gyr*A	topoisomerase 4	*Streptococcus mitis* *S. pneumoniae*	+	319
*pbp*1b	transpeptidase resistance	*Streptococcus pneumoniae*	+	320
*pbp*2	transpeptidase resistance	*Neisseria meningitidis* *Neisseria gonorrhoeae*	+	321-325
*pbp*2B	transpeptidase resistance	*Streptococcus pneumoniae*	+	326
*pbp*2X	transpeptidase resistance	*Streptococcus pneumoniae*	+	327, 328
pcn	bacteriocin	*Streptococcus pneumoniae*	+	285
*pen*A	resistance	*Neisseria gonorrhoeae*	+	329-332
*pen*A (pbp2)	resistance	*Neisseria gonorrhoeae*	+	333-336
*pgl*E	pilin glycosylation protein	*Neisseria*	+	296
*pgl*G	pilin glycosylation protein	*Neisseria*	+	296
*pgl*H	pilin glycosylation protein	*Neisseria*	+	296
*pgt*A / *pgl*A	pilin glycosylation protein	*Neisseria*	+	296
*pho*A	alkaline phosphatase	*Escherichia coli*	-	337
*pho*N	phosphatase	*Salmonella*	-	338
*pil*A	pilin biogenesis protein	*Gram negative bacteria*	+ / -	339
		Pseudomonas	+	
*pil*A	pilin biogenesis protein	*Clostridium perfringens*	-	340
*pil*C	pilin biogenesis protein	*Neisseria*	+	296
*pil*E	surface protein	*Neisseria gonorrhoeae*	+	341
*pil*S	surface protein	*Neisseria meningitidis*	+	296
pln	toxin	*Lactobacillus plantarum*		342
por	outer membrane protein; porin	*Neisseria gonorrhoeae*	+	343
*por*A	outer membrane protein	*Neisseria*	+	296
*psp*A	surface protein, immune system evasion	*Streptococcus pneumoniae*	+	32, 344-346
*psp*A	surface protein, immune system evasion	*Staphylococcus aureus* *Clostridium perfringens*	+ -	340
RNaseP	RNAseP	*Pyrococcus horikoshii ot3*	?	347

Gene	Function	Organism	Naturally transformable[1)]	Reference
*sia*D	capsule biosynthesis protein	*Neisseria*	+	296
*sfb*I	virulence factor	*Streptococcus pyogenes*	?	348
ska	streptokinase	*Streptococus pyogenes*	?	57, 349
skn	Streptokinase	*Streptococcus pyogenes*	?	349
*tbp*B	transferrin binding protein B	*Neisseria menigitidis* *N. sicca* *N. mucosa* *N. flava* *N. subflava* *N. cinerea* *N. flavescens* *N. polysaccharea*	+	350, 351
*tbp*BA operon	transferrin binding protein(s)	*Mannheimia haemolytica* *Mannheimia glucosida* *Bibersteinia trehalosi*	?	352
tet(M)	resistance	*Enterococcus faecalis*	-	270
		Streptococcus pneumoniae	+	
		Staphylococcus aureus	+	
		Ureaplasma urealyticum	?	
		Neisseria	+	
tet(O) *tet*(W)	resistance	*Mixed fecal samples*	+ / -	235
tet(O) *tet*(W)	resistance	*Megasphaera elsdenii*	?	273
tet(O/W/32/O/W/O) *tet*(W/32/O) *tet*(O/W)	resistance	*Bifidobacterium thermophilum,*	?	237
		Lactobacillus johnsonii	+	
tet(S/M)	resistance	*Streptococcus bovis*	+	272
transposon	resistance	*E. coli ESS1*	-	353, 354
		Enterobacter cloacae	-	
transposon	resistance	*Paracoccus marcusii*	?	355
*vac*A	toxin	*Helicobacter pylori*	+	356, 357
*vls*E	surface protein	*Borrelia burgdorferi*	?	358
wsp	surface protein	*Wolbachia*	?	359

3 Mosaic Gene Formation of Aminoglycoside Phosphotransferases in *Acinetobacter baylyi*

3.1 Summary

We have shown an anchor sequence mediated integration of the antibiotic resistance marker gene *aph(3')-IIa* (*npt*II) into a pre-existing aminoglycoside phosphotransferase gene (*aph(3')-Va*) leading to a characteristic alteration of the antibiotic resistance profile of an *Acinetobacter baylyi* recipient strain. The frequency of this homology-directed illegitimate recombination was approx. 10^{-7}; the length of the terminal region of microhomology was 10 – 12 nucleotides. Formation of *nptII* – *aph(3')-Va* mosaic patterns could not be demonstrated in our model system.

However, our results substantiate that anchor sequence-mediated gene transfer by homology-directed illegitimate recombination is of relevance considering the fact that most transgenic plant genes so far are of microbial origin and contain vector backbone sequences which could function as homologous recombination partners with chromosomal sequences already present in competent bacteria.

3.2 Background

The phosphotransferase gene *aph(3')-IIa* used in transgenic plants (or in other genetically modified organisms) as antibiotic resistance marker[252] may be available for DNA uptake by competent environmental bacteria upon envelope disintegration and chromosomal DNA release of the carrier cells.[30, 360] *Aph(3')-IIa* derived DNA fragments may be involved in recombination processes with similar genomic or episomal DNA regions already present in acceptor bacteria which are primed for the uptake of free DNA. If the insertion site represents the coding region of an aminoglycoside phosphotransferase different from *aph(3')-IIa*, a mosaic gene pattern will emerge potentially leading to aminoglycoside phosphotransferases with changed or expanded antibiotic inactivation profiles. Currently there is no soild experimental evidence available in favour or disproving this hypothesis. To close this knowledge gap an experimental approach for testing the potential of *aph(3')-IIa* to generate mosaic sequence patterns in related phosphotransferase genes was developed. The *aph(3')-IIa* gene was used as donor DNA and exposed to naturally transformable *Acinetobacter baylyi* cells harbouring an aminoglycoside phosphotransferase gene with a substantial grade of sequence similarity to *aph(3')-IIa*. This acceptor strain showed a distinct and characteristic aminoglycoside antibiotic resistance profile.

A donor DNA vector containing the *nptII* wild type sequence (GenBank # V00618) plus an *A. baylyi* homologous anchor region was constructed. This linear donor DNA molecule was used in natural transformation assays. If DNA uptake and recombination occurred, the transformed bacteria were expected to be identifiable due to significant changes in the minimum inhibitory concentration (MIC) or an altered resistance profile to aminoglycosides. Any transformants (i.e. growing colonies on selective agar showing MIC changes or an altered resistance profile) would then be subject to PCR and sequencing of relevant genome regions for further characterization of mosaic formations. As *nptII* fragment acceptor gene the aminoglycoside phosphotransferase *aph(3')-Va* (GenBank #

K00432) was chosen as model system to test for mosaic pattern formation induced by *nptII* derived DNA fragments *in vivo*.

As the applied aminoglycoside phosphotransferase genes are not naturally isolated from *A. baylyi*, it was considered that a stronger promoter could be needed to detect expression of the genes in *A. baylyi*. A previously constructed *aph(3')-Va* containing *A. baylyi* strain intended as receptor for *aph(3')-IIa* fragments in the *in vivo* model did not grow after re-cultivation. A thorough analysis of the genomic insertion region revealed the absence of an *aph(3')-Va* specific promoter in the region immediately preceding this gene. A completely new *aph(3')-Va* expression cassette containing a highly efficient phage T5 promoter was to be constructed. The resulting new *A. baylyi* acceptor strain was used for all subsequent transformation experiments using *aph(3')-IIa* containing fragments as DNA donor.

3.3 Aims

1. To provide information about the potential of the ARM gene *aph(3')-IIa* to form mosaics upon recombination with other aminoglycoside phosphotransferase genes an *in vivo* model for *nptII* uptake and recombination with similar chromosomal DNA regions in *Acinetobacter baylyi* was to be developed.

2. The resulting *A. baylyi* transformants were to be characterized for their antibiotic resistance profiles and the potential DNA insertion sites were to be sequenced to provide genetic evidence for the observed changes in the antibiotic resistance pattern of the transformed strains. The obtained sequences of the genomic region of interest should provide information about occurred partial gene transfers leading to mosaicism in the targeted *aph(3')-Va* gene.

3.4 *In vivo* Detection of Partial Horizontal Gene Transfers Among *Aph(3')-IIa* Similar Gene Sequences

3.4.1 Introduction

Natural genetic transformation plays a crucial role in the biology of important human pathogens like *Streptococcus pneumoniae* or *Neisseria meningitidis* by boosting genome variability allowing rapid adaptation to changing environmental conditions and quick evasion of selection pressure.[15, 79, 234, 361-363] Biologically relevant inter- and intraspecies transfer of DNA sequences was thought for a considerable period of time to encompass only the exchange of full length genes by homologous recombination.[76] This process usually does not result in the generation of novel sequence compositions but - rather conservatively - only in the redistribution of already existing genes between microorganisms.[73] However, it became clear that an interaction of two DNA molecules remarkably divergent in their nucleotide sequences is possible providing means to not only shuffle intact genes within bacterial populations but also to modify pre-existing coding sequences greatly enhancing the propensity for genetic variation.[239] Homologous recombination leads in these cases to the formation of mosaic genes.[14, 285, 307] These partial gene (or genetic) transfers are of significant biological relevance as exemplified by mosaic gene encoded altered penicillin binding proteins with reduced affinity to ß-lactam antibiotics in certain *S. pneumoniae* strains which constitute a serious threat for infected patients and public health.[58, 76, 364, 365]

The efficiency of homologous recombination drastically decreases in a log-linear relationship with increasing sequence dissimilarity of the involveld DNA molecules virtually dropping below the limit of detection at a sequence divergence above 30%.[205, 212, 271, 366] This observation is clearly restricting the capacity for mosaic gene formation to relatively similar nucleic acids interaction partners usually limiting potential DNA donors to closely related members of the same species or strain.[140, 367]

Similar and more dissimilar DNA sequences can also be chromosomally integrated by homology-directed illegitimate recombination.[56] This process relies on the presence of a single homologous anchor sequence in the donor DNA molecule which mediates a transient pairing between donor and receptor DNA. It was shown that this homologous anchor sequence can be as short as 153 bp for efficient integration of foreign – substantially dissimilar - DNA into the host genome of *S. pneumoniae*.[242] Another prerequisite for this process is the presence of a short segment of partial sequence identity in the foreign DNA region of the donor molecule with the usually chromosomal DNA receptor region termed "microhomology".[56] This region of microhomology can be as short as 3 – 10 identical nucleotides between donor and recipient and is the target for the formation of novel joints by illegitimate recombinations. The process leads to the insertion of a foreign DNA fragment released from any additional restrictions for sequence homology followed by the deletion of an endogenous region producing insertion-deletions (InDels).[56, 213]

Based upon already performed analyses the aminoglycoside phosphotransferase gene *aph(3')-Va* was selected as target for *nptII* containing DNA fragments for the following reasons:

The similarity of the APH(3')-IIa (= NPTII) aminoglycoside-modifying enzyme compared to the other aminoglycoside phosphotransferases at the amino acid level was in total below 61%.[254] An identity <57.7% in terms of nucleotide sequence could be established as compared by VectorNTI 7.1 analysis. There were no extended contiguous stretches of 100% homologous sequences between *aph(3')-IIa*

and all major other aminoglycoside phosphotransferases with published sequences. Only microhomologies were detected between *aph(3')-IIa* and these other aminoglycoside phosphotransferase genes. Based on these microhomologies the *aph(3')-Va* gene (GenBank # K00432) was chosen as potential *aph(3')-IIa* fragment acceptor. *Aph(3')-Va* has 50.2% of overall nucleotide identity with *nptII*, several homologous blocks of 9 -13 bp and two microhomology blocks of 35 bp (with 30 bp identical) and 14 bp (with 12 bp identical). An *aph(3')-Va* expression cassette containing a strong phage promoter was constructed by overlapping (fusion) PCR and transferred into *A. baylyi*.

Acinetobacter baylyi was chosen as a model system because it is a naturally competent and highly recombinogenic soil bacterium relevant for the study of plant-derived DNA – soil microbial community interactions.[368] The gram negative bacterium has simple culture requirements, a compact genome with a published sequence and can easily be genetically manipulated by the addition of linear DNA fragments to the culture medium. *A. baylyi* strains carrying the *aph(3')-Va* expression cassette were transformed with newly designed *nptII* donor DNA containing an appropriate anchor sequence homologous to an *A. baylyi* genomic region adjacent to the *aph(3')-Va* expression cassette insertion locus. The aminoglycoside resistance profile and the corresponding inhibitory concentrations of the resulting transformant zones using disc diffusion tests were determined. The genomic region putatively responsible for the alterations of the resistance status of the transformed cells was sequenced.

3.4.2 Materials and Methods

Construction of the nptII donor DNA

The donor DNA construct is constituted by the *nptII*$^{+}$ gene downstream of a homologous *Acinetobacter baylyi* derived DNA anchor (*lifO* gene). The expected size is 2043 bp.

a) Amplification of the *lifO* and *nptII*$^{+}$ genes containing a common restriction enzyme target site (EcoRI = GAATTC) with Phusion DNA polymerase (Thermo Scientific; Vienna, Austria). Plasmid pSDKH1 was used as template for the PCR amplification.

Reaction mix (total volume = 30 µl):

6 µl buffer HF (5x)

0.75 µl dNTPs (10 mM each)

0.3 µl each primer (100 µM)

0.3 µl Phusion (0.02 U/ml)

1.2 µl $MgCl_2$ (50 mM)

1 µl template

3 µl DMSO (100%)

17.15 µl water

Primers:

lifO: lifO-F = 5’ TCAGGCACAAGATTTATGGCA 3’

lifO-REcoRI = 5’ CGAATTCTCAAACCGCATATTATTTCC 3’

nptII^{+}: nptII-FEcoRI = 5’ CGAATTCATGATTGAACAAGATGGATTGC 3’

nptII-R = 5’ TCAGAAGAACTCGTCAAGAAGG 3’

PCR program:

98°C/30 sec

Cycles: 30

98°C/10 sec

58°C/10 sec

72°C/1 min

Final extention:

72°C/30 sec

b) Purification of the products with the Genejet PCR purification kit (Thermo Scientific).

c) Digestion of both amplicons with EcoRI (FastDigest; Thermo Scientific) at 37°C for 1 hour, followed by inactivation at 80°C for 5 min.
d) Purification of the products with the Genejet PCR purification kit.
e) Ligation of the products with 5U of T4 DNA ligase (Thermo Scientific) for 1 hour at 16.6°C followed by an inactivation of the ligase at 65°C for 10 min.
f) Amplification of the ligated product (donor DNA) with the expected size with primers and program as mentioned in a).
g) Purification of the donor DNA with the Genejet PCR purification kit.
h) Determination of the nucleotide sequence of the donor DNA.
The PCR product containing the donor DNA was sequenced by Eurofins (Ebersberg, Germany) with 4 primers, covering the full length of the construct.

Sequencing primers: lifO-F= 5' TCAGGCACAAGATTTATGGCA 3'

lifO-F2= 5' TCGGGTACAAGCATTTGAATC 3'

nptII-F = 5' GGATGATCTGGACGAAGAGC 3'

nptII-R = 5' AGGCTTTACACTTTATGCTTC 3'

Construction of a new Acinetobacter baylyi recipient strain – creation of an aph(3')-Va expression cassette by overlapping PCR

The expression cassette was assembled from four separate PCR amplicons*:*

Fragment A: *lifO* (homologous region with *A. baylyi* BD413)

Fragment B: phage promoter pT5

Fragment C: *aph(3')-Va*

Fragment D: *lipA* (homologous region with *A. baylyi* BD413)

a) Production of the 4 amplicons (fragments A - D) in single PCR reactions.
The reaction mix was the same as described above.

Primers:

A11 = 5' ATTATAGCAACCAGTCCAGGGCAG 3'

A12 = 5' TCAAACCGCATATTATTTCCGTTAAAAC 3' → Fragment A (986 bp)

B21 = 5' GTCGAGAAATCATAAAAAATTTATTT 3'

B22 = 5' GGTTAATTTCTCCTCTTTAATGAAT 3'→ Fragment B (118 bp)

C21 = 5’ ATGGACGACAGCACGTTGCGCCGGAA 3’

C22 = 5’ AGCCCGGTACCTCGCCGCCAACCCGTA 3’→ Fragment C (974 bp)

D21 = 5’ CTCTGTCGGTGCAGCGTATAGTCTATC 3’

D22 = 5’ TGTGTTGGTGTTGAGCGTGTTTC 3’→ Fragment D (622 bp)

PCR program:

98°C/30 sec

Cycles: 30

98°C/10 sec

58°C/10 sec

72°C/1 min

Final extention:

72°C/2 min

b) Purification of the amplicons with Genejet gel extraction kit (Thermo Scientific).
c) Amplification of fragments A to D with the overlapping primers. The reaction mix was the same as described above.

Primers:

A11 (see above)

A12-B = 5’ AAGCAAATAAATTTTTTATGATTTCTCGACTCAAACCGCATATTATTTCCGTTAAAAC 3’

B21-A = 5’ TTAATTTAGCTTTACCCGTGTTTTAACGGAAATAATATGCGGTTTGAGTCGAGAA ATCATAAAAAATTTATTT 3’

B22-C = 5’ GCGGGTACTTCCGGCGCAACGTGCTGTCGTCCATGGTTAATTTCTCCTCTTTAATG AAT 3’

C21-B = 5’ CAGAATTAATTCATTAAAGAGGAGAAATTAACCATGGACGACAGCACGTTGCGCC GGAA 3’

C22-D = 5’ GGAAAGATTGCATTAAATTTACCAGCACCTTCAGTGGATAGACTATACGCTGCAC CGACAGAGAGCCCGGTACCTCGCCGCCAACCCGTA 3’

D22-C = 5’ TTCCGCGGCTTGGCCGGAGCCGTCAGAGGCCGTGGTACGGGTTGGCGGCGAGGT ACCGGGCTCTCTGTCGGTGCAGCGTATAGTCTATC 3’

D22 (see above)

Fragment A' (1016 bp) = A11 + A12-B

Fragment B' (171 bp) = B21-A + B22-C

Fragment C' (1070 bp) = C21-B + C22-D

Fragment D' (684 bp) = D22-C + D22

The PCR program was the same as above.

d) Purification of the amplicons with the Genejet gel extraction kit.

e) Overlapping PCR:

- Amplification of product A' + B' (1138 bp) using Phusion DNA polymerase with primers A11 + B22-C

- Amplification of product C' + D' (1630 bp) with Phusion DNA polymerase with primers C21-B + D22

The reaction mix was the same as described above.

The first PCR program was performed without primers.

PCR program:

98°C/30 sec

Cycles: 15

98°C/10 sec

60°C/10 sec

72°C/1 min

Final extention:

72°C/2 min

Addition of 0.3 µl of each primer 100 µM, followed by

98°C/30 sec

Cycles: 15

98°C/10 sec

60°C/10 sec

72°C/1 min

Final extention:

72°C/3 min

- Purification of both amplicons with the Genejet gel extraction kit.

- Amplification of product A'B' + C'D' (=AD, 2700 bp) with Phusion Hot Start II (Thermo Scientific) with primers A11 + D22

Reaction mix (total volume = 20 µl):

10.8 µl water

4 µl buffer HF 5x

0.4 µl dNTPs 10 mM each

2 µl of each template

0.6 µl DMSO 100%

0.2 µl Phusion Hot Start II 2U/µl

The first PCR program was performed without primers.

PCR program:

98°C/30 sec

Cycles: 15:

98°C/10 sec

72°C/90 sec

Final extention:

72°C/5 min

Addition of 0.2 µl of each primer 100 µM, followed by

98°C/30 sec

Cycles: 20:

98°C/10 sec

60°C/10 sec

72°C/90 sec

Final extention:

72°C/5 min

- Purification of the amplicon by gel extraction with the Genejet kit.

- Presence of all fragments in the overlapped product AD was checked by PCR with primers A11, A12, B21, B22, C21, C22, D21, D22 as described above.

f) Natural transformation of *A. baylyi* BD413 with the *aph(3')-Va* containing expression cassette consisting of fragments A-D

A. baylyi BD413 was naturally transformed by liquid transformation at 30°C with the PCR amplicon AD (= *aph(3')-Va* expression cassette). Briefly, 5 ml culture of the recipient *A. baylyi* BD413 was prepared from one single colony, and incubated at 30°C with agitation at 150 rpm overnight (ON). The following day, a 1:100 dilution was done in a final volume of 5 ml and aerated (150 rpm) at 30°C for approx. 6 hrs. Cells were pelleted by centrifugation for 10 min at 4000 rpm, and resuspended in 4 volumes of fresh LB; each ml should contain 2.5 x 10^8 cells. Aliquots with 1 ml were transferred to glass tubes and PCR amplicon AD was added, the solution aerated (150 rpm) at 30°C for 90 min, and plated in LB with neomycin (20 µg/mL). Plates were incubated at 30°C. Transformed cells were checked 2 days later. A negative control, where water replaced the DNA in the transformation assay, was also performed and selected in plates with the same concentration of antibiotic.

g) Determination of the antimicrobial susceptibility of the transformants

The antimicrobial susceptibility of selected transformants to neomycin (10 µg) and kanamycin (30 µg) (Oxoid; Basingstoke, United Kingdom) was detected by the disc diffusion method.

The minimal inhibitory concentration (MIC) for kanamycin was determined by Etest (Biomerieux; Marcy-l'Etoile, France).

h) Determination of the nucleotide sequence of the *aph(3')-Va* expression cassette fragment AD in one of the transformants

The fragment AD incorporated in the genome of transformant AR7 was sequenced (by Eurofins).

Three fragments, covering the full length of the construct, were amplified with primers A11 + B22 (fragment AB), B21 + C22 (fragment BC), and C21 + D22 (fragment CD). The recipient cell *A. baylyi* BD413 was the negative control and the donor DNA was the positive control. The amplicons obtained from AR7 were sequenced with the same primers.

Natural transformation of aph(3')-Va-containing A. baylyi AR7 by nptII$^+$-containing donor DNA

A. baylyi AR7 was naturally transformed (liquid transformation) at 30°C with 1 µg of the constructed donor DNA and selection was done in LB with kanamycin (5 and 10 µg/mL). Recipients were

enumerated by plating dilution 10^{-5} in LB. A negative control, where water replaced the DNA in the transformation assay, was also done and plated in the two different concentrations of antibiotic. Three transformation assays were done on different days, and each assay was performed in triplicate.

The transformation frequency was calculated and is given as the number of transformants divided by the number of viable recipient cells:

Transformation frequency = number of transformants / number of recipient cells

Characterization of transformants

Changes in the phenotype and genotype of the transformants were analysed.

a) Antimicrobial susceptibility testing

Susceptibility of transformants to neomycin (10 µg) and kanamycin (30 µg) was determined by the disc diffusion method.

b) PCR-detection of acquired DNA

Transformants with changed antimicrobial susceptibility profile were screened by PCR with Phusion Hot Start II (same reaction mix as described for Phusion DNA polymerase) for DNA acquisition. The combinations of primers used were the following:

- A11 + D22 (covering the total region between fragment A and D);

- B21 + C22 (fragment BC, originally present in the recipient cell AR7);

- C21 + C22 (fragment C, containing the *aph(3')-Va* gene).

- A11 + nptII-R-int (positive if *nptII* has been acquired and inserted downstream the homologous region – fragment A – present in the recipient and in the donor DNA);

- nptII-F-int + D22 (positive if *nptII* has been acquired and inserted upstream the homologous region – fragment D – present in the recipient and in the donor DNA).

nptII-F-int = 5' GAAGGGACTGGCTGCTATTG 3'

nptII-R-int = 5' AATATCACGGGTAGCCAACG 3'

An additional reaction was done with primers A11 + B22 in transformant 10.

PCR program:

98°C/30 sec

Cycles: 30:

98°C/10 sec

58°C/10 sec

72°C/2 min

Final extention:

72°C/5 min

c) Nucleotide sequencing of the acquired DNA

Selected amplicons obtained in the previous PCRs were purified with the Genejet PCR purification kit and sequenced by Eurofins, with the same primers as used for the production of the PCR amplicons. Additional primers were designed to complete the sequence.

A21 = 5’ AAAGCCAAGCAACTCAAAGC 3’

A22 = 5’ GAAACAGAACGCCATCAGTG 3’

C23 = 5’ CTGCCCGTGGAGGACTG 3’

C13 = 5’ GTCGTCGAGGTCCACCAAG 3’

D3 = 5’ CCATTGACAGAAGATTCACC 3’

3.4.3 Results

Construction of nptII$^+$-containing donor DNA

Sequencing of the donor DNA revealed the expected nucleotide sequence, containing 735 bp of homologous region with the *A. baylyi* BD413 genome followed by the *nptII$^+$* gene (795 bp) and 442 bp corresponding to the pSDKH1/Tn5 backbone.

Construction of a new Acinetobacter baylyi strain containing the aph(3')-Va gene (recipient strain)

Transformation of *A. baylyi* BD413 with amplicon AD, containing the promoter pT5 and the *aph(3')-Va* gene flanked by two *A. baylyi* homologous regions, produced several transformants in LB with neomycin 20 µg/ml, while there was no CFUs growing on the negative control assay. Figure 24 shows the unmodified *lifO – lipA* insert region for the *aph(3')-Va* expression cassette.

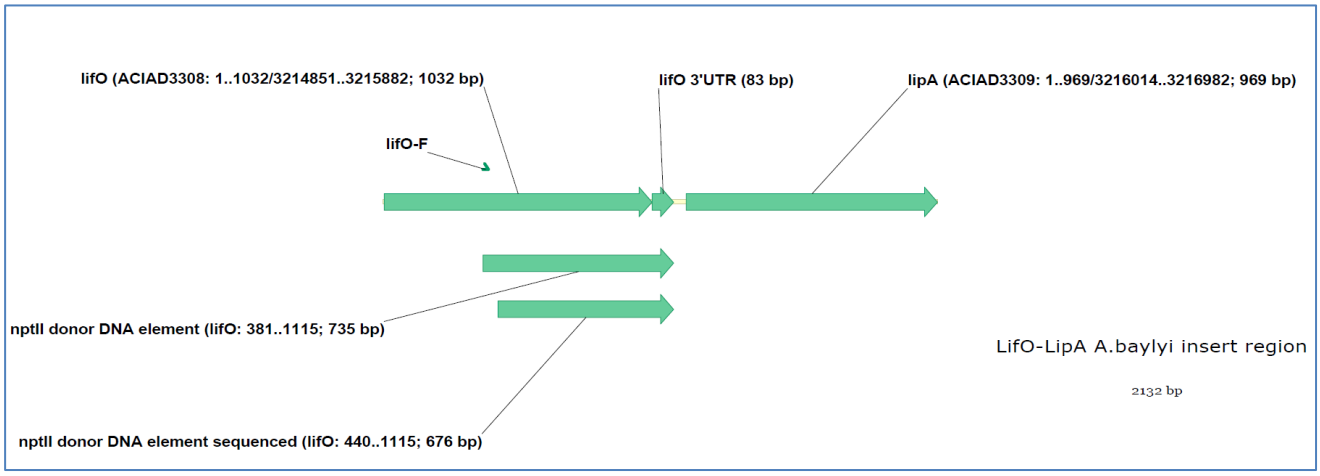


Figure 24: *Aph(3')-Va* expression cassette insertion region of *A. baylyi* BD413.

Four transformants were screened by PCR for *aph(3')-Va* gene presence. Two transformants were positive (transformant 2-3 and 2-7), while two were negative for the gene. The two positive isolates were additionally screened for the presence of the A, B and D fragments; they were positive for all the fragments.

As amplification of the total fragment AD was not successful with primers A11 + D22, three PCRs, covering the full length of the expected acquired DNA, were performed. Transformant 2-7 was positive for the three fragments AB, BC and CD, and the amplicons were furthered sequenced. Sequencing results revealed the presence of the four fragments, in the expected order and with the expected nucleotide sequence, with the exception of two mismatches in the *lifO* fragment.

The kanamycin MIC for this transformant was 1 µg/ml, as determined by Etest. This transformant was named **AR7** and was used as recipient in the following transformation experiments with the *nptII$^+$*-donor DNA.

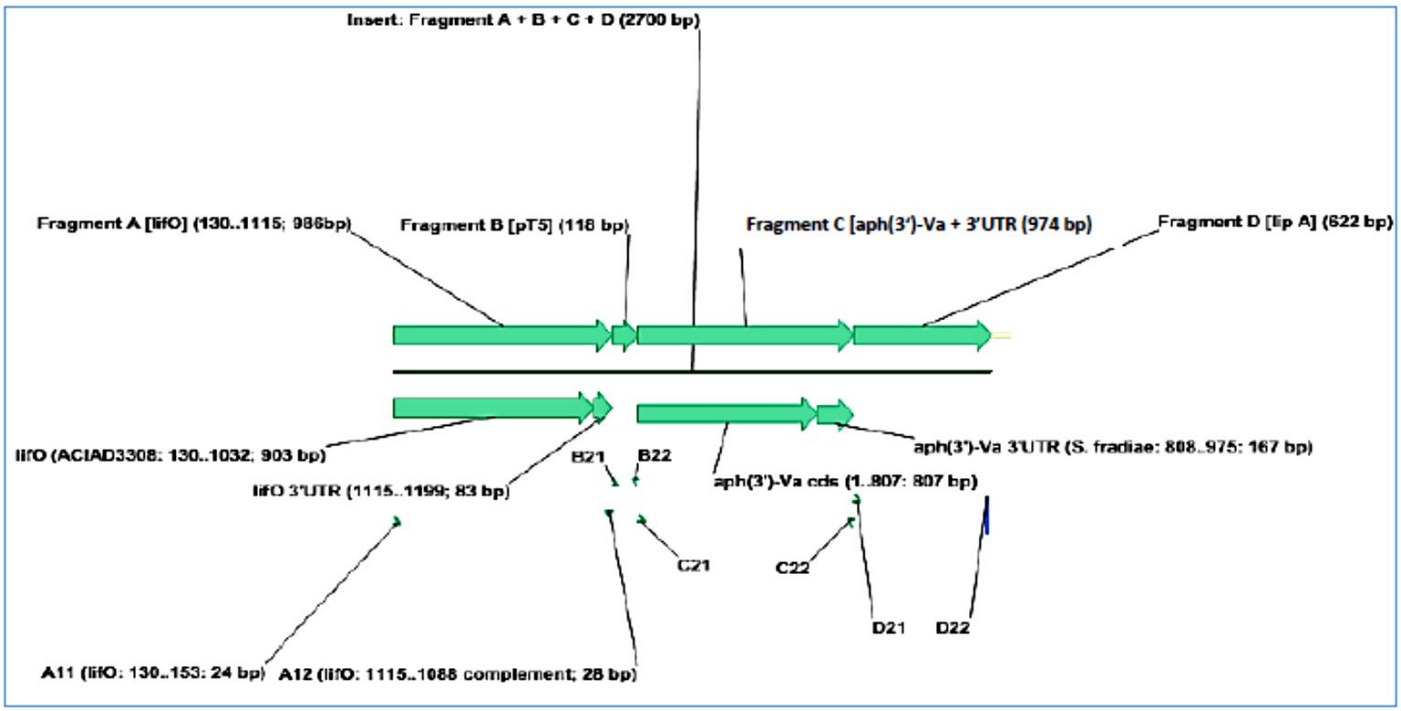


Figure 25: Schematic representation of the *aph(3')-Va* expression cassette in *A. baylyi* AR7.

The insert region of the recipient strain *A. baylyi* **AR7** (including fragment amplification and sequencing primers) is depicted.

Natural transformation of A. baylyi AR7

Natural transformation of the recipient *A. baylyi* AR7 (ACIAD3308::pT5-*aph(3')-Va*) with the *nptII*$^{+}$-containing donor DNA produced transformants in LB supplemented with kanamycin 5 µg/mL, but not with kanamycin 10 µg/mL. No CFUs were detected in the negative control.

The transformation frequency was 1.5 x 10^{-7} (Table 13).

Table 13. Natural transformation of *A. baylyi* AR7 with *nptII*$^{+}$ donor DNA.

Selection	**Mean no. of transformants (CFU)**	**Mean no. of recipients (CFU)**	**Transformants per recipients**
kanamycin 5 µg/mL	1.9 x 10^{0}	1.3 x 10^{7}	1.5 x 10^{-7}

Characterization of transformants

a) Antimicrobial susceptibility of the transformants:

The antimicrobial susceptibility profile of the recipient *A. baylyi* AR7 and transformants is shown in table 14:

Table 14. Antimicrobial susceptibility inhibition zones (mm) of *A. baylyi* AR7 transformants.

Isolate	Neomycin (mm)	Kanamycin (mm)
AR7 (receptor strain)	0	25
Transformant 1	14	9
Transformant 2	14	10
Transformant 3	14	10
Transformant 4	13	9
Transformant 5	14	10
Transformant 6	14	14
Transformant 7	14	10
Transformant 8	14	10
Transformant 9	15	11
Transformant 10	0	10
Transformant 11	14	9
Transformant 12	15	11
Transformant 13	16	12
Transformant 14	15	10
Transformant 15	15	10
Transformant 16	14	10
Transformant 17	14	11

The majority of the transformants showed a reduced resistance to neomycin, while resistance to kanamycin significantly increased; transformant 10 is an exception, maintaining the same resistance pattern to neomycin as AR7 but acquiring resistance to kanamycin.

b) Sequence analysis of the putative mosaic target region (*aph(3')-Va* expression cassette):

The amplicons obtained to check the acquisition of DNA by the transformants are represented in Table 15.

Table 15. PCR-detection of fragments acquired by *A. baylyi* AR7 transformants.

Isolate	A11 + D22[1)]	B21 + C22	C21 + C22	A11 + nptII-R-int	nptII-F-int + D22
AR7	-	+	+	-	-
Transformant 1	-	-	-	-	-
Transformant 2	-	-	-	-	-
Transformant 3	-	-	-	+	-
Transformant 4	+	-	-	-	-
Transformant 5	+	-	-	+	+
Transformant 6	+	-	-	+	+
Transformant 7	+	-	-	+	+
Transformant 8	+	-	-	+	+
Transformant 9	-	n.d.	-	-	+
Transformant 10	-	+	+	+	-
Transformant 11	-	n.d.	-	-	+
Transformant 12	-	n.d.	-	-	+
Transformant 13	-	n.d.	-	+	-
Transformant 14	+	n.d.	-	+	+
Transformant 15	-	n.d.	-	+	-
Transformant 16	-	n.d.	-	+	+
Transformant 17	-	n.d.	-	-	-

[1)] amplification primers; + amplification; - no amplification; n.d. not determined

Additionally, transformant 10 produced an amplicon with primers A11 + B22.

c) Sequence analysis of the acquired DNA

The nucleotide sequence obtained for transformants Tr4, Tr5, Tr7, Tr8 and Tr14 were similar (see Figures 26 – 30). The *npt*II gene has replaced the pT5 promoter and the *aph(3')-Va* gene, probably by recombination at the 3'end of both genes, where microhomologies are present. As the pT5 promoter was lost, expression of the acquired gene is probably due to expression from the promoter of the preceding gene, *lif*O. The 3'UTR of the *aph(3')-Va* gene is maintained in all of these transformants. Figures 26 – 30 display the recombination region and the acquired DNA.

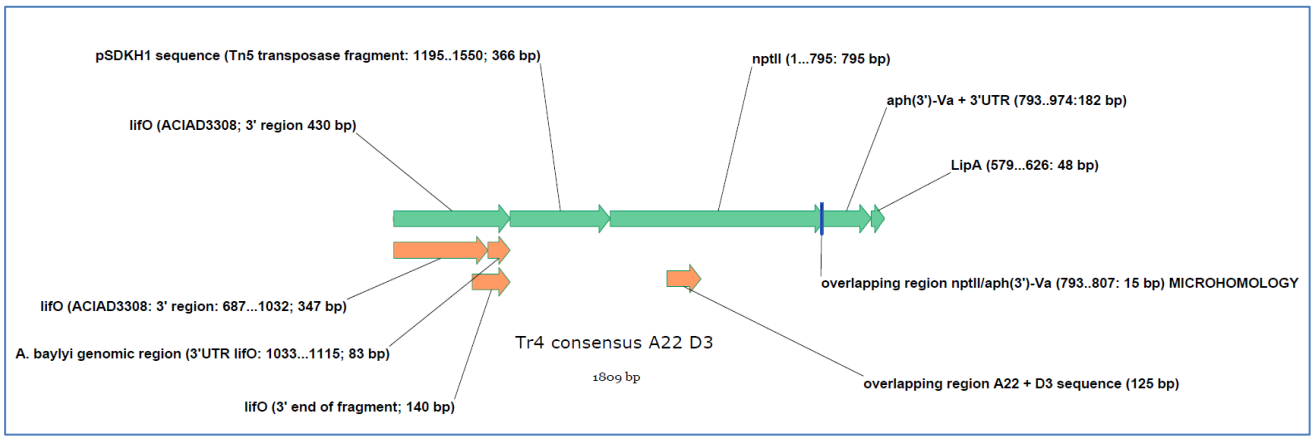


Figure 26: Schematic representation of the acquired DNA in transformant Tr4.

Recombination has probably occurred between the lifO gene of both the donor and the recipient and the 3'end of both the nptII donor gene and the *aph(3')-Va* recipient gene. The *aph(3')-Va* 3'UTR and the lipA are the sequences present in the recipient. The acquisition of the *npt*II gene replaces the promoter pT5 and the *aph(3')-Va* present in the recipient.

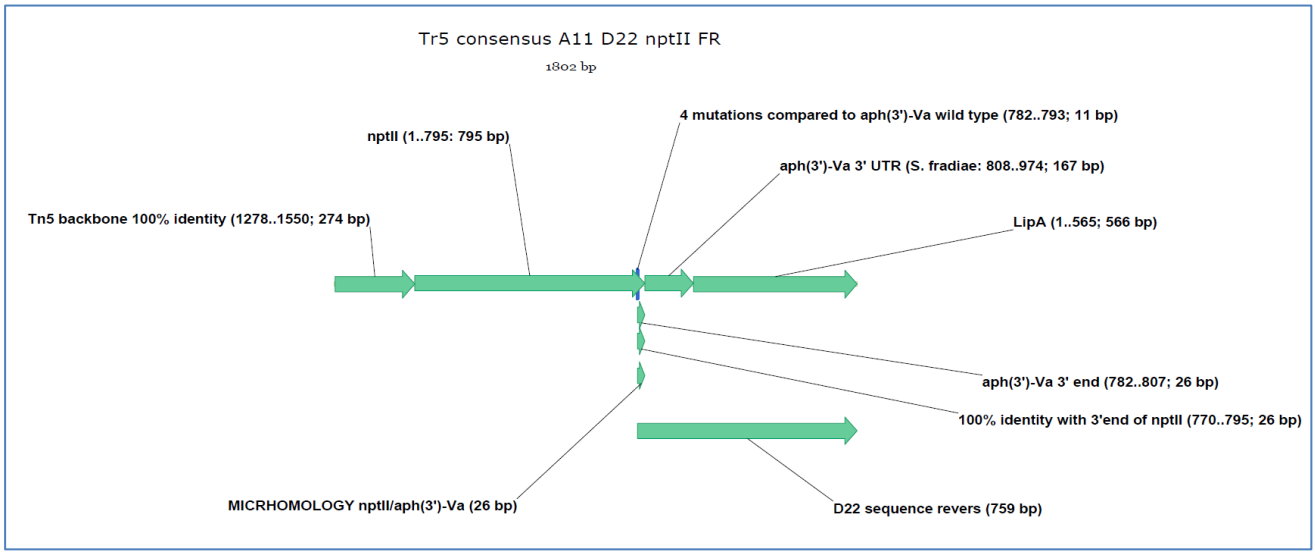


Figure 27: Schematic representation of the acquired DNA in transformant Tr5.

For details see legend of Figure 26.

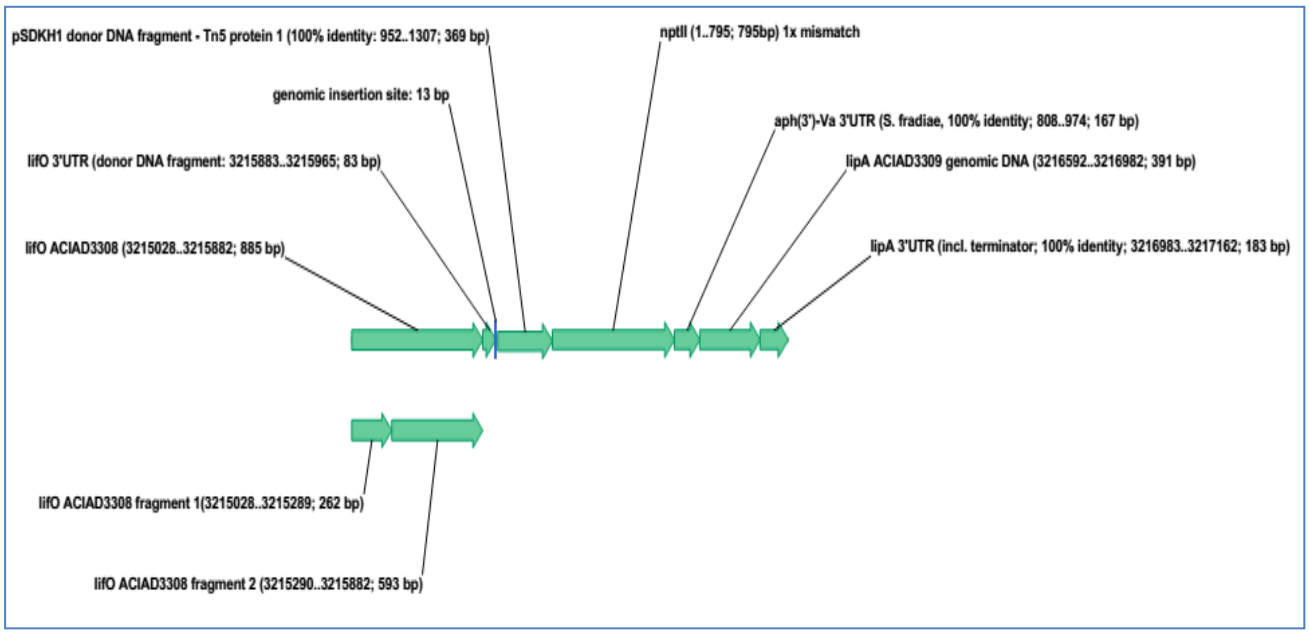


Figure 28: Schematic representation of the acquired DNA in transformant Tr7.

For details see legend of Figure 26.

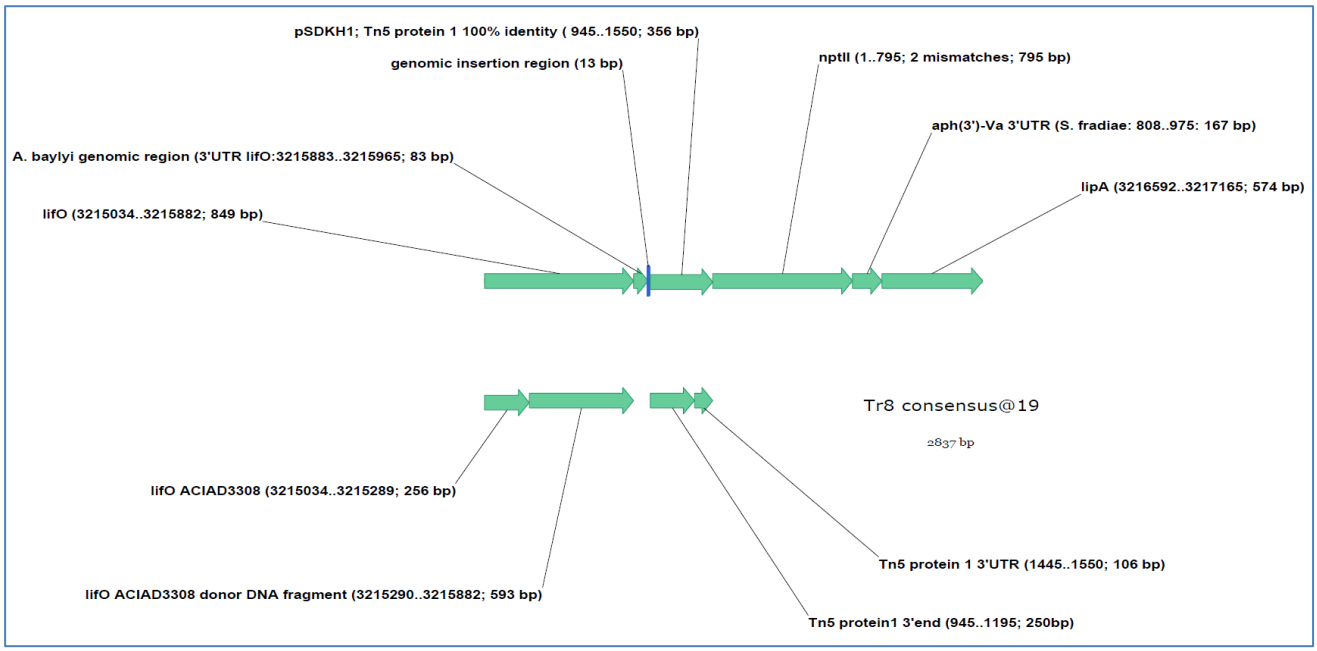


Figure 29: Schematic representation of the acquired DNA in transformant Tr8.

For details see legend of Figure 26.

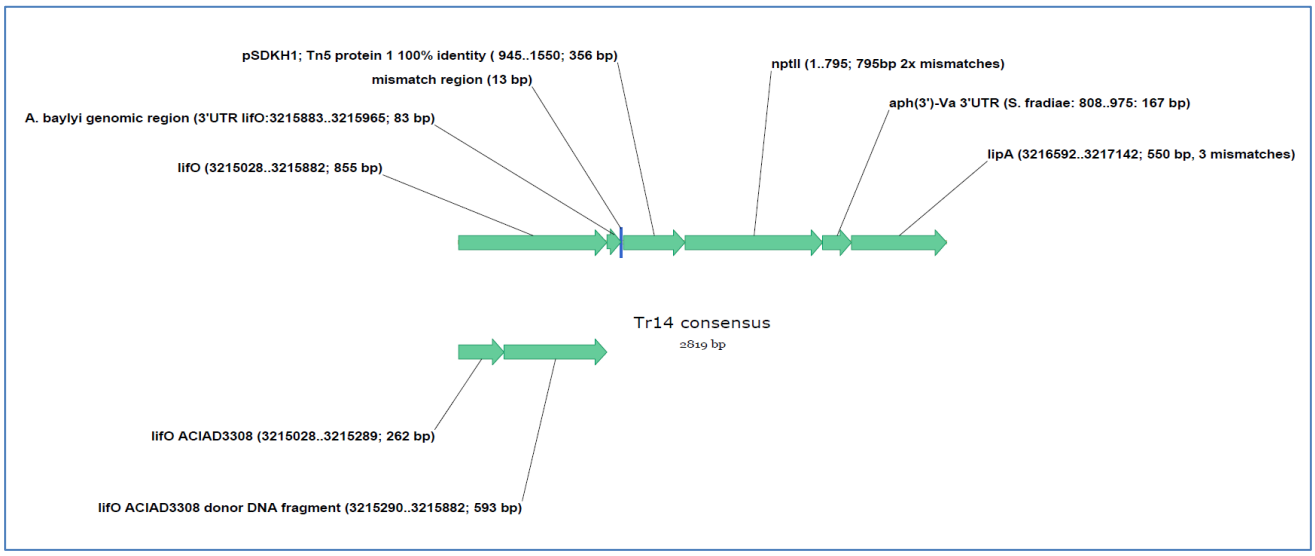


Figure 30: Schematic representation of the acquired DNA in transformant Tr14.

For details see legend of Figure 26.

Sequencing of the transformant Tr10 revealed that this transformant maintained the same inserted region as the recipient, with the pT5 promoter and the *aph(3')-Va* gene, without any mismatches in the nucleotides (Figure 31) . The changes of the antimicrobial susceptibility to aminoglycosides in this transformant are probably due to recombination in other region of the recipient genome.

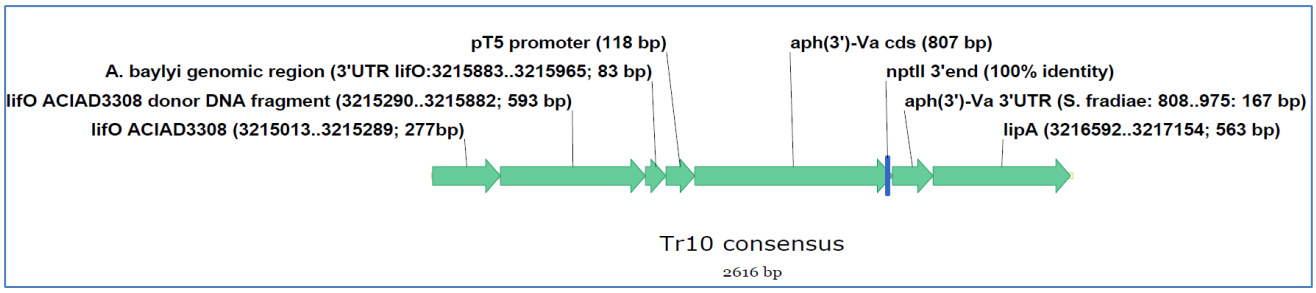


Figure 31: Schematic representation of the acquired DNA in transformant Tr10.

The result confirms correct overlapping PCR during the construction of the *aph(3')-Va* expression cassette.

3.4.4 Discussion

Acinetobacter baylyi is a common soil and water bacterium[369] and its ability to undergo natural transformation has been demonstrated both *in vitro*[370] and *in situ*[371]. Although not demonstrated yet, probably due to experimental challenges, it is expected that natural transformation occurs in the field. It has also been shown that DNA can persist in the environment over long periods of time[360, 372]. This offers the possibility for the occurrence of natural transformation of *A. baylyi* with the DNA present in the surroundings.

Accordingly to de Vries and Wackernagel[373], in the presence of a single homologous region that can serve as anchor, short stretches of sequence identity (3 – 8 bp) between donor and recipient DNA are enough for homology-facilitated illegitimate recombination (HFIR) events in *Acinetobacter baylyi* BD413. These authors have shown that HFIR occurred at a frequency of 1×10^{-8}. In our experiments we have observed a transformation frequency 10-fold higher. Our donor DNA carried a 735 bp homologous region. It is estimated that during the uptake of DNA by *A. baylyi* 500 bp are degraded due to exonuclease activity.[374] Thus, a shorter homologous region can still be enough for homologous recombination to occur. In fact, 183 bp are sufficient as a homology anchor in *A. baylyi.*[373] These numbers have to be taken into consideration when evaluating the possibility of homology between bacteria and plant DNA. Although it is unlikely that chromosomal plant DNA provides extended contiguous regions of complete sequence homology with bacterial DNA transgenic inserts often contain genes of bacterial origin or remnants of vector backbone sequences.[17, 30, 65] In both cases at least parts of these sequences may serve as anchor for interaction with bacterial DNA mediating the integration of more dissimilar sequences via HFIR. In the present case Tn5 derived vector backbone sequences may interact with endogenously present transposon sequences in competent receptor bacteria facilitating the transfer of an intact *nptII* gene. The presence of approx. 10 - 12 nucleotides identical to the 3' terminal end of *npt*II appeared to be sufficient for a successful integration of this resistance gene into the bacterial genome.

Transformation of the *aph(3')-Va*-containing recipient with *nptII*-carrying donor DNA produced changes in the antimicrobial susceptibility of the analysed transformants. Accordingly to CLSI guidelines[375], and for all microorganisms where this antibiotic is included, an inhibition halo of ≤ 13 mm, between 14-17 mm, and ≥ 18 mm represents resistance, intermediate resistance and susceptibility to kanamycin (30 µg disk), respectively. Taking these in consideration, the recipient *A. baylyi* AR7 can be considered susceptible to kanamycin, while all transformants are resistant, with the exception of transformant Tr6, which shows an intermediate profile. In the NCCLS guidelines from 2004[376], isolates were considered susceptible to neomycin (30 µg disk) when showing a diameter zone of ≥ 17 mm, intermediate between 13 and 16 mm, and resistant with ≤ 12 mm. According to this, the recipient *A. baylyi* AR7 would be considered resistant to neomycin, whereas all the transformants showed an intermediate resistance profile. Altogether these results show that the uptake of the *nptII* donor molecule results in kanamycin resistance acquisition without reverting to neomycin susceptibility.

Sequencing analysis of the acquired DNA by transformants revealed that uptake of *nptII* by natural transformation is not followed by the formation of mosaic genes with an existing *aph(3')-Va* gene, but there is a replacement of the gene. The fact that these two genes contain a nucleotide similar 3'end facilitates the acquisition of the entire *nptII* gene, and the possibility of the isolate to acquire resistance, or at least reduced susceptibility, to more antibiotics than initially.

Despite the absence of mosaic gene formation, we observed that the *nptII* gene can in fact be acquired by recipient cells containing genes sharing microhomologies in the presence of a homologous region. In theory, and although this occurs at low frequency, it can be of importance as it results in the initial acquisition of a new resistance gene, which can potentially be further spread among the soil bacterial population.[377] However, a homologous region seems to be essential, and to date there is no evidence for the stable acquisition of plant DNA in the absence of a supplied homologous DNA anchor.[378] Nonetheless, *in vitro* studies have shown that double illegitimate recombination in *A. baylyi* occurs at a frequency of 7×10^{-13}.[379] This appears to be an extremely low number indicating a very low likelihood and a negligible risk for potential adverse effects mediated by this process. However, this event theoretically may easily take place several hundred times in a single crop field of standard size.[16]

We have shown an anchor sequence mediated integration of the antibiotic resistance marker gene *nptII* into a pre-existing aminoglycoside phosphotransferase gene leading to a characteristic alteration of the antibiotic resistance profile of the recipient strain. Formation of *nptII – aph(3')-Va* mosaic patterns could not be demonstrated in our model system.

3.4.5 Conclusions

According to the results obtained from our *Acinetobacter baylyi* transformation system we cannot substantiate the hypothesis of an involvement of *aph(3')-IIa / nptII* in the formation of mosaic patterns with a related aminoglycoside phosphotransferase gene. Although the limit of detection for the detection of induced gene mosaicism in our system was high (approx. 10^{-8}) the risk for mosaic gene formation is negligible and certainly superimposed by the more frequent transfer of the intact gene by homology-directed illegitimate recombination.

Our results substantiate that anchor sequence mediated gene transfer by homology directed-illegitimate recombination is of concern considering the fact that most transgenic plant genes so far are of microbial origin and contain vector backbone sequences which could function as homologous recombination partners with chromosomally present sequences in competent bacteria.

4 Modelling Mosaic Gene Formation in Natural Environments

4.1 Summary

Two models were developed to investigate the probability and the impact of a successful mosaic gene formation event in the intestinal tract of pigs fed with transgenic plants. The simulations were run for different bacterial generation times, applying varying values for recombination rates and selection coefficients, taking into account the bacterial cell number, growth rate and movement in the gut content, DNA degradation and DNA diffusion, and the number of ingested transgenic plant DNA molecules.

The first model investigates the time until the first successful mosaic gene formation event occurs in the intestinal tract of pigs, and further the probability of such an event occurring within the life span of a single pig and in the pig population of an entire country. The results of the probabilistic framework show that the expected time for a mosaic gene formation event to occur far exceeds the life span of a single pig when applying realistic recombination rates (i.e. very low recombination rates as expected to occur in natural environments). However, under strong positive selection pressure recombination rates of 10^{-12} to 10^{-11} are suffcient to generate a mosaic gene within the life span of the animal. Considering a fattening period of 120 days the results for small to moderate values of the recombination rate and the selection coefficient indicate again only a remote probability of mosaic gene formation for a single pig. Using the annual pig production of Austria as basis for the calculations the model shows that even for very small recombination rates (i.e. 10^{-18}– 10^{-16}), a mosaic gene formation event in the population under investigation becomes very likely if the selective pressure is sufficiently high.

The second model focusses on an agricultural liquid manure tank in which the manure of a number of pigs is collected over a period of time. It is assumed that a successful formation of a mosaic gene has occurred in the digestive tract of a pig and that bacteria carrying the novel genetic material are excreted and are introduced into the model environment. Once it has entered the model environment, the novel genetic material is propagated within the bacterial host population in the presence of a selective advantage. The results show that for sufficiently large values of the selection coefficient, the multiplication effect of the novel genetic material within the bacterial host population becomes rather strong, resulting in a considerable number of bacterial cells carrying the novel gene which are eventually released into the environment.

Both models show that the results are extremely dependent on the selection pressure, while in the first model a strong dependence is further observed on the recombination rate and – to a lower extent – to the number of ingested transgenic plant DNA copies per day.

In conclusion it is remarkable that the model indicates a negligible risk for mosaic gene formation for individual animals. However, considering the phenomenon on a broader scale (e.g., annual pig production per country) the formation of mosaic genes is implied to be an event of high probability.

4.2 Aims

1. The aim of the present work is to propose and test models for the estimation of the frequency of the formation of mosaic genes and their fixation in the observed bacterial population under simulated environmental conditions.

2. The duration until a successful mosaic gene formation event (i.e. providing a selective advantage for the carrier) occurs is to be characterized by considering the following questions:

How long would a single pig need to ingest feed from transgenic plants before a mosaic gene formation event is expected to occur?

What is the probability of a mosaic gene formation event occurring during the fattening period of a single feeder pig?

What is the probability of at least one mosaic gene formation event occurring during the fattening period of feeder pigs if we consider the entire feeder pig population in Austria?

3. The propagation of a mosaic gene within the bacterial population in a liquid manure tank is characterized by

the time until 50% of the bacterial population have acquired the novel gene,

the number of bacterial cells that have acquired the novel gene six months after the introduction into the environment.

4.3 Model Environment A: Modelling mosaic gene formation in the digestive tract of pigs

4.3.1 Introduction

Based on results of Townsend et al.[29], we construct a cohesive probabilistic framework, modelling the transfer of genetic material from the plant matter ingested by a fattening pig into the intestinal bacterial population using simulation models. We determine the time until a mosaic gene formation event occurs in the bacterial population under consideration, as dependent on the fitness of the genetic trait. The model takes into account fluctuations in bacterial populations as encountered in a gastrointestinal environment. Data relevant for the simulations (DNA degradation and diffusion rates, bacterial generation times, mean retention time of the gastrointestinal content of the model organism, homologous recombination rates, selection coefficients, average cell numbers of bacterial strains in the relevant environment) was retrieved from the literature (see section 1.5) and applied in the model.

The aim of the present work is to propose and test a model for the estimation of the frequency of the formation of mosaic genes under simulated environmental conditions. We characterize the duration until a mosaic gene formation event occurs by considering the following questions:

How long would a single pig need to ingest feed from transgenic plants before a mosaic gene formation event is expected to occur?

What is the probability of a mosaic gene formation event occurring during the fattening period of a single feeder pig?

What is the probability of at least one mosaic gene formation event occurring during the fattening period of feeder pigs if we consider the entire feeder pig population in Austria?

4.3.2 Materials and Methods

4.3.2.1 Model environment

We consider the formation of mosaic genes in the intestinal tract of a single feeder pig of an approximate weight of 100 – 120 kg. The model environment consists of two separate bacterial populations:

Intestinal lining: stationary bacterial population
Intestinal content: bacterial population moves with the intestinal content

Additionally, we consider transgenic plant DNA in the feed (maize) which is ingested by the pig; see Figure 32. The feed is assumed to contain transgenes (e.g. *nptII*) which may have the potential to be involved in the formation of mosaics with similar genes (e.g. aminoglycoside phosphotransferase genes) already present in receptor bacteria. If the new genes provide a selective benefit to the host, fixation may result and the novel gene is propagated in the bacterial population.

The model encompasses of the following processes (see also **Figure** 34):

1. DNA diffusion in the intestinal content
2. DNA degradation in the intestinal content
3. Recombination; formation of mosaic genes
4. Propagation of novel gene in the bacterial population
5. Movement of the intestinal content

Item 4 (propagation of novel gene in the bacterial population) is included in the description for the sake of completeness. It is, however, not used in the computations reported in Section 4.3.3.

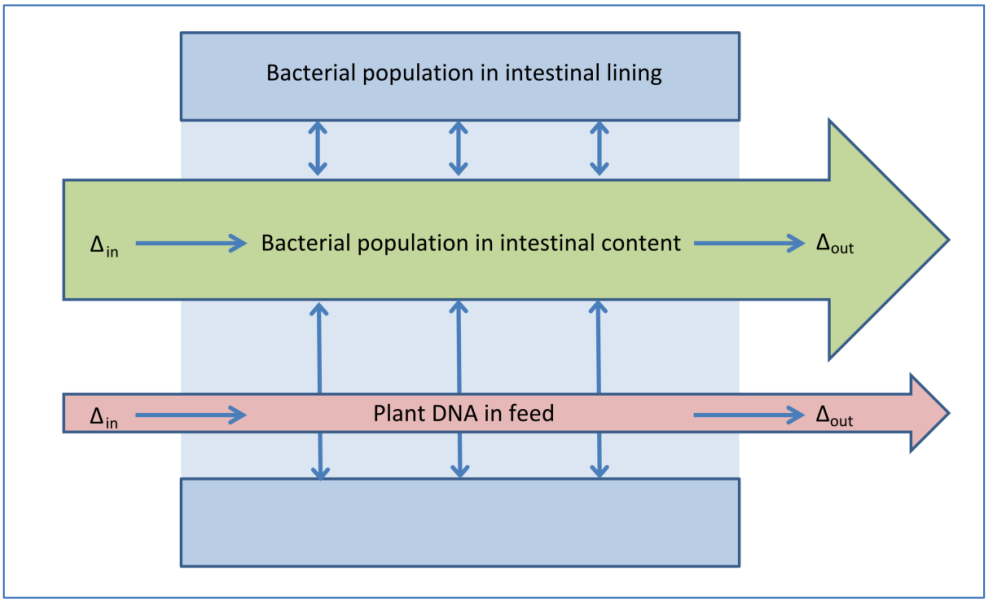


Figure 32: Schematic depiction of model environment.

4.3.2.2 Discretization

4.3.2.2.1 Spatial discretization

The intestinal tract (small intestine and large intestine) is approximated by considering a cylindrical tube of length L_{INT}. The intestinal tract is spatially discretized in the form of a one-dimensional grid consisting of N grid cells. The width of a grid cell is then given by

$$\Delta x = L_{INT} / N.$$

For each grid cell we record the total number of bacteria in the gut lining and in the intestinal content, the number of bacteria carrying the new gene and the number of transgenic plant DNA copies (see Figure 33).

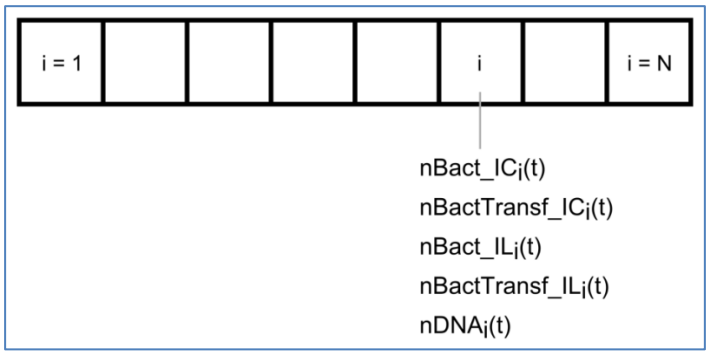


Figure 33: Spatial discretization of the model environment.

The total number of bacteria in the intestinal content (nBact_IC) and intestinal lining (nBact_IL), the number of bacteria that have acquired the novel gene in the intestinal content (nBactTransf_IC) and intestinal lining (nBactTransf_IL) and the number of transgene plant DNA copies (nDNA) for each grid cell i at time t are depicted.

4.3.2.2.2 Temporal discretization

Due to practical reasons, the temporal discretization is realized in such a manner that the intestinal content moves one grid cell ($i \rightarrow i + 1$) in each time step. The intestinal content thus passes through the entire intestinal tract in N time steps. The temporal discretization is therefore determined by the spatial discretization and the retention time Δt_{ret}:

$$\Delta t = \Delta t_{ret} / N.$$

In each time step, the processes listed in Section 4.3.2.1 are sequentially executed (see Section 4.3.2.2).

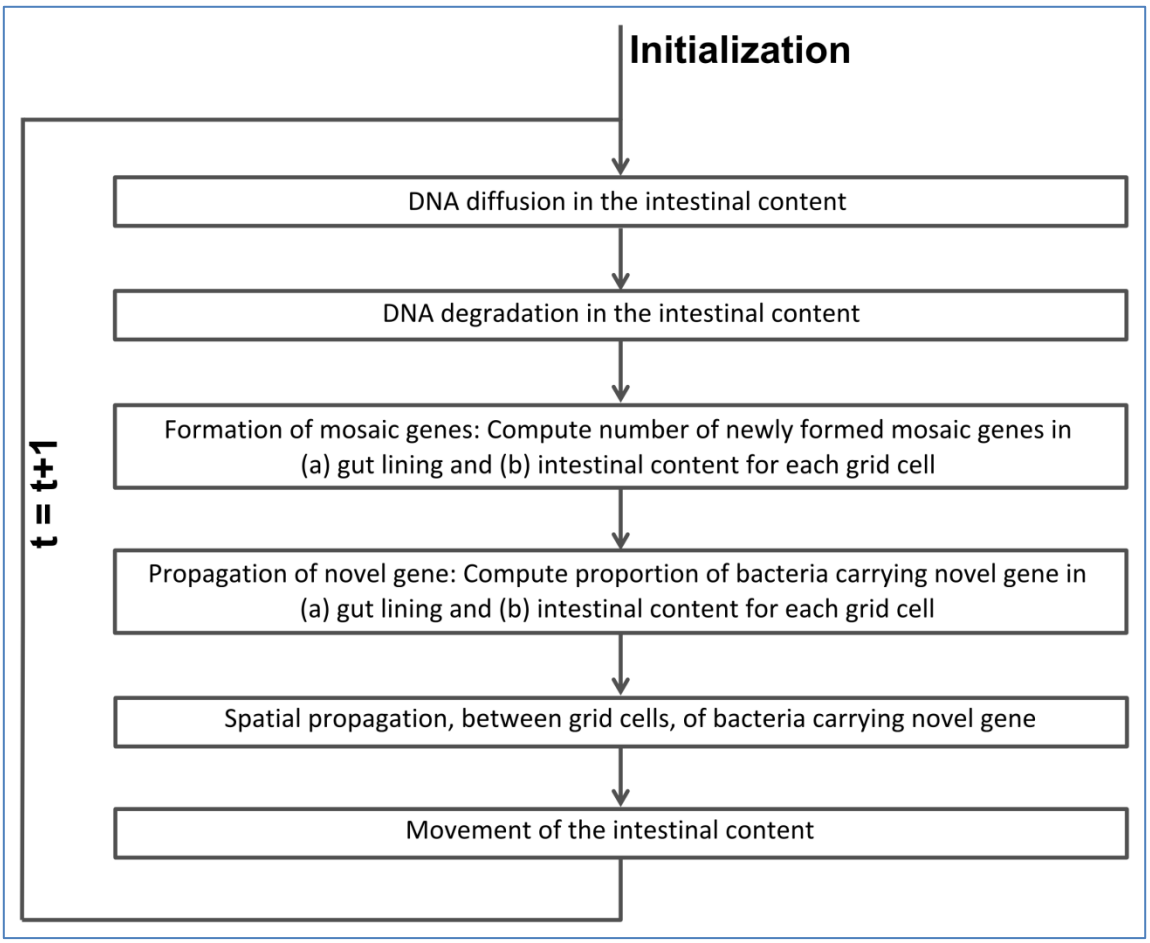


Figure 34: Sequential flow of full simulation model.

Note: The process of DNA diffusion is realized using a diffusion equation (see Section 4.3.2.3.1). This equation is numerically solved using an explicit Euler method, which is numerically stable if the following condition is met:

$$2 \cdot D \cdot \frac{\Delta t}{\Delta x^2} \leq 1,$$

where D denotes the diffusion coefficient in m^2/s. As Δt and Δx are linked, this poses a constraint (upper bound) on the number of grid cells N for which a numerically stable solution can be achieved:

$$2 \cdot D \cdot \frac{\Delta t}{\Delta x^2} = 2 \cdot D \cdot \frac{\Delta t_{ret}}{N} \cdot \frac{N^2}{L_{INT}^2} \leq 1,$$

i.e., $N \leq L_{INT}^2 / (2 \cdot D \cdot \Delta t_{ret})$.

4.3.2.3 Sub models

4.3.2.3.1 DNA diffusion

The diffusion of DNA within the intestinal content is modelled using a one-dimensional diffusion equation with Neumann boundary conditions:

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = D \cdot \frac{\partial^2 u(x,t)}{\partial x^2}, \\ \frac{\partial}{\partial x} u(0,t) = \frac{\partial}{\partial x} u(L,t) = 0, \end{cases}$$

where *u(x,t)* denotes the number of DNA copies at time *t* and position *x* and *D* denotes the diffusion coefficient in m^2 / s. We solve the diffusion equation numerically using an explicit Euler equation: Let $t_l = \Delta t \cdot l$ and $u_j^{(l)}$ denote the value at the *j*-th grid cell at time t_l. The numerical solution is then determined by

$$u_j^{(l+1)} = u_j^{(l)} + D \cdot \frac{\Delta t}{\Delta x^2} \cdot \left(u_{j+1}^{(l)} - 2 \cdot u_j^{(l)} + u_{j-1}^{(l)} \right)$$

for $j = 1,\ldots,N$, where $u_0^{(l)} = u_1^{(l)}$ and $u_{N+1}^{(l)} = u_N^{(l)}$ for all time steps $l > 0$. The numerical solution of a diffusion process is depicted in Figure 35.

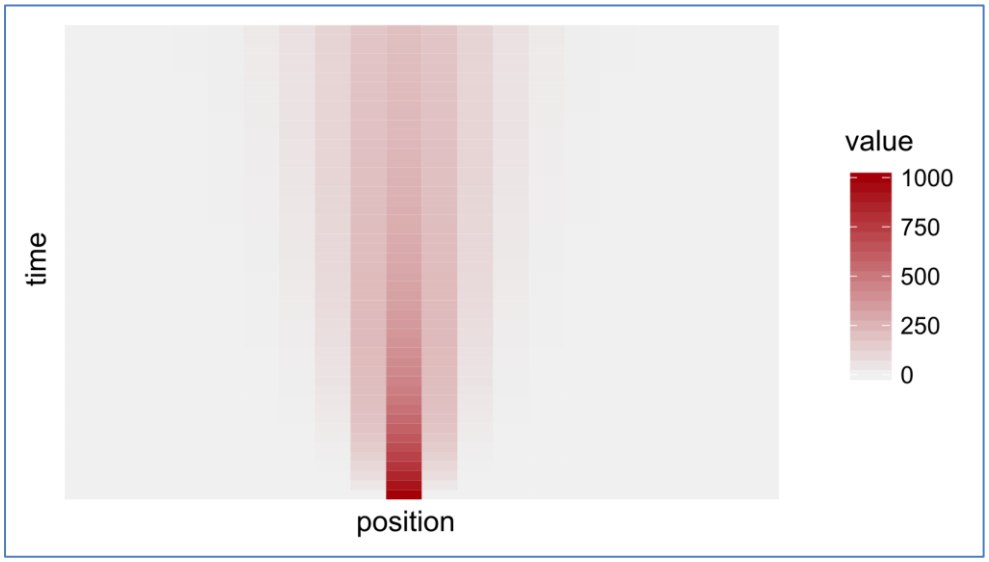


Figure 35: Example of a diffusion process.

4.3.2.3.2 DNA degradation

Following Pote et al.[380] who analyzed plant DNA decay in soil, the degradation of the DNA of the transgenic maize in the digestive tract was simulated using a double exponential decay function of the form:

$$nDNA(t) = A \cdot \exp(-k_a \cdot t) + B \cdot \exp(-k_b \cdot t),$$

where *nDNA(t)* denotes the number of DNA copies at time *t*. The application of this formula is justified as data obtained from DNA degradation studies in human intestinal contents indicate a

similar biphasic behavior. For given parameters *A, B,* k_a and k_b, we can discretize the process and compute scaling factors for a given temporal grid $t_1, t_2, \ldots, t_N$ where $t_l = \Delta t \cdot l$ for $l = 1, \ldots, N$. The scaling factors τ_j are chosen, such that

$$nDNA(t_j) = nDNA(t_{j-1}) \cdot \tau_j \text{ for all } j > 1.$$

As the intestinal content moves from grid cell *j-1* to grid cell *j* in each time step Δt (see Section 4.3.2.3.6) and the DNA present in grid cell *j* has been inside the intestinal tract for *j-1* time steps (neglecting the effects of DNA diffusion at this point), the DNA degradation process can be computed using the formula

$$nDNA_j(t_{i+1}) = nDNA_j(t_i) \cdot \tau_j \text{ for all } j > 1 \text{ and } i > 0.$$

4.3.2.3.3 Formation of mosaic genes

The process of the formation of mosaic genes in a competent host population is modelled following the methodology described in Townsend et al.[29] The formation of mosaic genes is viewed as an infrequent and stochastically independent event. Thus, the time span until such an event occurs is modeled using an exponential distribution. The expected value is determined by

- the number of exposed bacteria (n_{Bexp}) and
- the rate at which such events occur per exposed bacterium (*r*).

The time until a mosaic gene formation occurs is hence exponentially distributed with parameter

$$\lambda = r \cdot n_{Bexp},$$

(expected value = $1 / \lambda$); see also Nielsen & Townsend, 2001.[381] As not all mosaic gene formation events lead to a fixation in the host population, we characterize the probability of a new gene establishing in the population via

$$P(fixation) = \frac{1 - \exp(-2 \cdot m)}{1 - \exp(-2 \cdot N_{pop} \cdot m)},$$

where *m* denotes the relative Malthusian fitness parameter and N_{pop} is the size of the bacterial host population. We can therefore deduce that the time until a mosaic gene formation event occurs that will eventually lead to fixation is an exponentially distributed random variable with parameter

$$\lambda = r \cdot n_{Bexp} \cdot \frac{1 - \exp(-2 \cdot m)}{1 - \exp(-2 \cdot N_{pop} \cdot m)}.$$

In the modelling we make use of two very convenient properties of the exponential distribution:

1. If $X_1 \sim \exp(\lambda_1)$ and $X_2 \sim \exp(\lambda_2)$, then $Y := \min(X_1, X_2) \sim \exp(\lambda_1 + \lambda_2)$: If we assume a stationary bacterial population size and number of exposed bacteria in each grid cell, then the time until the first mosaic gene formation event occurs is exponentially distributed in each grid cell *i* with parameter λ_i. If we are interested in the time until the first grid cell experiences a successful mosaic gene formation event (minimum over all grid cells), we find that this is also an exponentially distributed random variable.
2. If the time between events is exponentially distributed with parameter λ, then the number of events in an interval of length Δt is Poisson distributed with parameter $\Delta t \cdot \lambda$: In the simulation model, the

number of newly transformed bacteria can be computed in each time step and grid cell *i* using a Poisson distribution with expected value

$$\lambda_i \cdot \frac{\Delta t}{\Delta g} = r \cdot n_{Bexp,i} \cdot \frac{1 - \exp(-2 \cdot m)}{1 - \exp\left(-2 \cdot N_{pop,i} \cdot m\right)} \cdot \frac{\Delta t}{\Delta g}$$

where Δg denotes the bacterial generation time (in the same unit as Δt), and $N_{pop,i}$ and $n_{Bexp,i}$ denote the size of the bacterial population and the number of exposed bacteria in grid cell *i*, respectively.

4.3.2.3.4 Propagation of the novel gene in the bacterial population

Once a novel genetic trait has entered the competent host population (and fixation is attained), the propagation of the novel gene in the population can be modelled deterministically. Let p_0 be the proportion (value between 0 and 1) of bacteria with the novel gene at time $t = 0$. Following Hartl & Clark[382] and Nielsen & Townsend[26] we find that the proportion of bacteria with the novel gene at time t_g *(= p(t_g))* can be characterized as follows:

$$\frac{p(t_g)}{1 - p(t_g)} = \frac{p_0}{1 - p_0} \cdot \exp\left(m \cdot t_g\right),$$

where the time t_g is measured in bacterial generations ($t_g = t / \Delta g$) and *m* denotes the relative Malthusian fitness parameter; see, e.g., Figure 36.

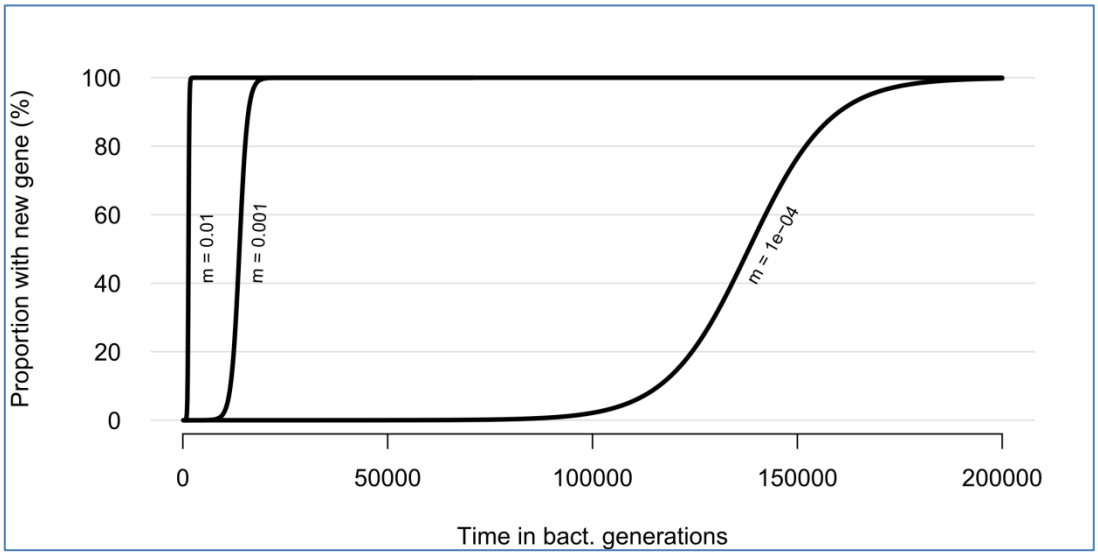


Figure 36: Propagation of a novel gene in a bacterial population under variable selection pressure.

An initial value p_0 = 1E-06 and different values of the fitness parameter *m* were used for calculation.

In order to characterize the change in the proportion of bacteria with the novel gene over one time step Δt, we compute

$$p(t + \Delta t) = \frac{\theta(t)}{1 + \theta(t)},$$

where

$$\theta(t) := \frac{p(t)}{1 - p(t)} \cdot \exp(m \cdot \Delta t / \Delta g).$$

4.3.2.3.5 Spatial propagation between grid cells of bacteria carrying the novel gene

As discussed in Section 4.3.2.3.4, once a mosaic gene formation event has occurred in the host population, the proportion of the bacteria that have acquired the novel gene increases if the gene provides a selective benefit to the host. In that sense, one can observe a propagation of the novel gene within the host population. This leads not only to an increase in pure numbers but also to a spatial propagation of the bacterial population carrying the novel genetic material. This spatial propagation is realized in the model using a diffusion equation, as discussed in Section 4.3.2.3.1.

4.3.2.3.6 Movement of intestinal content

In each time step, the intestinal content moves one grid cell to the right ($j \rightarrow j + 1$). With it, the bacteria in the intestinal content and the transgene plant DNA in the feed move from grid cell j to grid cell $j + 1$:

$$nBact\_IC_j(t_{i+1}) = nBact\_IC_{j-1}(t_i)$$

$$nBactTransf\_IC_j(t_{i+1}) = nBactTransf\_IC_{j-1}(t_i)$$

$$nDNA_j(t_{i+1}) = nDNA_{j-1}(t_i)$$

for all time steps $i > 0$ and grid cells $j = 1,\ldots,N$.

Table 16: Parameters and notation used in the simulation model.

Parameter	Description	Unit
$nBact\_IC_j(t)$	Number of bacteria in intestinal content in grid cell *j* at time *t*	–
$nBactTransf\_IC_j(t)$	Number of bacteria that have acquired the novel gene in intestinal content in grid cell *j* at time *t*	–
$nBact\_IL_j(t)$	Number of bacteria in intestinal lining in grid cell *j* at time *t*	–
$nBactTransf\_IL_j(t)$	Number of bacteria that have acquired the novel gene in intestinal lining in grid cell *j* at time *t*	–
$nDNA_j(t)$	Number of transgene plant DNA copies in intestinal content in grid cell *j* at time *t*	–
L_{INT}	Length of the intestinal tract	m
N	Number of grid cells in one dimensional space	–
Δx	Width of a grid cell	m
Δt	Length of a time step	s
Δt_{ret}	Retention time of the intestinal content	s
n_{Bexp}	Number of exposed bacteria	–
r	Rate at which mosaic gene formation events occur	Per exposed bacterium per bacterial generation
m	relative Malthusian fitness parameter	Per bacterium per bacterial generation
Δg	Bacterial growth rate/generation time	s
D	DNA diffusion coefficient	m^2 / s

4.3.2.4 Parameters

In the following section, the parameters and data on which the simulation is based are derived and discussed. A summary of the model parameters is given in Table 18: .

4.3.2.4.1 Reference animal

All parameters refer to a single feeder pig with an approximate body weight of 100 – 120 kg, which was chosen as a reference animal for the simulation.

4.3.2.4.2 Number of grid cells

For the simulation model, the intestinal tract is discretized, using N = 100 spatial grid cells.

4.3.2.4.3 Length and surface area of the gastrointestinal tract

According to Lærke & Hedemann[383] the small intestine of a fully grown pig is between 16 and 21 m long and has a surface area of the mucosa of 5.8 m^2. In the model, we consider the lower bound for the length. For the large intestine of a pig of 100 kg body weight, a length of 5 m is reported. No data on the surface area of the mucosa of the large intestines could be found. It can, however, be assumed that it is not feasible to approximate the area by considering the large intestines as a smooth cylinder due to the presence of microvilli. We approximate the surface area of the large

intestines in pigs by assuming that the large intestines in pigs are similar to those in humans and that we can scale the surface area according to the difference in lengths. The large intestines of humans have a surface area of approximately 2 m^2 and a length of approximately 1.8 m.[384] We therefore assume that the surface area of the large intestines in a 100 kg pig is 2 / 1.8 · 5 = 5.56 m^2.

4.3.2.4.4 Retention time

Le Goff et al. describe the mean retention time (Δt_{ret}) of the total gastrointestinal tract in growing pigs to be approximately 33 h (=1.188· 10^5 seconds).[385]

4.3.2.4.5 Number of DNA copies in the feed

The length of the fattening period, as well as the total feed consumption during the fattening period is reported for a daily weight gain of 750 g, 850 g and 950 g.[386] For the simulation model, we consider an average weight gain of 850 g / day, which yields the following parameters:

Duration of the fattening period (30–120 kg body weight (BW)): 106 days
Total feed consumption: 245 kg
Average feed consumption per day: 245 / 106 = 2.31 kg / day

Maize contains 6.08 · 10^8 copies of a typical transgenic insert per gram dry weight.[225] For the calculation of the copy number the following data were used: an average genome size for maize of 2.292 Mbp, an average transgenic insert size of 5000 bp, a percentage of recombinant DNA/genome of 0.00022%, recombinant DNA per g maize: 0.0033 µg, 1 bp = approx. 654 Dalton. According to Agrarmarkt Austria, the total feed can consist of up to 60% maize.[387] Therefore a feeding pig ingests 245 / 106 · 1000 · 0.6 · 6.08 · 10^8 = 8.43 · 10^{11} copies per day. If we make the simplifying assumption of a continuous food ingestions, we find that a feeding pig on average ingests 8.43 · 10^{11} · Δt_{ret} / N DNA-copies of maize per time step, where Δt_{ret} denotes the retention time in days and N denotes the number of grid cells.

4.3.2.4.6 DNA degradation in the digestive tract

The degradation of the DNA in the digestive tract was simulated using a double exponential decay function of the form:

$$nDNA(t) = A \cdot \exp(-k_a \cdot t) + B \cdot \exp(-k_b \cdot t).$$

Table 17: Fitted parameters for the double exponential decay function of DNA in the digestive tract.

Parameter	Fitted value
A	9.0838 · 10^4
k_a	4.4367 · 10^{-2}
B	9.0844 · 10^4
k_b	1.4854 · 10^{-2}

The model was fitted using data from human intestinal simulations reported in Martin- Orúe et al.[388] The fitted parameters are reported in Table 17, the fitted curve is displayed in Figure 37.

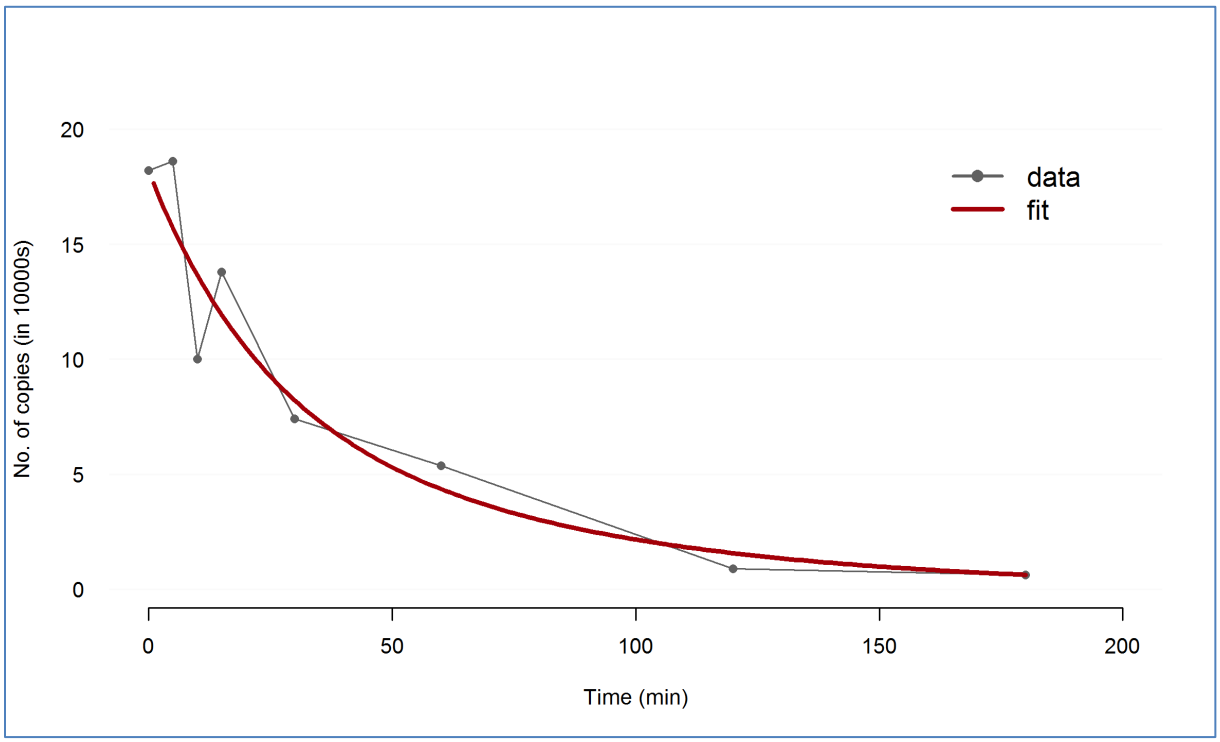


Figure 37: DNA decay in the digestive tract fitted using a double exponential function.

4.3.2.4.7 Bacterial population in the intestinal content

As no data are available on the number of bacteria in the intestinal content (gut lumen) of feeder pigs, we considered data from the human digestive system. According to Prakash et al., the small intestines of humans contain an average of $10^4 – 10^7$ bacterial cells per g of gastro-intestinal tract (GIT) content and the large intestines contain an average of $10^{10} – 10^{12}$ bacterial cells per gram of GIT content.[389]

If we assume an average daily intake of 245 / 106 = 2.31 kg feed (see Section 4.3.2.4.5) and a retention time of 33 h (see Section 4.3.2.4.4), the average content of feed in the intestinal tract is given by 2.31 / 24 · 33 = 3.18 kg. At this point we assume that 50% of the intestinal content is made up of bacteria; compared with data from Stephen & Cummings.[390] According to Lærke & Hedemann, the small intestine of a small adult animal has a capacity of approximately 20 l. The large intestine of a pig of 100 kg BW has a capacity of approximately 10 l.[383] Assuming that the distribution of the intestinal content behaves according to the ratio of the capacities of the small and large intestines (20 l : 10 l = 2 : 1), we can compute ranges for the numbers of bacteria in the gut lumen:

Min: $7.22 \cdot 1000 \cdot 0.5 \cdot (2 / 3 \cdot 10^4 + 1 / 3 \cdot 10^{10}) = 5.3 \cdot 10^{12}$

(Geometric) mean: $7.22 \cdot 1000 \cdot 0.5 \cdot (2 / 3 \cdot 10^{5.5} + 1 / 3 \cdot 10^{11}) = 5.3 \cdot 10^{13}$

Max: $7.22 \cdot 1000 \cdot 0.5 \cdot (2 / 3 \cdot 10^7 + 1 / 3 \cdot 10^{12}) = 5.3 \cdot 10^{14}$

4.3.2.4.8 Bacterial population in the gut lining

In order to estimate the number of bacterial cells in the biofilm of the intestinal mucosa, we make the following qualified assumptions:

1. The width of the biofilm is one single cell.
2. The diameter of a single bacterial cell is 1 µm.
3. A bacterial cell is circular.

The area occupied by one cell is then given by $0.5^2 \cdot \pi = 0.785$ µm^2. If we assume an optimal packing density of $\pi/(2 \cdot \sqrt{3}) = 0.969$ by arranging the circles in a hexagonal lattice, one square meter of surface area is occupied by

$$\frac{10^{12}}{0.785} \cdot \frac{\pi}{2 \cdot \sqrt{3}} = 1.15 \cdot 10^{12}$$

bacterial cells. For a total surface area of 11.36 m^2, the number of bacterial cells in the gut lining is therefore given by $1.15 \cdot 10^{12} \cdot 11.36 = 1.31 \cdot 10^{13}$ bacterial cells.

Table 18: Summary table of the model parameters.

Parameter	Value
Number of grid cells	100
Length of the intestinal tract	21 m (small intestine: 16 m, large intestine: 5 m)
Surface area of mucosa	11.36 m^2 (small intestine: 5.8 m^2, large intestine: 5.56 m^2)
Retention time of intestinal content	33 h = $1.188 \cdot 10^5$ s
Number of DNA-copies in maize ingested per day	$8.43 \cdot 10^{11}$
DNA degradation	Double exponential decay, fitted using values from Martín-Orúe et al.[388]
Number of bacterial cells in gut lumen	$5.3 \cdot 10^{13}$ (range: $5.3 \cdot 10^{12}$ – $5.3 \cdot 10^{14}$)
Number of bacterial cells in gut lining	$1.31 \cdot 10^{13}$
DNA diffusion	Consider broad range following Moradigaravand & Engelstädter[83]
Recombination rate in gut lumen	10^{-17} – 10^{-5}
Recombination rate in gut lining	10^{-17} – 10^{-5}
Bacterial growth rate; generation time	optimal lab conditions: 20 min (1200 s); in intestine: 6 – 24 h ($2.16 \cdot 10^4$ – $8.64 \cdot 10^4$ s)
Selection coefficient	weak positive selection: 10^{-12} – 10^{-3} strong positive selection: 10^{-3} – 1

4.3.2.4.9 The relevance of *Campylobacter jejuni* in the gastrointestinal tract of pigs

Campylobacter jejuni is the dominant cause of human cases of bacterial gastroenteritis. Disease symptoms range from mild, watery to a severe, inflammatory diarrhea.[391] Pigs appear to be a natural reservoir of *Campylobacter* spp. with a prevalence of 50 - 100% and excretion levels from 10^2 to 10^7 CFU/g. Opposite to most other animals, pigs show a dominance of *C. coli* compared to *C. jejuni*.[392] However, *C. jejuni* may constitute up to 87% of the *Campylobacter* detected on hog farms.[393] In outdoor pigs *C. jejuni* is more prevalent possibly due to closer contact with the environment and wildlife.[394]

An analysis of pig herds on farms with cattle and poultry production revealed a prevalence of *C. jejuni* in the fecal content of pigs of 2.3%. *C. coli* could be detected in 90.1% of the pigs.[395] Individual pigs positive for *C. jejuni* showed also high levels of *C. jejuni* excretion in feces. A mixed infection with *C. coli* was common in pigs infected with *C. jejuni*.[395]

Intraspecies recombination plays an important role in generating genetic diversity among *C. jejuni* strains.[396] *C. jejuni* and *C. coli* are reported to be naturally transformable showing transformation frequencies of approximately 10^{-4} and 10^{-3}, respectively.[8] According to these data they are primary targets for DNA uptake in natural environments. However, the frequencies were obtained in plate transformation assays in Mueller-Hinton medium without any further manipulations to provide a proof of principle for natural transformability. Due to the complex regulation of competence induction[3] it is not straight forward to infer natural transformability of *C. jejuni* in the gut of pigs.[397] Although, natural transformation of *C. jejuni* in the gastrointestinal tract of poultry[398] has been reported transformation frequencies of *Campylobacter spp.* from the porcine gastrointestinal environment are not yet available.

Taking all these data into account it was decided to avoid focusing on a special bacterial species and to develop a model for mosaic gene formation in the porcine gastrointestinal tract which allows generally valid conclusions.

4.3.3 Results

In the simulation study, we consider a stationary exposure of the host bacteria in the intestinal content and the intestinal lining to transgenic plant DNA in the intestinal content. Assuming a constant (=non-varying) intake of plant DNA, we simulate the DNA diffusion and degradation process over a time span of $10 \cdot \Delta t_{ret}$ until the number of DNA copies becomes stationary in each grid cell. Using the property of the exponential distribution described in Section 4.3.2.3.2, we can compute the time until the first grid cell experiences a successful mosaic gene formation event.

Table 19 and Table 20 show the statistical expectation of the time (in days) until the first successful mosaic gene formation event occurs in the intestinal tract (not distinguished between lining and content) for a single pig for bacterial generation times of 6 h and 12 h, respectively, and varying values for the recombination rate and the selection coefficient. For the computations, a DNA diffusion rate of 1E-8 was used. The remaining parameters were set according to Table 18: The results are to be interpreted as the expected time that a single pig would need to be fed in order to experience a successful mosaic gene formation event in the intestinal tract. As can be seen, the values are very large for small to moderate values of the recombination rate and the selection coefficient, and they by far exceed the life span of a feeder pig.

Table 19: Duration until the first mosaic gene formation occurs in the intestinal tract of a single pig (Δg=6h).

Different values for the recombination rate and the selection coefficient were used for calculation. The generation time was set to **6 h**, the DNA diffusion rate to 1E-8. The expected duration is indicated in days.

Recombination rate	**Selection coefficient**						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	1.8E+18	3.8E+16	3.8E+14	3.8E+12	3.8E+10	3.9E+08	8.9E+06
1E-17	1.8E+17	3.8E+15	3.8E+13	3.8E+11	3.8E+09	3.9E+07	8.9E+05
1E-16	1.8E+16	3.8E+14	3.8E+12	3.8E+10	3.8E+08	3.9E+06	8.9E+04
1E-15	1.8E+15	3.8E+13	3.8E+11	3.8E+09	3.8E+07	3.9E+05	8.9E+03
1E-14	1.8E+14	3.8E+12	3.8E+10	3.8E+08	3.8E+06	3.9E+04	885
1E-13	1.8E+13	3.8E+11	3.8E+09	3.8E+07	3.8E+05	3866	89
1E-12	1.8E+12	3.8E+10	3.8E+08	3.8E+06	3.8E+04	387	9
1E-11	1.8E+11	3.8E+09	3.8E+07	3.8E+05	3828	39	< 1
1E-10	1.8E+10	3.8E+08	3.8E+06	3.8E+04	383	4	< 1
1E-09	1.8E+09	3.8E+07	3.8E+05	3827	38	< 1	< 1
1E-08	1.8E+08	3.8E+06	3.8E+04	383	4	< 1	< 1
1E-07	1.8E+07	3.8E+05	3827	38	< 1	< 1	< 1
1E-06	1.8E+06	3.8E+04	383	4	< 1	< 1	< 1
1E-05	1.8E+05	3827	38	< 1	< 1	< 1	< 1

Table 20: Duration until the first mosaic gene formation occurs in the GIT of a single pig (Δg=12h).

Different values for the recombination rate and the selection coefficient were used for calculation. The generation time was set to **12 h**, the DNA diffusion rate to 1E-8. The expected duration is indicated in days.

Recombination rate	Selection coefficient						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	3.7E+18	7.7E+16	7.7E+14	7.7E+12	7.7E+10	7.7E+08	1.8E+07
1E-17	3.7E+17	7.7E+15	7.7E+13	7.7E+11	7.7E+09	7.7E+07	1.8E+06
1E-16	3.7E+16	7.7E+14	7.7E+12	7.7E+10	7.7E+08	7.7E+06	1.8E+05
1E-15	3.7E+15	7.7E+13	7.7E+11	7.7E+09	7.7E+07	7.7E+05	1.8E+04
1E-14	3.7E+14	7.7E+12	7.7E+10	7.7E+08	7.7E+06	7.7E+04	1771
1E-13	3.7E+13	7.7E+11	7.7E+09	7.7E+07	7.7E+05	7731	177
1E-12	3.7E+12	7.7E+10	7.7E+08	7.7E+06	7.7E+04	773	18
1E-11	3.7E+11	7.7E+09	7.7E+07	7.7E+05	7655	77	2
1E-10	3.7E+10	7.7E+08	7.7E+06	7.7E+04	766	8	< 1
1E-09	3.7E+09	7.7E+07	7.7E+05	7654	77	< 1	< 1
1E-08	3.7E+08	7.7E+06	7.7E+04	765	8	< 1	< 1
1E-07	3.7E+07	7.7E+05	7654	77	< 1	< 1	< 1
1E-06	3.7E+06	7.7E+04	765	8	< 1	< 1	< 1
1E-05	3.7E+05	7654	77	< 1	< 1	< 1	< 1

In Table 21 and Table 22, the probability of a single pig experiencing a successful mosaic gene formation event within an assumed fattening period of 120 days is reported for bacterial generation times of 6 h and 12 h, respectively, and varying values for the recombination rate and the selection coefficient. The computations are again based on a DNA diffusion rate of 1E-8. It can be seen that for small to moderate values of the recombination rate and the selection coefficient, the probability of mosaic gene formation are very small for a single pig.

Table 21: Probability of a mosaic gene formation event in a pig (120 day fattening period; Δg=6h).

Different values of the recombination rate and the selection coefficient were used for calculation. The generation time was set to **6 h**, the DNA diffusion rate to 1E-8.

Recombination rate	Selection coefficient						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	< 1E-16	3.1E-15	3.1E-13	3.1E-11	3.1E-09	3.1E-07	1.4E-05
1E-17	6.7E-16	3.1E-14	3.1E-12	3.1E-10	3.1E-08	3.1E-06	1.4E-04
1E-16	6.6E-15	3.1E-13	3.1E-11	3.1E-09	3.1E-07	3.1E-05	1.4E-03
1E-15	6.5E-14	3.1E-12	3.1E-10	3.1E-08	3.1E-06	3.1E-04	0.013
1E-14	6.5E-13	3.1E-11	3.1E-09	3.1E-07	3.1E-05	0.003	0.127
1E-13	6.5E-12	3.1E-10	3.1E-08	3.1E-06	3.1E-04	0.031	0.742
1E-12	6.5E-11	3.1E-09	3.1E-07	3.1E-05	0.003	0.267	> 0.9999
1E-11	6.5E-10	3.1E-08	3.1E-06	3.1E-04	0.031	0.955	> 0.9999
1E-10	6.5E-09	3.1E-07	3.1E-05	0.003	0.269	> 0.9999	> 0.9999
1E-09	6.5E-08	3.1E-06	3.1E-04	0.031	0.957	> 0.9999	> 0.9999
1E-08	6.5E-07	3.1E-05	0.003	0.269	> 0.9999	> 0.9999	> 0.9999
1E-07	6.5E-06	3.1E-04	0.031	0.957	> 0.9999	> 0.9999	> 0.9999
1E-06	6.5E-05	0.003	0.269	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-05	6.5E-04	0.031	0.957	> 0.9999	> 0.9999	> 0.9999	> 0.9999

Table 22: Probability of a mosaic gene formation event in a pig (120 day fattening period; Δg=12h).

Different values for the recombination rate and the selection coefficient were used for calculation. Generation time was set to **12 h**, the DNA diffusion rate to 1E-8.

Recombination rate	Selection coefficient						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	< 1E-16	1.6E-15	1.6E-13	1.6E-11	1.6E-09	1.6E-07	6.8E-06
1E-17	3.3E-16	1.6E-14	1.6E-12	1.6E-10	1.6E-08	1.6E-06	6.8E-05
1E-16	3.2E-15	1.6E-13	1.6E-11	1.6E-09	1.6E-07	1.6E-05	6.8E-04
1E-15	3.3E-14	1.6E-12	1.6E-10	1.6E-08	1.6E-06	1.6E-04	0.007
1E-14	3.3E-13	1.6E-11	1.6E-09	1.6E-07	1.6E-05	0.002	0.066
1E-13	3.3E-12	1.6E-10	1.6E-08	1.6E-06	1.6E-04	0.015	0.492
1E-12	3.3E-11	1.6E-09	1.6E-07	1.6E-05	0.002	0.144	> 0.9999
1E-11	3.3E-10	1.6E-08	1.6E-06	1.6E-04	0.016	0.788	> 0.9999
1E-10	3.3E-09	1.6E-07	1.6E-05	0.002	0.145	> 0.9999	> 0.9999
1E-09	3.3E-08	1.6E-06	1.6E-04	0.016	0.791	> 0.9999	> 0.9999
1E-08	3.3E-07	1.6E-05	0.002	0.145	> 0.9999	> 0.9999	> 0.9999
1E-07	3.3E-06	1.6E-04	0.016	0.791	> 0.9999	> 0.9999	> 0.9999
1E-06	3.3E-05	0.002	0.145	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-05	3.3E-04	0.016	0.791	> 0.9999	> 0.9999	> 0.9999	> 0.9999

Approximately 5.4 million pigs were slaughtered in Austria in 2013.[399] We therefore attempt to put the formation probabilities into perspective by computing the probability of a mosaic gene formation event occurring in at least one of 5 Mio. pigs during the fattening period of 120 days. The results are listed in Table 23 (generation time = 6 h) and Table 24 (generation time = 12 h) and show that even for very small recombination rates (1E-18 – 1E-16), a mosaic gene formation event in the population becomes likely if the selective pressure is sufficiently high.

Table 23: Probability of a mosaic gene formation event occurring in at least one of 5 Mio. pigs (Δg=6h).

Different values of the recombination rate and the selection coefficient were used for calculation. Generation time was set to **6 h**, DNA diffusion rate to 1E-8.

Recombination rate	Selection coefficient						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	5.6E-10	1.6E-08	1.6E-06	1.6E-04	0.016	0.788	> 0.9999
1E-17	3.3E-09	1.6E-07	1.6E-05	0.002	0.145	> 0.9999	> 0.9999
1E-16	3.3E-08	1.6E-06	1.6E-04	0.016	0.791	> 0.9999	> 0.9999
1E-15	3.3E-07	1.6E-05	0.002	0.145	> 0.9999	> 0.9999	> 0.9999
1E-14	3.3E-06	1.6E-04	0.016	0.791	> 0.9999	> 0.9999	> 0.9999
1E-13	3.3E-05	0.002	0.145	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-12	3.3E-04	0.016	0.791	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-11	0.003	0.145	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-10	0.032	0.791	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-09	0.279	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-08	0.962	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-07	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-06	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-05	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999

Table 24: Probability of a mosaic gene formation event occurring in at least one of 5 Mio. pigs (Δg=12h).

Different values of the recombination rate and the selection coefficient were used for calculation. Generation time was set to **12 h**, DNA diffusion rate to 1E-8.

Recombination rate	Selection coefficient						
	1E-12	**1E-10**	**1E-08**	**1E-06**	**1E-04**	**0.01**	**1**
1E-18	1.7E-10	7.8E-09	7.8E-07	7.8E-05	0.008	0.540	> 0.9999
1E-17	1.7E-09	7.8E-08	7.8E-06	0.001	0.075	> 0.9999	> 0.9999
1E-16	1.6E-08	7.8E-07	7.8E-05	0.008	0.543	> 0.9999	> 0.9999
1E-15	1.6E-07	7.8E-06	0.001	0.075	> 0.9999	> 0.9999	> 0.9999
1E-14	1.6E-06	7.8E-05	0.008	0.543	> 0.9999	> 0.9999	> 0.9999
1E-13	1.6E-05	0.001	0.075	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-12	1.6E-04	0.008	0.543	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-11	0.002	0.075	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-10	0.016	0.543	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-09	0.151	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-08	0.805	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-07	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-06	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
1E-05	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999

Figure 38 shows the probability of a mosaic gene formation event occurring in at least one of 5 million pigs during their respective fattening period of 120 days in the form of surface plots. The plots illustrate the dependence of the probability of a formation event on the bacterial generation time (note that the x- and y-axes in the plots are displayed on a logarithmic scale). The probability of an event increases with a decreasing bacterial generation time, although, compared to the dependence on the selection pressure, the effect seems to be weaker.

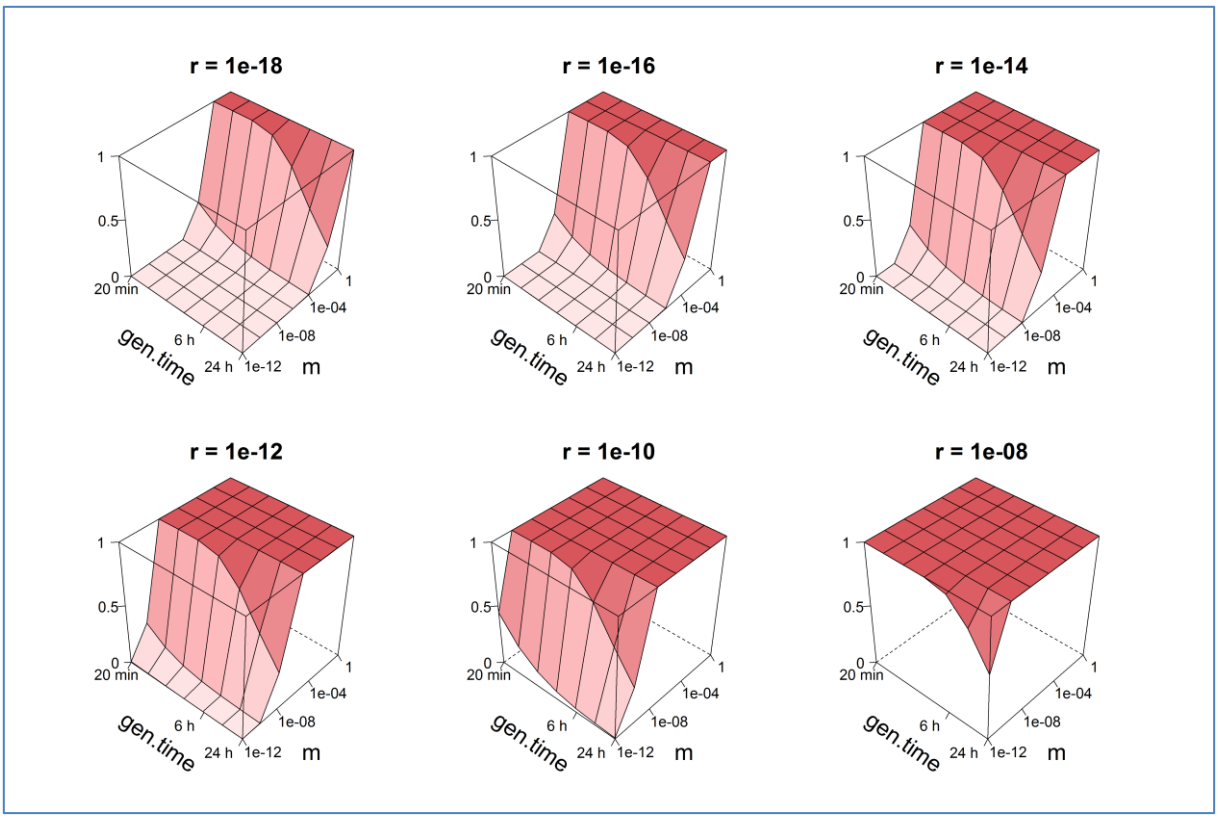


Figure 38: Surface plot of the probability of a mosaic gene formation in the annual pig production of Austria.

The probability for at least one event in 5 million pigs for different values of the recombination rate (r), the selection coefficient (m) and the bacterial generation time is calculated. The DNA diffusion rate was set to 1E-8.

4.3.3.1 System sensitivity

As can be seen, the results reported above strongly depend on the recombination rate and the value of the selection coefficient and, to some extent on the bacterial growth rate/generation time. In a sensitivity analysis, we further examined the dependence on the size of the bacterial populations and on the DNA diffusion coefficient.

In a univariate analysis, the number of bacterial cells in the intestinal lining was varied between 10^{10} and 10^{15}, the number of bacteria in the intestinal content was varied between 10^{10} and 10^{15} and the value of the DNA diffusion coefficient was varied between 10^{-12} and 10^{-7}. In the considered value ranges, the system did not seem to be significantly influenced by the parameter changes.

The system is, however, sensitive to the number of ingested transgenic plant DNA copies. Figure 39 shows the probability of a mosaic gene formation event occurring during the fattening period of a single feeder pig for varying values of the recombination rate and the number of transgenic plant DNA copies ingested per day. The computations are based on a selection coefficient of $m = 10^{-8}$, a bacterial generation time of 12 h and a DNA diffusion coefficient of 10^{-8}. The figure shows a positive association between the number of DNA copies and the probability of mosaic gene formation. This trend is also visible if the values of the selection coefficient, the generation time and the DNA diffusion are varied (not depicted in the figure). The results show that variations/uncertainties in the number of ingested DNA copies can have a noticeable effect on the outcome and should, therefore, be taken under consideration.

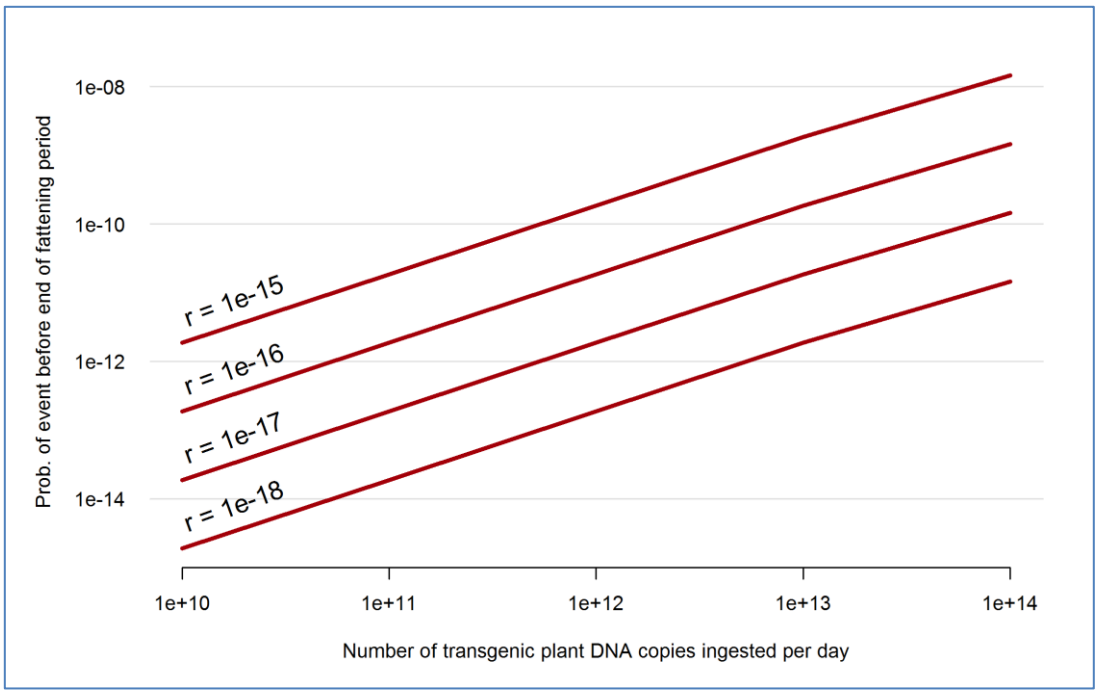


Figure 39: Probability of mosaic gene formation occurring in a pig within a 120 day fattening period.

Different values for the number of DNA copies ingested/d and the recombination rate were used for calculation. Selection coefficient = 1E-8, generation time = 12 h, DNA diffusion = 1E-8

4.3.4 Conclusions

The results of the probabilistic framework show that the expected time for a mosaic gene formation event to occur far exceeds the life span of a single pig applying realistic recombination rates (i.e. very low recombination rates as expected to occur in natural environments). However, under strong positive selection pressure, recombination rates of 10^{-12} to 10^{-11} are suffcient to generate a mosaic gene within the life span of the animal. Considering a fattening period of 120 days the results for small to moderate values of the recombination rate and the selection coefficient indicate again only a remote probability of mosaic gene formation for a single pig. Using the annual pig production of Austria as basis for the calculations our model shows that even for very small recombination rates (i.e. 10^{-18}– 10^{-16}), a mosaic gene formation event in the population under investigation becomes very likely if the selective pressure is sufficiently high.

The model shows that the results are extremely dependent on the recombination rate and the selection pressure and - to a lower extent – to the number of ingested transgenic plant DNA copies per day.

In conclusion it is remarkable that the model indicates a negligible risk for mosaic gene formation for individual animals. However, considering the phenomenon on a broader scale (e.g. annual pig production per country) there seems to be a realistic possibility of a successful mosaic gene formation.

4.4 Model Environment B: Modelling the propagation of mosaic genes in liquid manure tanks

4.4.1 Introduction

We consider a liquid manure tank on a pig farm, in which the manure of a number of pigs is collected over a period of time. We assume that a successful formation of a mosaic gene has occurred in the digestive tract of a pig and that bacteria carrying the novel genetic material are excreted and are introduced into the model environment at time t_0. We assume that one bacterial cell carrying the mosaic gene is introduced. We further assume that the tank is at 50% of its capacity at the time of introduction and assume a stationary number of bacteria throughout the modelling process, i.e., we neglect the effects of a reduction of the pathogen load due to thermophilic digestion on the one hand and the addition of bacteria due to a further filling of the tank with fresh manure on the other hand.

Once it has entered the model environment, the novel genetic material is propagated within the bacterial host population in the presence of a selective advantage. We model the propagation as described in Section 4.3.2.3.4 and characterize the time until 50% of the host population has acquired the novel gene for different parameter values (capacity of manure tank, bacterial generation time, selection coefficient). As agricultural holdings are required to ensure a storage capacity for manure for up to six months, we further compute the number of bacterial cells in the manure tank that have acquired the novel gene after a period of six months.

4.4.2 Materials and Methods

4.4.2.1 Sub models

4.4.2.1.1 Propagation of the novel gene in the bacterial population

The propagation of the novel gene in the host population is modelled as described in Section 4.3.2.3.4. The time t_{50} until 50% of the host population has acquired the novel gene can then be computed using

$$t_{50} = log\left(\frac{1-p_0}{p_0}\right) \cdot \frac{1}{m} \cdot \Delta g,$$

where Δg is the bacterial generation time. In our model, we assume that exactly one bacterial cell carrying the mosaic gene enters the manure tank. The formula above is then equivalent to

$$t_{50} = log(N-1) \cdot \frac{1}{m} \cdot \Delta g,$$

where N denotes the size of the bacterial host population, i.e., the number of bacterial cells in the manure tank.

4.4.2.2 Parameters

4.4.2.2.1 Holding capacity of a manure tank

The holding capacities of Austrian manure tanks generally vary between 10 m^3 and 10 000 m^3. The required capacity depends on the size of the holding. Most commonly, manure tanks with capacities between 2000 m^3 and 5000 m^3 are in use.

4.4.2.2.2 Number of bacterial cells in a manure tank

In a study by Cotta et al.[400], the number of bacterial cells in liquid manure storage systems for pigs was found to lie between 10^9 and 10^{10} cells per ml of liquid manure. In our simulation study we will use the geometric mean of $10^{9.5} = 3.16 \cdot 10^9$.

4.4.2.2.3 The relevance of *Acinetobacter* spp. for manure

There are two species of the genus *Acinetobacter* described to be naturally transformable: *Acinetobacter baumannii* and *Acinetobacter baylyi.*[401] *A. baumannii* is an opportunistic human pathogen causing severe problems in hospital derived infections because this species is reported to has become increasingly multidrug resistant.[402] This feature relies largely on the acquisition of resistance genes via horizontal gene transfer.[12] *Acinetobacter baylyi* is a typical soil bacterium of low pathogenicity used as model system to study DNA uptake from the environment.[403]

A. baumannii is yielding naturally induced transformation frequencies ranging from 10^{-3} to 10^{-8} in plate assays.[126] *A. baylyi* shows a transformation frequency of approximately 10^{-3} in liquid culture.[8, 404] Transformation of *A. baylyi* in soil has been demonstrated using plant DNA derived antibiotic resistance genes for restoration of resistance in *A. baylyi* carrying a mutated version of a similar resistance gene (i.e. marker rescue).[403] Piggery manure is a reservoir of antibiotic resistance genes and corresponding plasmids[405-407] and is usually contaminated with antibiotics at sub-lethal or higher concentrations. [408-411] There is no data accessible dealing with natural transformability of *Acinetobacter spp.* in manure.

Taking these observations into account it was decided to avoid focusing on a special bacterial species and to develop a model for mosaic gene dissemination in manure which allows generally valid conclusions.

4.4.3 Results

4.4.3.1 Time until 50% of the population have acquired the gene

Table 25–Table 27 report the time in days until 50% of the bacterial population in a liquid manure tank have acquired the novel gene for tank capacities of 10 m^3, 1000 m^3 and 10 000 m^3. As can be deduced from the formula stated in Section 4.4.2.1.1, the value of t_{50} depends linearly on the bacterial generation time (i.e., doubling of the generation time leads to a doubling of t_{50}) and inversely on the selection coefficient (i.e., halving of the selection coefficient leads to a doubling of t_{50}). This is reflected in the results listed below. The dependence on the size of the host population, i.e., the carrying capacity of the tank, is only logarithmic and hence rather weak compared to the other parameters described above. It can be seen that for sufficiently large values of the selection coefficient, t_{50} lies within a practically relevant range.

Table 25: Time (in days) for fixation of the trait (liquid manure tank of 10 m^3 capacity).

The time until until 50% of the bacterial population in a liquid manure tank of 10 m^3 capacity has acquired the new gene for different values of the selection coefficient and the bacterial generation time is displayed.

Selection coefficient	Bacterial generation time			
	20 min	6 h	12 h	24 h
1E-12	5.2E+11	9.3E+12	1.9E+13	3.7E+13
1E-10	5.2E+09	9.3E+10	1.9E+11	3.7E+11
1E-08	5.2E+07	9.3E+08	1.9E+09	3.7E+09
1E-06	5.2E+05	9.3E+06	1.9E+07	3.7E+07
1E-04	5180.487	9.3E+04	1.9E+05	3.7E+05
0.01	51.805	932.488	1864.975	3729.951
1	0.518	9.325	18.650	37.300

Table 26: Time (in days) for fixation of the trait (liquid manure tank of 1000 m^3 capacity).

The time until 50% of the bacterial population in a liquid manure tank of 1000 m^3 capacity has acquired the new gene for different values of the selection coefficient and the bacterial generation time is displayed.

Selection coefficient	Bacterial generation time			
	20 min	6 h	12 h	24 h
1E-12	5.8E+11	1.0E+13	2.1E+13	4.2E+13
1E-10	5.8E+09	1.0E+11	2.1E+11	4.2E+11
1E-08	5.8E+07	1.0E+09	2.1E+09	4.2E+09
1E-06	5.8E+05	1.0E+07	2.1E+07	4.2E+07
1E-04	5820.094	1.0E+05	2.1E+05	4.2E+05
0.01	58.201	1047.617	2095.234	4190.468
1	0.582	10.476	20.952	41.905

Table 27: Time (in days) for fixation of the trait (liquid manure tank of 10 000 m^3 capacity).

The time until 50% of the bacterial population in a liquid manure tank of 10 000 m^3 capacity has acquired the new gene for different values of the selection coefficient and the bacterial generation time is displayed.

Selection coefficient	Bacterial generation time			
	20 min	6 h	12 h	24 h
1E-12	6.1E+11	1.1E+13	2.2E+13	4.4E+13
1E-10	6.1E+09	1.1E+11	2.2E+11	4.4E+11
1E-08	6.1E+07	1.1E+09	2.2E+09	4.4E+09
1E-06	6.1E+05	1.1E+07	2.2E+07	4.4E+07
1E-04	6139.898	1.1E+05	2.2E+05	4.4E+05
0.01	61.399	1105.182	2210.363	4420.726
1	0.614	11.052	22.104	44.207

4.4.3.2 Number of bacterial cells with novel gene after six months

Figure 40 shows the number of bacterial cells carrying the novel gene in a manure tank six months after a single cell carrying the mosaic gene was introduced into the environment. In the figure, results are shown for different bacterial generation times as well as for different values for the capacity of the tank and the selection coefficient (*m*). Note that all three axes are depicted on a logarithmic scale. The number of bacterial cells carrying the new gene depends weakly on the capacity of the tank. A strong dependence can, however, be observed on the selection coefficient. In the presence of strong positive selection ($m \approx 10^{-4}$ or larger), a considerable multiplication effect of the novel genetic material is noticeable.

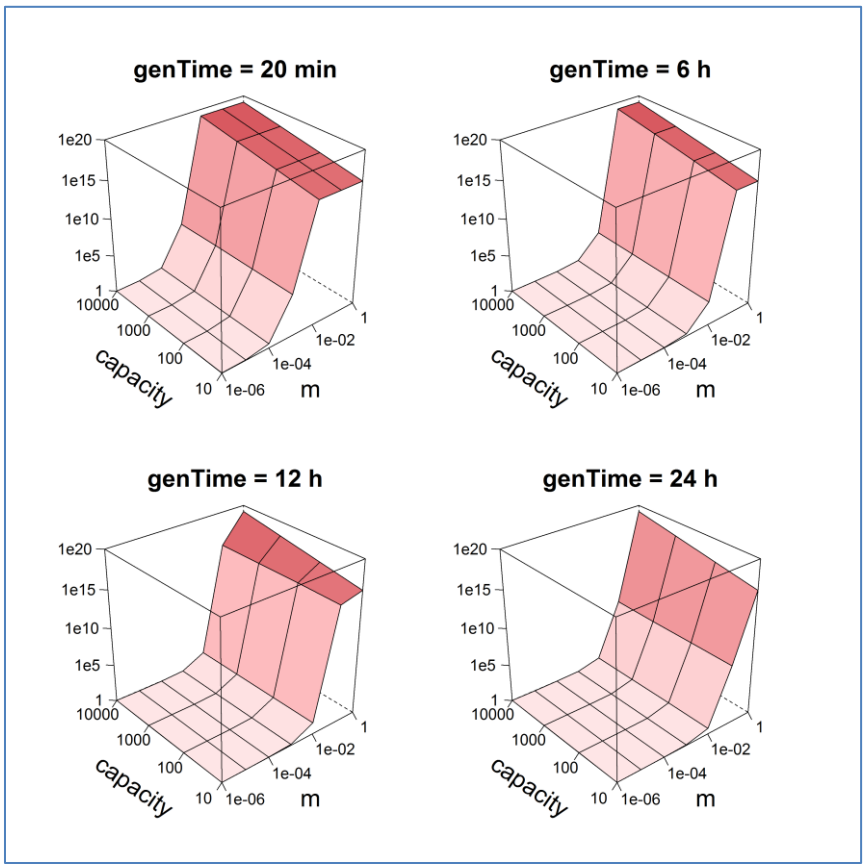


Figure 40: Surface plot of the number of bacterial cells carrying the novel gene in a manure tank.

Simulation was performed for the situation occurring 6 months after the introduction of a single altered cell, for different values of the bacterial generation time, the capacity of the tank in m^3 and the selection coefficient (*m*).

4.4.4 Conclusions

The results of the simulation modelling show that a single copy of a mosaic gene that is introduced into the bacterial host population in an agricultural liquid manure tank is propagated within the host population in the presence of a selective advantage. Under strong positive selection pressure (selection coefficients of 10^{-4} or larger), the effect becomes rather strong, resulting in a considerable number of bacterial cells carrying the novel gene which are eventually released into the environment.

5 Risk Assessment and Recommendations

5.1 Risk Assessment

In the case of mosaic genes the hazard is the formation of genetic entities coding for altered proteins which allow bacteria to develop novel resistances to antimicrobial agents, to escape the immune defence of the eukaryotic host, to increase their virulence or to help pathogens to escape vaccination schemes. The risk of mosaic genes to induce adverse effects on human and animal health is, thus, not restricted to the issue of antibiotic resistance development and spread; however, the impact of mosaic gene formation in the area of antibiotic resistance is more comprehensively documented in the scientific literature, the effects of mosaic gene formation are worsening the generally already deteriorating situation in antibiotic resistance and the repercussions on public health are immanent and requests for actions.

The risk for adverse effects on human and animal health raised by mosaic gene formation is therefore assessed with a focus on antibiotic resistance.

5.1.1 General Considerations

The formation of mosaic genes may take place directly in human or animal pathogens or may occur in non-pathogenic environmental or commensal bacteria. In the latter cases the generated mosaic genes may reach pathogenic bacteria and be integrated into their genomes via a cascade of different horizontal gene transfer events utilizing transformation, conjugation or transduction of the complete gene and various intermediate bacterial hosts.[412-414] The common result of both scenarios is an increase in morbidity and mortality of infected patients due to prolonged hospitalizations, more treatment failures, and an increase in personal suffering culminating in an elevation of the financial burden in public health.[415, 416] A similar situation is encountered in animal husbandry where mosaic genes may be the cause for a more frequent application and higher dosages of antibiotics in prophylaxis and therapy eventually leading to products, manure and wastes more contaminated with antimicrobial agents or directly to substantial losses of revenues for the producer due to treatment failures and resulting slaughter of total animal flocks and herds.[405, 408, 417]

The formation of mosaic antibiotic resistance genes expands the environmental resistome and increases the variability of the global resistance gene pool enhancing the chance for the development of clinically relevant resistant strains.[418-421]

This development may be rapid in environments under strong selection pressure (e.g. in clinical settings with patients treated with antibiotics or prophylactic or therapeutic application of antimicrobial agents on a farm scale level by veterinarians), but due to substantial knowledge gaps concerning the modes of action of selection prevailing in the respective habitats, neither the period of time necessary for a mosaic gene to become clinically relevant nor the kind of the resulting new resistance functions are predictable.[16, 26-29]

5.1.2 Transgenic Plants

Antibiotic resistance marker genes from transgenic plants may interfere with mosaic gene formation by providing the substrate for transformation and recombination with already endogenously present similar antibiotic resistance determinants. The incoming DNA – under field conditions usually damaged, mutated and/or genetically modified - may lead to sequence rearrangements and mutations in the chromosomal resistance gene.[77, 422] Transgenic plant DNA is still in many events of microbiological origin or contains prokaryotic vector backbone fragments (as remnants of the genetic modification process in the laboratory) which may provide the required anchor sequences for homology-directed illegitimate recombination.[17, 213] Upon disruption of the cell wall due to the decay of plant material transgenic DNA is part of the extracellular fraction of environmental DNA, may be stable in soil for a prolonged period of time or may survive the passage through the mammalian gastrointestinal tract.[217, 360] In all cases transgenic DNA of microbiological origin has the opportunity to get into contact with environmental bacteria in the exposed habitat, is a substrate for DNA uptake by competent bacterial cells and may take part in the reshuffling and recombination of DNA fragments involved in the generation of segmented genes consisting of sequence regions of different phylogenetic origin.[19] The affected bacterial strain may become more resistant to an antibiotic compound requiring higher dosage with potentially more adverse effects on the patient or expand its antibiotic inactivation spectrum by alterations of the substrate specificity of the enzyme or binding protein.

5.1.3 Frequency of Mosaic Gene Formation in Natural Habitats

The formation of mosaic genes induced by transgenic plant DNA has to be seen in the context of the abundance of horizontal gene transfers and mosaic gene formations taking already place in naturally occurring bacterial populations in any case. Mosaic gene formation is a naturally occurring process in all bacteria rendering probably all genes of the bacterial chromosome as potential targets for segmented gene rearrangements.[49, 50] It can be reasonably assumed that the segmented exchange of gene fragments is a relatively abundant process taking into account that many of these fragmented transfers are silent or deleterious leading to no detectable phenotype for the experimenter.[73] There are essentially no restrictions for the fragment lengths to be exchanged nor are there any gene or protein domain boundaries constraining the genomic target sites for fragment exchange to functional units.[48, 49] The frequency of mosaic gene formation in bacterial populations is species-specific and substantially dependent on transformability, sequence divergence of the involved DNA molecules, on the efficiency of the recombination and mismatch repair apparatus and to a substantial extent on the selection pressure providing the means for maintenance and expansion of the mosaic gene in the population.[3, 204, 239] However, there are no experimental data available dealing explicitly with the frequency of mosaic gene formations in naturally occurring bacterial populations to date.

5.1.4 Exposure Levels of Natural Habitats with Transgenic Plant Gene Derived Mosaic Genes

It is scientific consensus that the transfer of plant DNA to competent soil or gut bacteria is an extremely rare event and that this process had not been demonstrated to occur in the field without providing optimized conditions for gene transfer and recombination.[20] It is therefore justified to assume an extremely low exposure level of natural habitats with mosaic genes generated with transgenic plant DNA involvement. There are theoretically calculated transformation frequencies for full length gene transfers from plant to bacteria of 10^{-17} or even lower.[20] Jack Heinemann has calculated the probability for the formation of a clinically relevant mosaic penicillin binding protein in *Streptococcus pneumoniae* to be approximately 10^{-24}.[16] Nevertheless this completely unlikely event took place and became – within a period of four to five decades of ß-lactam induced selection – a major cause of therapy failure in pneumococcal pneumonia in clinical settings.[16] It is important to bear in mind that only a single successful mosaic gene formation event may be sufficient to create a pathogenic bacterial strain with a novel resistance function of clinical relevance.

The low probability of the formation of a mosaic gene is supported by the results obtained in the simulated gastrointestinal environment of pigs for a single animal. The data suggest that the formation of a mosaic gene will not occur during the average life span of an individual pig with high confidence. However, if the simulation is transferred to a more global level taking into account the total annual pig production in Austria, it could be demonstrated that this process of a biologically relevant mosaic gene formation becomes more likely. The data indicate that under a sufficiently high selection pressure mosaic gene formation and its fixation in the bacterial population of all pigs is to be expected within the given time period of one year. It is to be expected that animals in large scale animal husbandry are experiencing antibiotic selection pressure during their life time.[423-425] These results indicate that an extremely rare event indicative for a generally low exposure level may become significant due to the large number of exposed bacteria (i.e. in several millions of porcine gastrointestinal tracts). Moreover, it has to be stressed that antibiotic resistance genes are auto-replicative environmental pollutants. They are not degraded or diluted over time as common contaminants in natural habitats but are usually amplified on replicative units (i.e. plasmids, transposons) and by bacterial cell division potentially aggravating the problem over time. A profoundly low probability for an event is, thus, not predictive for long-term adverse effects on human or animal health in the case of antibiotic resistance genes and bacterial populations. The laboratory model for the formation of aminoglycoside phosphotransferase mosaic genes in *Acinetobacter baylyi* did not provide evidence for a segmented integration of *nptII*-derived donor DNA. However, an anchor sequence dependent full length integration of an intact copy of *nptII* could be demonstrated utilizing a short 12 - 15 bp region of microhomology with the chromosomal target region. The event occurred with a frequency of approximately 10^{-7}. This observation is another indication that prokaryotic vector sequences remaining in the transgenic plant are of importance because they are capable to provide homologous anchor sequences necessary for the integration of otherwise heterologous DNA into the bacterial chromosome. The analysis of sequence databases for the presence of mosaic genes revealed the utility of sequence comparisons for assessing the risks of antibiotic resistance marker genes. *NptII* could be established as comparably stable genetic unit in bacterial populations for which the respective sequence information was available. Only a single recombination event could be identified to have occurred in evolutionary terms in the available GeneBank sequence data sets. Recombination of the codon optimized plant DNA derived copy of

epsps with similar *epsps* gene homologs in soil or plant associated bacterial communities appear to be also unlikely considering the high grade of sequence divergence among plant and bacterial *epsps* genes. However, with both analyses uncertainties concerning the representativeness of the deposited GenBank sequences have to be taken into consideration for the risk assessment. Considering the constantly growing number of deposited genes and whole microbial genome sequences the performed *in silico* analysis for recombination breakpoints is only valid for the number of sequences available at the time of modelling. The analysis may lead to different results if new sequence data become available.

In conclusion the exposure level of naturally occurring bacterial populations with transgenic plant gene derived mosaic genes is expected to be low. However, a low probability of an event is not predictive for the absence of long term adverse effects on human or animal health.[27]

5.1.5 Knowledge Gaps

The following knowledge gaps have to be taken into account for the risk assessment of mosaic genes:

1. Only a small proportion of bacterial species have been demonstrated experimentally to be transformable. The actual abundance of transformable species in natural environments is unknown.

2. The conditions which induce competence in natural environments are largely unknown for the broad majority of bacterial species carrying genomic signatures of competence genes.

3. Experimental data of the actual frequency of mosaic gene formation in natural environments are not available.

4. Modus, strengths and duration of selection pressure in natural habitats necessary for selection and maintaining the mosaic gene in the bacterial population are highly variable or not known at all.

5. GenBank deposited bacterial sequences for recombination analysis may not be representative for naturally occurring bacterial populations exposed to transgenic plant DNA.

6. Sequence databases are inherently incomplete. However, software algorithms for the detection of recombination events are extremely sensitive for the correct and representative collection of sequence datasets for *in silico* recombination breakpoint detection to obtain meaningful results.

5.1.6 Conclusions

The likelihood for the formation of mosaic antibiotic resistance genes with transgenic plant DNA involvement is low. It is justified to assume that similar mosaic gene formation processes will occur with a significantly higher rate already naturally in bacterial populations compared to processes involving transgenic plant DNA (i.e. the naturally occurring background rate will be comparably high). It is therefore questionable whether the relative contribution of transgenic plant DNA for the formation of mosaic genes will be high enough to be of biological relevance. However, this assumption remains to be verified experimentally. There are substantial knowledge gaps concerning the actual frequency of mosaic gene formation and the kind of selection pressure prevailing in natural habitats. Moreover, a low likelihood of such an event is not predictive for an absence of

adverse long-term effects induced by this event. Devastating adverse effects on public health of an extremely unlikely mosaic penicillin resistance gene formation could be demonstrated in a clinical setting. It is therefore advisable that risk assessment bodies take the formation of mosaic genes on their agenda and take it into account on a routine basis for the risk assessment of transgenic plants containing microbial-derived DNA. Additionally, it would be necessary to increase research efforts on this topic to narrow the still prevailing knowledge gaps and reduce the uncertainties currently linked with the risk assessment of mosaic genes.

5.2 Recommendations

Although still major knowledge gaps concerning the issue of mosaic gene formation in natural habitats exist the implementation of the following strategies is recommended:

1. The formation of mosaic genes in the field of transgenic plants and the generation and dissemination of novel antibiotic resistance determinants is a recently emerging problem not yet in the focus of the official risk assessment bodies. It is, thus, recommended to increase the awareness on this topic and to sensitize EFSA and the European Commission for the impact of mosaic gene formation for the risk assessment of transgenic plants.

2. In this respect EFSA should be asked to develop guidance on the proper risk assessment of mosaic genes and to provide strategies and recommendations to minimize the formation of mosaic genes in the interplay between transgenic plant DNA and exposed bacterial populations.

3. The results obtained in this project highlight the importance of the availability of sequence information about the transgenic insert including its flanking genomic regions. In light of the scheduled revision of *Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed* it would be reasonable to include mandatory bioinformatic sequence similarity analyses in the regulatory framework comparing transgenic sequences with the pool of homologous sequences already endogenously present in exposed bacterial populations to check for a possible elevated recombination potential of the transgenic insert with naturally occurring counterparts. However, it is obvious that the quality of these analyses is dependent on the quality of the accessible (GenBank) database entries.

4. Additionally, it should be taken into consideration to include mandatory *in silico* recombination analyses using software algorithms designed to detect potential recombination breakpoints in the transgene and similar sequences in exposed bacterial populations as amendment to Regulation (EU) No 503/2013 if this regulation is to be revised. Recombination analysis provides valuable information about the genetic stability of the involved sequences in evolutionary terms and their potential to engage in mosaic gene formation. Highly recombinogenic transgene sequences should be avoided or at least modified for optimal codon usage in plants (see next paragraph). It would be also advisable to sensitize the applicants on the issue of mosaic gene formation as potential hazard and try to obtain their committment to provide appropriate recombination analysis data on a routine basis.

5. In this respect the optimization of bacterial transgene sequences to the codon usage of the transformed plant is a valuable tool to reduce sequence similarities between transgene and potential acceptor DNA sequences present in bacterial populations. This procedure is usually already applied by the applicant to support a high expression rate of the transgene in the plant. The obtained results indicate that this strategy is also beneficial for reducing the potential for homologous recombination with similar counterparts in exposed bacterial populations. Codon usage optimization should be strongly promoted.

6. It is recommended to raise the awareness concerning the potential impact of extremely rare events especially in the case of antibiotic resistance gene or gene fragment transfers in the risk assessment community: a low likelihood of such an event is not predictive for an absence of any adverse long-term effects induced by this event. In the case of mosaic resistance gene formation the

model simulations show that this extremely rare event is counterbalanced by an extremely high number of exposed cells, a long term exposure and a sufficiently strong selection pressure.

7. It is important to realize that the currently available technology for the detection of gene or gene fragment transfers is incapable for the detection of rare events in natural environments. Detection is in any case temporarily delayed until the event reaches a frequency in the population sufficient for the detection with the commonly available arsenal of methods. Supporting research projects with a focus on improving the sensitivity of methods to detect rare horizontal sequence transfer events in natural environments would be beneficial.

8. Selection pressure appeared to be the key driver for fixation of a mosaic gene in a bacterial population in the calculated model systems. It is important to note that also subinhibitory antibiotic concentrations (i.e. a low selection pressure) which do not kill bacteria but are readily present in natural environments are major drivers for selection and maintenance of antibiotic resistant strains. Therefore it is recommended that a further release of DNA fragments capable to interfere with mosaic antibiotic resistance gene development should be minimized as far as possible.

9. In light of the precautionary principle bacterial transgenes with high similarities to existing counterparts in exposed naturally occurring bacterial populations should be avoided in genetically modified plants to minimize the potential for mosaic gene formation. This is especially necessary in the case of antibiotic resistance marker genes and an additional argument in support of Article 4 (2) of Directive 2001/18/EC, which requires a step-by-step phasing out of antibiotic resistance marker genes in GMOs which may have adverse effects on human health and the environment by the end of 2004 (concerning GMOs released for marketing according to part C) and by the end 2008 (concerning the deliberate release of GMOs into the environment according to part B of the directive).

6 References

1. Chen I, Dubnau D. DNA uptake during bacterial transformation. *Nature Reviews Microbiology* 2004; **2**: 241-9.
2. Bakkali M. Could DNA uptake be a side effect of bacterial adhesion and twitching motility? *Arch Microbiol* 2013; **195**: 279-89.
3. Johnston C, Martin B, Fichant G *et al.* Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol* 2014; **12**: 181-96.
4. EFSA. Scientific opinion on guidance for the risk assessment of genetically modified plants used for non-food or non-feed purposes. *The EFSA Journal* 2009; **1164**: 1-42.
5. EFSA. Statement of the Scientific Panel on Genetically Modified Organisms on the safe use of the nptII antibiotic resistance marker gene in genetically modified plants. http://www.efsa.europa.eu/EFSA/efsa_locale-1178620753824_1178620775641.htm. *The EFSA Journal* 2007: 1-7.
6. EFSA. Opinion of the Scientific Panel on Genetically Modified Organisms on the use of antibiotic resistance genes as marker genes in genetically modified plants. *The EFSA Journal* 2004; **48**: 1-18.
7. Zheng Y, Roberts RJ, Kasif S. Segmentally Variable Genes:A New Perspective on Adaptation. *PLoS Biol* 2004; **2**: 452-64.
8. Lorenz MG, Wackernagel W. Bacterial gene transfer by natural transformation in the environment. *Microbiol Mol Biol Rev* 1994; **58**: 5563-602.
9. Guerriero G. Putative chitin synthases from Branchiostoma floridae show extracellular matrix-related domains and mosaic structures. *Genomics Proteomics Bioinformatics* 2012; **10**: 197-207.
10. Bourkal'tseva MV, Krylov SV, Kropinski AM *et al.* Bacteriophage phi297, a new species of Pseudomonas aeruginosa temperate phages with a mosaic genome: Potential use in phage therapy. *Russ J Genet* 2011; **47**: 794-8.
11. Johansen BK, Wasteson Y, Granum PE *et al.* Mosaic structure of Shiga-toxin-2-encoding phages isolated from Escherichia coli O157:H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology* 2001; **147**: 1929-36.
12. Geysen D, Bazarusanga T, Brandt J *et al.* An unusual mosaic structure of the PIM gene of Theileria parva and its relationship to allelic diversity. *Mol Biochem Parasitol* 2004; **133**: 163-74.
13. Boc A, Makarenkov V. Towards an accurate identification of mosaic genes and partial horizontal gene transfers. *Nucleic Acids Res* 2011; **39**.
14. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol* 1992; **34**: 126-9.
15. Spratt B. Resistance to antibiotics mediated by target alterations. *Science* 1994; **264**: 388-93.
16. Heinemann JA, Traavik T. Problems in monitoring horizontal gene transfer in field trials of transgenic plants. *Nat Biotechnol* 2004; **22**: 1105-9.
17. Kleter GA, Peijnenburg AACM, Aarts HJM. Health considerations regarding horizontal transfer of microbial transgenes present in genetically modified crops. *J Biomed Biotechnol* 2005: 326-52.
18. Pontiroli A, Ceccherini MT, Pote J *et al.* Long-term persistence and bacterial transformation potential of transplastomic plant DNA in soil. *Res Microbiol* 2010; **161**: 326-34.
19. Domingues S, Nielsen KM, da Silva GJ. Various pathways leading to the acquisition of antibiotic resistance by natural transformation. *Mobile Genetic Elements* 2012; **2**: 257-60.
20. EFSA. Consolidated presentation of the joint scientific opinion of the GMO and BIOHAZ Panels on the "Use of Antibiotic Resistance Genes as Marker Genes in Genetically Modified Plants" and the Scientific Opinion of the GMO Panel on "Consequences of the Opinion on the Use of Antibiotic Resistance Genes as Marker Genes in Genetically Modified Plants on Previous EFSA Assessments of Individual GM Plants". *The EFSA Journal* 2009; **1108**: 1-8.
21. Bennett PM, Livesey CT, Nathwani D *et al.* An assessment of the risks associated with the use of antibiotic resistance genes in genetically modified plants: report of the Working Party of the British Society for Antimicrobial Chemotherapy. *J Antimicrob Chemother* 2004; **53**: 418-31. Epub 2004 Jan 28.

22. Flavell RB, Dart E, Fuchs RL *et al.* Selectable marker genes: safe for plants? *Biotechnology (N Y)* 1992; **10**: 141-4.
23. Nap JP, Bijvoet J, Stiekema WJ. Biosafety of kanamycin-resistant transgenic plants. *Transgenic Res* 1992; **1**: 239-49.
24. Goldstein DA, Tinland B, Gilbertson LA *et al.* Human safety and genetically modified plants: a review of antibiotic resistance markers and future transformation selection technologies. *J Appl Microbiol* 2005; **99**: 7-23.
25. Demanèche S, Sanguin H, Pote J *et al.* Antibiotic-resistant soil bacteria in transgenic plant fields. *Proc Natl Acad Sci U S A* 2008; **105**: 3957-62.
26. Nielsen KM, Townsend JP. Monitoring and modeling horizontal gene transfer. *Nat Biotechnol* 2004; **22**: 1110-4.
27. Pettersen AK, Bohn T, Primicerio R *et al.* Modeling suggests frequency estimates are not informative for predicting the long-term effect of horizontal gene transfer in bacteria. *Environ Biosafety Res* 2005; **4**: 223-33.
28. Nielsen KM, Bohn T, Townsend JP. Detecting rare gene transfer events in bacterial populations. *Front Microbiol* 2014; **4**: 415.
29. Townsend JP, Bohn T, Nielsen KM. Assessing the probability of detection of horizontal gene transfer events in bacterial populations. *Frontiers in Antimicrobials, Resistance and Chemotherapy* 2012; **3**: doi:10.3389/fmicb.2012.00027. 2012.
30. Nielsen KM. An assessment of factors affecting the likelihood of horizontal transfer of a recombinant plant DNA to bacterial recipients in the soil and phytosphere. *Collection of Biosafety Reviews (ICGEB, Italy)* 2003; **1**: 96-149.
31. Boc A, Makarenkov V. Towards an accurate identification of mosaic genes and partial horizontal gene transfers. *Nucleic Acids Res* 2011; **39**: e144.
32. Hollingshead SK, Becker R, Briles DE. Diversity of PspA: Mosaic genes and evidence for past recombination in Streptococcus pneumoniae. *Infect Immun* 2000; **68**: 5889-900.
33. Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 2002; **19**: 2226-38.
34. Zhaxybayeva O, Lapierre P, Gogarten JP. Genome mosaicism and organismal lineages. *Trends Genet* 2004; **20**: 254-60.
35. Boc A, Philippe H, Makarenkov V. Inferring and validating horizontal gene transfer events using bipartition dissimilarity. *Syst Biol* 2010; **59**: 195-211.
36. Davies RL, Whittam TS, Selander RK. Sequence Diversity and Molecular Evolution of the Leukotoxin (lktA) Gene in Bovine and Ovine Strains ofMannheimia (Pasteurella)haemolytica. *J Bacteriol* 2001; **183**: 1394-404.
37. Doolittle WF. Phylogenetic classification and the universal tree. *Science* 1999; **284**: 2124-9.
38. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000; **405**: 299-304.
39. Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* 2005; **3**: 679-88.
40. van Hoek AH, Mevius D, Guerra B *et al.* Acquired antibiotic resistance genes: an overview. *Front Microbiol* 2011; **2**: 203.
41. Koonin EV. Horizontal gene transfer: the path to maturity. *Mol Microbiol* 2003; **50**: 725-7.
42. Doolittle WF, Boucher Y, Nesbo CL *et al.* How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans R Soc Lond B Biol Sci* 2003; **358**: 39-57; discussion -8.
43. Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 2008; **62**: 53-70.
44. Ochman H, Lerat E, Daubin V. Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci U S A* 2005; **102 Suppl 1**: 6595-9.
45. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* 2005; **3**: 711-21.

46. Kloesges T, Popa O, Martin W *et al.* Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol Biol Evol* 2011; **28**: 1057-74.
47. Hao W. Extensive genomic variation within clonal bacterial groups resulted from homologous recombination. *Mobile Genetic Elements* 2013; **3**: e23463.
48. Le PT, Pontarotti P, Raoult D. Alphaproteobacteria species as a source and target of lateral sequence transfers. *Trends Microbiol* 2014; **22**: 147-56.
49. Chan CX, Darling AE, Beiko RG *et al.* Are protein domains modules of lateral genetic transfer? *PLoS One* 2009; **4**: e4524.
50. Dagan T, Martin W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci U S A* 2007; **104**: 870-5.
51. Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res* 2012; **40**: W573-9.
52. Martin DP. RDP3 Instruction Manual. http://web.cbio.uct.ac.za/~darren/RDPManual.pdf. 2010.
53. Zhaxybayeva O, Doolittle WF. Lateral gene transfer. *Curr Biol* 2011; **21**: R242-R6.
54. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science* 1997; **278**: 631-7.
55. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* 1999; **96**: 3801-6.
56. Prudhomme M, Libante V, Claverys JP. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in Streptococcus pneumoniae. *Proc Natl Acad Sci U S A* 2002; **99**: 2100-5.
57. Dowson CG, Barcus V, King S *et al.* Horizontal gene transfer and the evolution of resistance and virulence determinants in Streptococcus. *J Appl Microbiol* 1997; **26**: 42S-51S.
58. Doern GV. Antimicrobial use and the emergence of antimicrobial resistance with Streptococcus pneumoniae in the United States. *Clin Infect Dis* 2001; **33 Suppl 3**: S187-92.
59. Keeling PJ, Palmer JD. Lateral transfer at the gene and subgenic levels in the evolution of eukaryotic enolase. *Proc Natl Acad Sci U S A* 2001; **98**: 10745-50.
60. Lilley AK, Bailey MJ, Barr M *et al.* Population dynamics and gene transfer in genetically modified bacteria in a model microcosm. *Mol Ecol* 2003; **12**: 3097-107.
61. Molbak L, Licht TR, Kvist T *et al.* Plasmid transfer from Pseudomonas putida to the indigenous bacteria on alfalfa sprouts: characterization, direct quantification, and in situ location of transconjugant cells. *Appl Environ Microbiol* 2003; **69**: 5536-42.
62. du Plessis M, Bingen E, Klugman KP. Analysis of penicillin-binding protein genes of clinical isolates of Streptococcus pneumoniae with reduced susceptibility to amoxicillin. *Antimicrob Agents Chemother* 2002; **46**: 2349-57.
63. Gasson M, Burke D. Scientific perspectives on regulating the safety of genetically modified foods. *Nat Rev Genet* 2001; **2**: 217-22.
64. de Vries J, Heine M, Harms K *et al.* Spread of recombinant DNA by roots and pollen of transgenic potato plants, identified by highly specific biomonitoring using natural transformation of an Acinetobacter sp. *Appl Environ Microbiol* 2003; **69**: 4455-62.
65. Nielsen KM, Bones AM, Smalla K *et al.* Horizontal gene transfer from transgenic plants to terrestrial bacteria--a rare event? *FEMS Microbiol Rev* 1998; **22**: 79-103.
66. Nielsen KM, van Elsas JD, Smalla K. Transformation of Acinetobacter sp. strain BD413 (pFG4-delta-nptII) with transgenic plant DNA in soil microcosms and effects of kanamycin on selection of transformants. *Appl Environ Microbiol* 2000; **66**: 1237-42.
67. Kay E, Vogel TM, Bertolla F *et al.* In situ transfer of antibiotic resistance genes from transgenic (transplastomic) tobacco plants to bacteria. *Appl Environ Microbiol* 2002; **68**: 3345-51.
68. Tepfer D, Garcia-Gonzales R, Mansouri H *et al.* Homology-dependent DNA transfer from plants to a soil bacterium under laboratory conditions: implications in evolution and horizontal gene transfer. *Transgenic Res* 2003; **12**: 425-37.

69. Schlüter K, Futterer J, Potrykus I. "Horizontal" gene transfer from a transgenic potato line to a bacterial pathogen (Erwinia chrysanthemi) occurs - if at all - at an extremely low frequency. *Biotechnology (N Y)* 1995; **13**: 1094-8.
70. Nielsen KM, van Weerelt MD, Berg TN *et al.* Natural transformation and availability of transforming DNA to Acinetobacter calcoaceticus in soil microcosms. *Appl Environ Microbiol* 1997; **63**: 1945-52.
71. Dubnau D. DNA uptake in bacteria. *Annual Rev Microbiol* 1999; **53**: 217-44.
72. Woegerbauer M, Jenni B, Thalhammer F *et al.* Natural genetic transformation of clinical isolates of Escherichia coli in urine and water. *Appl Environ Microbiol* 2002; **68**: 440-3.
73. Didelot X, Maiden MC. Impact of recombination on bacterial evolution. *Trends Microbiol* 2010; **18**: 315-22.
74. Fontaine L, Wahl A, Flechard M *et al.* Regulation of competence for natural transformation in streptococci. *Infect Genet Evol* 2014.
75. Smith JM, Dowson CG, Spratt BG. Localized sex in bacteria. *Nature* 1991; **349**: 29-31.
76. Chan CX, Beiko RG, Darling AE *et al.* Lateral transfer of genes and gene fragments in prokaryotes. *Genome Biol Evol* 2009; **1**: 429-38.
77. Overballe-Petersen S, Harms K, Orlando LA *et al.* Bacterial natural transformation by highly fragmented and damaged DNA. *Proc Natl Acad Sci U S A* 2013; **110**: 19860-5.
78. Lin EA, Zhang XS, Levine SM *et al.* Natural transformation of helicobacter pylori involves the integration of short DNA fragments interrupted by gaps of variable size. *PLoS Pathog* 2009; **5**: e1000337.
79. Spratt BG, Bowler LD, Zhang Q-Y *et al.* Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species. *J Mol Evol* 1992; **34**: 115-25.
80. Croucher NJ, Harris SR, Fraser C *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011; **331**: 430-4.
81. Dubnau D. Genetic competence in Bacillus subtilis. *Microbiol Rev* 1991; **55**: 395-424.
82. Stewart GJ, Carlson CA. The biology of natural transformation. *Annu Rev Microbiol* 1986; **40**: 211-35.
83. Moradigaravand D, Engelstadter J. The impact of natural transformation on adaptation in spatially structured bacterial populations. *BMC Evol Biol* 2014; **14**: 141.
84. Seitz P, Blokesch M. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. 2013.
85. Lo Scrudato M, Blokesch M. The regulatory network of natural competence and transformation of Vibrio cholerae. *PLoS Genet* 2012; **8**: e1002778.
86. Averhoff B. Shuffling genes around in hot environments: the unique DNA transporter of Thermus thermophilus. *FEMS Microbiol Rev* 2009; **33**: 611-26.
87. Israel DA, Lou AS, Blaser MJ. Characteristics of Helicobacter pylori natural transformation. *FEMS Microbiol Lett* 2000; **186**: 275-80.
88. Morikawa K, Takemura AJ, Inose Y *et al.* Expression of a cryptic secondary sigma factor gene unveils natural competence for DNA transformation in Staphylococcus aureus. *PLoS Pathog* 2012; **8**: e1003003.
89. Johnsborg O, Eldholm V, Havarstein LS. Natural genetic transformation: prevalence, mechanisms and function. *Res Microbiol* 2007; **158**: 767-78.
90. Lipscomb GL, Stirrett K, Schut GJ *et al.* Natural competence in the hyperthermophilic archaeon Pyrococcus furiosus facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl Environ Microbiol* 2011; **77**: 2232-8.
91. Sato T, Fukui T, Atomi H *et al.* Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *J Bacteriol* 2003; **185**: 210-20.
92. de Vries J, Wackernagel W. Microbial horizontal gene transfer and the DNA release from transgenic crop plants. *Plant Soil* 2004; **266**: 91-104.
93. Redfield RJ. Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation. *J Hered* 1993; **84**: 400-4.

94. Sinha S, Mell J, Redfield R. The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. *Mol Microbiol* 2013; **88**: 1106-19.
95. Claverys JP, Prudhomme M, Martin B. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol* 2006; **60**: 451-75.
96. Sinha S, Mell JC, Redfield RJ. Seventeen Sxy-dependent cyclic AMP receptor protein site-regulated genes are needed for natural transformation in Haemophilus influenzae. *J Bacteriol* 2012; **194**: 5245-54.
97. Goodman SD, Scocca JJ. Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae. *Proc Natl Acad Sci U S A* 1988; **85**: 6982-6.
98. Danner DB, Deich RA, Sisco KL *et al.* An eleven-base-pair sequence determines the specificity of DNA uptake in Haemophilus transformation. *Gene* 1980; **11**: 311-8.
99. Cameron AD, Redfield RJ. Non-canonical CRP sites control competence regulons in Escherichia coli and many other gamma-proteobacteria. *Nucleic Acids Res* 2006; **34**: 6001-14.
100. Mell JC, Shumilina S, Hall IM *et al.* Transformation of natural genetic variation into Haemophilus influenzae genomes. *PLoS Pathog* 2011; **7**: e1002151.
101. Berge M, Mortier-Barriere I, Martin B *et al.* Transformation of Streptococcus pneumoniae relies on DprA- and RecA-dependent protection of incoming DNA single strands. *Mol Microbiol* 2003; **50**: 527-36.
102. Snyder L, Peters JE, Henkin TM *et al.* Molecular mechanisms of homologous recombination. . *Molecular Genetics of Bacteria*. Washington, DC: ASM Press, 2013; 403-31.
103. Lee MS, Morrison DA. Identification of a new regulator in Streptococcus pneumoniae linking quorum sensing to competence for genetic transformation. *J Bacteriol* 1999; **181**: 5004-16.
104. Mohan S, Dubnau D. Transcriptional regulation of comC: evidence for a competence-specific transcription factor in Bacillus subtilis. *J Bacteriol* 1990; **172**: 4064-71.
105. Redfield RJ. sxy-1, a Haemophilus influenzae mutation causing greatly enhanced spontaneous competence. *J Bacteriol* 1991; **173**: 5612-8.
106. Meibom KL, Blokesch M, Dolganov NA *et al.* Chitin induces natural competence in Vibrio cholerae. *Science* 2005; **310**: 1824-7.
107. Barany F, Kahn ME, Smith HO. Directional transport and integration of donor DNA in Haemophilus influenzae transformation. *Proc Natl Acad Sci U S A* 1983; **80**: 7274-8.
108. Mejean V, Claverys JP. Use of a cloned DNA fragment to analyze the fate of donor DNA in transformation of Streptococcus pneumoniae. *J Bacteriol* 1984; **158**: 1175-8.
109. Mell JC, Hall IM, Redfield RJ. Defining the DNA uptake specificity of naturally competent Haemophilus influenzae cells. *Nucleic Acids Res* 2012; **40**: 8536-49.
110. Claverys JP, Martin B, Havarstein LS. Competence-induced fratricide in streptococci. *Mol Microbiol* 2007; **64**: 1423-33.
111. Claverys JP, Havarstein LS. Cannibalism and fratricide: mechanisms and raisons d'etre. *Nat Rev Microbiol* 2007; **5**: 219-29.
112. Johnsborg O, Eldholm V, Bjornstad ML *et al.* A predatory mechanism dramatically increases the efficiency of lateral gene transfer in Streptococcus pneumoniae and related commensal species. *Mol Microbiol* 2008; **69**: 245-53.
113. Morrison DA, Mortier-Barriere I, Attaiech L *et al.* Identification of the major protein component of the pneumococcal eclipse complex. *J Bacteriol* 2007; **189**: 6497-500.
114. Johnston C, Martin B, Granadel C *et al.* Programmed protection of foreign DNA from restriction allows pathogenicity island exchange during pneumococcal transformation. *PLoS Pathog* 2013; **9**: e1003178.
115. Johnston C, Polard P, Claverys JP. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. *Mob Genet Elements* 2013; **3**: e25582.
116. Hofer F. Transfer of lactose-fermenting ability in Lactobacillus lactis. *NZJ Dairy Sci Technol* 1985; **20**: 179-83.

117. Tannock GW. A special fondness for lactobacilli. *Appl Environ Microbiol* 2004; **70**: 3189-94.

118. Chen I, Christie PJ, Dubnau D. The ins and outs of DNA transfer in bacteria. *Science* 2005; **310**: 1456-60.
119. Zaccaria E, van Baarlen P, de Greeff A *et al.* Control of competence for DNA transformation in *Streptococcus suis* by genetically transferable pherotypes. *PLoS One* 2014; **9**: e99394.
120. Baltrus DA, Guillemin K. Multiple phases of competence occur during the Helicobacter pylori growth cycle. *FEMS Microbiol Lett* 2006; **255**: 148-55.
121. Dorer MS, Fero J, Salama NR. DNA damage triggers genetic exchange in Helicobacter pylori. *PLoS Pathog* 2010; **6**: e1001026.
122. Morrison DA, Guedon E, Renault P. Competence for Natural Genetic Transformation in the S. bovis Group Streptococci, S. infantarius and S. macedonicus. *J Bacteriol* 2013.
123. Charpentier X, Polard P, Claverys JP. Induction of competence for genetic transformation by antibiotics: convergent evolution of stress responses in distant bacterial species lacking SOS? *Curr Opin Microbiol* 2012; **15**: 570-6.
124. Charpentier X, Kay E, Schneider D *et al.* Antibiotics and UV radiation induce competence for natural transformation in Legionella pneumophila. *J Bacteriol* 2011; **193**: 1114-21.
125. Juni E, Heym GA. Transformation assay for identification of psychrotrophic achromobacters. *Appl Environ Microbiol* 1980; **40**: 1106-14.
126. Wilharm G, Piesker J, Laue M *et al.* DNA uptake by the nosocomial pathogen Acinetobacter baumannii occurs during movement along wet surfaces. *J Bacteriol* 2013; **195**: 4146-53.
127. Nielsen KM, Bones AM, Van Elsas JD. Induced natural transformation of Acinetobacter calcoaceticus in soil microcosms. *Appl Environ Microbiol* 1997; **63**: 3972-7.
128. Palmen R, Vosman B, Buijsman P *et al.* Physiological characterization of natural transformation in Acinetobacter calcoaceticus. *J Gen Microbiol* 1993; **139**: 295-305.
129. Demanèche S, Kay E, Gourbiere F *et al.* Natural transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in soil. *Appl Environ Microbiol* 2001; **67**: 2617-21.
130. Tonjum T, Bovre K, Juni E. Fastidious gram-negative bacteria: meeting the diagnostic challenge with nucleic acid analysis. *APMIS* 1995; **103**: 609-27.
131. Lorenz MG, Reipschlager K, Wackernagel W. Plasmid transformation of naturally competent Acinetobacter calcoaceticus in non-sterile soil extract and groundwater. *Arch Microbiol* 1992; **157**: 355-60.
132. Tonjum T, Bukholm G, Bovre K. Identification of Haemophilus aphrophilus and Actinobacillus actinomycetemcomitans by DNA-DNA hybridization and genetic transformation. *J Clin Microbiol* 1990; **28**: 1994-8.
133. Bosse JT, Sinha S, Schippers T *et al.* Natural competence in strains of Actinobacillus pleuropneumoniae. *FEMS Microbiol Lett* 2009; **298**: 124-30.
134. Bosse JT, Nash JH, Kroll JS *et al.* Harnessing natural transformation in Actinobacillus pleuropneumoniae: a simple method for allelic replacements. *FEMS Microbiol Lett* 2004; **233**: 277-81.
135. Sakai DK. Transformation of a nonpathogenic *Aeromonas salmonicida* mutant with intraspecific, interspecific and intergeneric bacterial DNA fragments in a laboratory model of river sediments. *Nippon Suisan Gakkaishi* 1987; **53**: 1141-9.
136. Huddleston JR, Brokaw JM, Zak JC *et al.* Natural transformation as a mechanism of horizontal gene transfer among environmental Aeromonas species. *Syst Appl Microbiol* 2013; **36**: 224– 34.
137. Essich E, Stevens SE, Jr., Porter RD. Chromosomal transformation in the cyanobacterium Agmenellum quadruplicatum. *J Bacteriol* 1990; **172**: 1916-22.
138. Stevens SE, Jr., Porter RD. Heterospecific transformation among cyanobacteria. *J Bacteriol* 1986; **167**: 1074-6.
139. Coukoulis H, Campbell LL. Transformation in Bacillus amyloliquefaciens. *J Bacteriol* 1971; **105**: 319-22.
140. Zawadzki P, Roberts MS, Cohan FM. The log-linear relationship between sexual isolation and sequence divergence in Bacillus transformation is robust. *Genetics* 1995; **140**: 917–32.

141. Roberts MS, Nakamura LK, Cohan FM. Bacillus mojavensis sp. nov., distinguishable from Bacillus subtilis by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *Int J Syst Bacteriol* 1994; **44**: 256-64.
142. Mareckova H. Transformation in Rhizobium japonicum. *Arch Mikrobiol* 1969; **68**: 113-5.
143. Thongdee M, Gallagher LA, Schell M *et al.* Targeted mutagenesis of Burkholderia thailandensis and Burkholderia pseudomallei through natural transformation of PCR fragments. *Appl Environ Microbiol* 2008; **74**: 2985-9.
144. Kang Y, Norris MH, Wilcox BA *et al.* Knockout and pullout recombineering for naturally transformable Burkholderia thailandensis and Burkholderia pseudomallei. *Nat Protoc* 2011; **6**: 1085-104.
145. Wang Y, Taylor DE. Natural transformation in Campylobacter species. *J Bacteriol* 1990; **172**: 949-55.
146. Frigaard NU, Bryant DA. Chromosomal gene inactivation in the green sulfur bacterium Chlorobium tepidum by natural transformation. *Appl Environ Microbiol* 2001; **67**: 2538-44.
147. Kennan RM, Dhungyel OP, Whittington RJ *et al.* The type IV fimbrial subunit gene (finA) of Dichelobacter nodosus is essential for virulence, protease secretion, and natural competence. *J Bacteriol* 2001; **183**: 4451–8.
148. Sinha S, Redfield RJ. Natural DNA uptake by Escherichia coli. *PLoS One* 2012; **7**: e35620.
149. Salomonsson E, Forsberg A, Roos N *et al.* Functional analyses of pilin-like proteins from Francisella tularensis: complementation of type IV pilus phenotypes in Neisseria gonorrhoeae. *Microbiology* 2009; **155**: 2546-59.
150. Kristensen BM, Sinha S, Boyce JD *et al.* Natural transformation of Gallibacterium anatis. *Appl Environ Microbiol* 2012; **78**: 4914-22.
151. Herriott RM, Meyer EM, Vogt M. Defined nongrowth media for stage II development of competence in Haemophilus influenzae. *J Bacteriol* 1970; **101**: 517-24.
152. Redfield RJ, Cameron AD, Qian Q *et al.* A novel CRP-dependent regulon controls expression of competence genes in Haemophilus influenzae. *J Mol Biol* 2005; **347**: 735-47.
153. Kleerebezem M, Boekhorst J, van Kranenburg R *et al.* Complete genome sequence of Lactobacillus plantarum WCFS1. *Proc Natl Acad Sci U S A* 2003; **100**: 1990-5.
154. Schmid S, Bevilacqua C, Crutz-Le Coq AM. Alternative sigma factor sigmaH activates competence gene expression in Lactobacillus sakei. *BMC Microbiol* 2012; **12**: 32.
155. Stone BJ, Kwaik YA. Natural competence for DNA transformation by Legionella pneumophila and its association with expression of type IV pili. *J Bacteriol* 1999; **181**: 1395–402.
156. Helmark S, Hansen ME, Jelle B *et al.* Transformation of Leuconostoc carnosum 4010 and evidence for natural competence of the organism. *Appl Environ Microbiol* 2004; **70**: 3695-9.
157. Meier PS, Troller R, Heiniger N *et al.* Unveiling electrotransformation of Moraxella catarrhalis as a process of natural transformation. *FEMS Microbiol Lett* 2006; **262**: 72-6.
158. Juni E. Simple genetic transformation assay for rapid diagnosis of Moraxella osloensis. *Appl Microbiol* 1974; **27**: 16-24.
159. Tribble GD, Rigney TW, Dao DH *et al.* Natural competence is a major mechanism for horizontal DNA transfer in the oral pathogen Porphyromonas gingivalis. *mBio* 2012; **3**.
160. Frischer ME, Williams HG, Bennison B *et al.* The naturally transformable marine bacterium WJT-1C formally identified as 'Vibrio' is pseudomonad. *Curr Microbiol* 1996; **33**: 287–91.
161. Carlson CA, Pierson LS, Rosen JJ *et al.* Pseudomonas stutzeri and related species undergo natural transformation. *J Bacteriol* 1983; **153**: 93-9.
162. Frischer ME, Thurmond JM, Paul JH. Natural plasmid transformation in a high-frequency-of-transformation marine Vibrio strain. *Appl Environ Microbiol* 1990; **56**: 3439-44.
163. Bertolla F, Van Gijsegem F, Nesme X *et al.* Conditions for natural transformation of Ralstonia solanacearum. *Appl Environ Microbiol* 1997; **63**: 4965-8.
164. Martin B, Quentin Y, Fichant G *et al.* Independent evolution of competence regulatory cascades in streptococci? *Trends Microbiol* 2006; **14**: 339-45.

165. Havarstein LS, Hakenbeck R, Gaustad P. Natural competence in the genus Streptococcus: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J Bacteriol* 1997; **179**: 6589-94.
166. Fontaine L, Boutry C, de Frahan MH *et al.* A novel pheromone quorum-sensing system controls the development of natural competence in Streptococcus thermophilus and Streptococcus salivarius. *J Bacteriol* 2010; **192**: 1444-54.
167. Asanuma N, Yoshii T, Kanada K *et al.* Involvement of two-component signal transduction system, ComDE, in the regulation of growth and genetic transformation, in the ruminal bacterium Streptococcus bovis. *Anaerobe* 2010; **16**: 405-11.
168. Mercer DK, Melville CM, Scott KP *et al.* Natural genetic transformation in the rumen bacterium Streptococcus bovis JB1. *FEMS Microbiol Lett* 1999; **179**: 485-90.
169. Abdulamir A, Hafidh R, Bakar F. The association of Streptococcus bovis/gallolyticus with colorectal tumors: The nature and the underlying mechanisms of its etiological role. *J Exp Clin Cancer Res* 2011; **30**: 11.
170. Correia FF, McKay TL, Farrow MF *et al.* Natural transformation of Streptococcus crista. *FEMS Microbiol Lett* 1996; **143**: 13-8.
171. Mercer DK, Scott KP, Bruce-Johnson WA *et al.* Fate of free DNA and transformation of the oral bacterium Streptococcus gordonii DL1 by plasmid DNA in human saliva. *Appl Environ Microbiol* 1999; **65**: 6-10.
172. Heng NCK, Tagg JR, Tompkins GR. Competence-Dependent Bacteriocin Production by Streptococcus gordonii DL1 (Challis). *J Bacteriol* 2007; **189**: 1468-72.
173. Morrison DA, Guedon E, Renault P. Competence for natural genetic transformation in the Streptococcus bovis group streptococci S. infantarius and S. macedonicus. *J Bacteriol* 2013; **195**: 2612-20.
174. Schlegel L, Grimont F, Collins MD *et al.* Streptococcus infantarius sp. nov., Streptococcus infantarius subsp. infantarius subsp. nov. and Streptococcus infantarius subsp. coli subsp. nov., isolated from humans and food. *Int J Syst Evol Microbiol* 2000; **50 Pt 4**: 1425-34.
175. Ween O, Teigen S, Gaustad P *et al.* Competence without a competence pheromone in a natural isolate of Streptococcus infantis. *J Bacteriol* 2002; **184**: 3426–32.
176. Papadimitriou K, Ferreira S, Papandreou NC *et al.* Complete Genome Sequence of the Dairy Isolate Streptococcus macedonicus ACA-DC 198. *J Bacteriol* 2012; **194**: 1838-9.
177. Gardan R, Besset C, Guillot A *et al.* The Oligopeptide Transport System Is Essential for the Development of Natural Competence in Streptococcus thermophilus Strain LMD-9. *J Bacteriol* 2009; **191**: 4647-55.
178. Guo Q, Ahn S-J, Kaspar J *et al.* Growth Phase and pH Influence Peptide Signaling for Competence Development in Streptococcus mutans. *J Bacteriol* 2014; **196**: 227-36.
179. Mashburn-Warren L, Morrison DA, Federle MJ. A novel double-tryptophan peptide pheromone controls competence in Streptococcus spp. via an Rgg regulator. *Mol Microbiol* 2010; **78**: 589-606.
180. Ronda C, Garcia JL, Lopez R. Characterization of genetic transformation in Streptococcus oralis NCTC 11427: expression of the pneumococcal amidase in S. oralis using a new shuttle vector. *Mol Gen Genet* 1988; **215**: 53-7.
181. Prudhomme M, Attaiech L, Sanchez G *et al.* Antibiotic stress induces genetic transformability in the human pathogen Streptococcus pneumoniae. *Science* 2006; **313**: 89-92.
182. Marks LR, Mashburn-Warren L, Federle MJ *et al.* Streptococcus pyogenes biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange. *J Infect Dis* 2014; **210**: 25-34.
183. Mashburn-Warren L, Morrison DA, Federle MJ. The cryptic competence pathway in Streptococcus pyogenes is controlled by a peptide pheromone. *J Bacteriol* 2012; **194**: 4589-600.
184. Fontaine L, Boutry C, de Frahan MH *et al.* A Novel Pheromone Quorum-Sensing System Controls the Development of Natural Competence in Streptococcus thermophilus and Streptococcus salivarius. *J Bacteriol* 2010; **192**: 1444-54.

185. Gaustad P. Genetic transformation in Streptococcus sanguis. Distribution of competence and competence factors in a collection of strains. *Acta Pathol Microbiol Scand B* 1979; **87B**: 123-8.
186. Gaustad P, Eriksen J, Henriksen SD. Genetic transformation in Streptococcus sanguis. Spontaneous and induced competence of selected strains. *Acta Pathol Microbiol Scand B* 1979; **87B**: 117-22.
187. Khan R, Rukke HV, Ricomini Filho AP *et al.* Extracellular Identification of a Processed Type II ComR/ComS Pheromone of Streptococcus mutans. *J Bacteriol* 2012; **194**: 3781-8.
188. Fontaine L, Goffin P, Dubout H *et al.* Mechanism of competence activation by the ComRS signalling system in streptococci. *Mol Microbiol* 2013; **87**: 1113-32.
189. Blomqvist T, Steinmoen H, Havarstein LS. Natural genetic transformation: A novel tool for efficient genetic engineering of the dairy bacterium Streptococcus thermophilus. *Appl Environ Microbiol* 2006; **72**: 6751-6.
190. Shestakov SV, Khyen NT. Evidence for genetic transformation in blue-green alga Anacystis nidulans. *Mol Genet Genomics* 1979; **107**: 372-5.
191. Tsinoremas NF, Kutach AK, Strayer CA *et al.* Efficient gene transfer in Synechococcus sp. strains PCC 7942 and PCC 6301 by interspecies conjugation and chromosomal recombination. *J Bacteriol* 1994; **176**: 6764-8.
192. Nishikawa K, Tanaka Y. A simple mutagenesis using natural competence in Tannerella forsythia. *J Microbiol Methods* 2013; **94**: 378-80.
193. Shaw AJ, Hogsett DA, Lynd LR. Natural competence in Thermoanaerobacter and Thermoanaerobacterium species. *Appl Environ Microbiol* 2010; **76**: 4713-9.
194. Onai K, Morishita M, Kaneko T *et al.* Natural transformation of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1: a simple and efficient method for gene transfer. *Mol Genet Genomics* 2004; **271**: 50-9.
195. Han D, Xu H, Puranik R *et al.* Natural transformation of Thermotoga sp. strain RQ7. *BMC Biotechnol* 2014; **14**: 39.
196. Pollack-Berti A, Wollenberg MS, Ruby EG. Natural transformation of Vibrio fischeri requires tfoX and tfoY. *Environ Microbiol* 2010; **12**: 2302-11.
197. Jeffrey WH, Paul JH, Stewart GJ. Natural transformation of a marineVibrio species by plasmid DNA. *Microb Ecol* 1990; **19**: 259-68.
198. Gulig PA, Tucker MS, Thiaville PC *et al.* USER friendly cloning coupled with chitin-based natural transformation enables rapid mutagenesis of Vibrio vulnificus. *Appl Environ Microbiol* 2009; **75**: 4936-49.
199. Kung SH, Almeida RP. Natural competence and recombination in the plant pathogen Xylella fastidiosa. *Appl Environ Microbiol* 2011; **77**: 5278-84.
200. Zhou J, Bowler LD, Spratt BG. Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of Neisseria meningitidis and commensal Neisseria species. *Mol Microbiol* 1997; **23**: 799-812.
201. Desjardins P, Picard B, Kaltenbock B *et al.* Sex in Escherichia coli does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J Mol Evol* 1995; **41**: 440-8.
202. Maslow JN, Whittam TS, Gilks CF *et al.* Clonal relationships among bloodstream isolates of Escherichia coli. *Infect Immun* 1995; **63**: 2409-17.
203. Michod RE, Bernstein H, Nedelcu AM. Adaptive value of sex in microbial pathogens. *Infect Genet Evol* 2008; **8**: 267-85.
204. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J* 2009; **3**: 199-208.
205. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science* 2007; **315**: 476-80.
206. Acinas SG, Klepac-Ceraj V, Hunt DE *et al.* Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 2004; **430**: 551-4.
207. Hanage WP, Fraser C, Spratt BG. Fuzzy species among recombinogenic bacteria. *BMC Biol* 2005; **3**: 6.

208. Venter JC, Remington K, Heidelberg JF *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; **304**: 66-74.
209. Hanage WP, Spratt BG, Turner KM *et al.* Modelling bacterial speciation. *Philos Trans R Soc Lond B Biol Sci* 2006; **361**: 2039-44.
210. Lawrence JG. Gene transfer in bacteria: speciation without species? *Theor Popul Biol* 2002; **61**: 449-60.
211. Vulic M, Dionisio F, Taddei F *et al.* Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc Natl Acad Sci U S A* 1997; **94**: 9763-7.
212. Majewski J, Zawadzki P, Pickerill P *et al.* Barriers to genetic exchange between bacterial species: Streptococcus pneumoniae transformation. *J Bacteriol* 2000; **182**: 1016-23.
213. de Vries J, Wackernagel W. Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. *PNAS* 2002; **99**: 2094-9.
214. Meier P, Wackernagel W. Mechanisms of homology-facilitated illegitimate recombination for foreign DNA acquisition in transformable Pseudomonas stutzeri. *Mol Microbiol* 2003; **48**: 1107-18.
215. van Elsas JD, Turner S, Bailey MJ. Horizontal gene transfer in the phytosphere. *New Phytol* 2003; **157**: 525-37

216. Kharazmi M, Bauer T, Hammes WP *et al.* Effect of food processing on the fate of DNA with regard to degradation and transformation capability in Bacillus subtilis. *Syst Appl Microbiol* 2003; **26**: 495-501.
217. Rizzi A, Raddadi N, Sorlini C *et al.* The stability and degradation of dietary DNA in the gastrointestinal tract of mammals: Implications for horizontal gene transfer and the biosafety of GMOs. *Crit Rev Food Sci Nutr* 2012; **52**: 142-61.
218. Stecher B, Denzler R, Maier L *et al.* Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proc Natl Acad Sci U S A* 2012; **109**: 1269-74.
219. Capozzi V, Spano G. Horizontal gene transfer in the gut: Is it a risk? *Food Res Int* 2009; **42**: 1501-2.
220. Van Reenen CA, Dicks LMT. Horizontal gene transfer amongst probiotic lactic acid bacteria and other intestinal microbiota: What are the possibilities? A review. *Arch Microbiol* 2011; **193**: 157-68.
221. Nielsen KM, Berdal KG, Kruse H *et al.* An assessment on potential long-term health effects caused by antibiotic resistance marker genes in genetically modified organisms based on antibiotic usage and resistance patterns in Norway. *VKM-Report* 2005: 1-62.
222. Bauer F, Hertel C, Hammes WP. Transformation of Escherichia coli in foodstuffs. *Syst Appl Microbiol* 1999; **22**: 161-8.
223. Straub JA, Hertel C, Hammes WP. The fate of recombinant DNA in thermally treated fermented sausages. *Eur Food Res Technol* 1999; **210**: 62-7.
224. Alexander TW, Sharma R, Okine EK *et al.* Impact of feed processing and mixed ruminal culture on the fate of recombinant EPSP synthase and endogenous canola plant DNA. *FEMS Microbiol Lett* 2002; **214**: 263-9.
225. Jonas DA, Elmadfa I, Engel KH *et al.* Safety considerations of DNA in food. *Ann Nutr Metab* 2001; **45**: 235-54.
226. Ottow JCG. *Mikrobiologie von Böden. Biodiversität, Ökophysiologie und Metagenomik.*: Springer-Verlag Berlin-Heidelberg, 2011.
227. Nielsen KM, van Elsas JD. Stimulatory effects of compounds present in the rhizosphere on natural transformation of Acinetobacter sp BD413 in soil. *Soil Biology & Biochemistry* 2001; **33**: 345-57.
228. Davison J. Genetic exchange between bacteria in the environment. *Plasmid* 1999; **42**: 73-91.
229. Nielsen KM, Van Elsas JD, Smalla K. Dynamics, horizontal transfer and selection of novel DNA in bacterial populations in the phytosphere of transgenic plants. *Annals of Microbiology* 2001; **51**: 79-94.

230. Pontiroli A, Rizzi A, Simonet P *et al.* Visual evidence of horizontal gene transfer between plants and bacteria in the phytosphere of transplastomic tobacco. *Appl Environ Microbiol* 2009; **75**: 3314-22.
231. Kay E, Bertolla F, Vogel TM *et al.* Opportunistic colonization of Ralstonia solanacearum-infected plants by Acinetobacter sp. and its natural competence development. *Microb Ecol* 2002; **43**: 291-7.
232. Brautigam M, Hertel C, Hammes WP. Evidence for natural transformation of Bacillus subtilis in foodstuffs. *FEMS Microbiol Lett* 1997; **155**: 93-8.
233. Bowler LD, Zhang QY, Riou JY *et al.* Interspecies recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicillin resistance in N. meningitidis: natural events and laboratory simulation. *J Bacteriol* 1994; **176**: 333-7.
234. Dowson CG, Coffey TJ, Spratt BG. Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to beta-lactam antibiotics. *Trends Microbiol* 1994; **2**: 361-66.
235. Patterson AJ, Rincon MT, Flint HJ *et al.* Mosaic tetracycline resistance genes are widespread in human and animal fecal samples. *Antimicrob Agents Chemother* 2007; **51**: 1115-8.
236. Roberts AP, Chandler M, Courvalin P *et al.* Revised nomenclature for transposable genetic elements. *Plasmid* 2008; **60**: 167-73.
237. van Hoek AH, Mayrhofer S, Domig KJ *et al.* Mosaic tetracycline resistance genes and their flanking regions in Bifidobacterium thermophilum and Lactobacillus johnsonii. *Antimicrob Agents Chemother* 2008; **52**: 248-52.
238. Zhaxybayeva O, Gogarten JP, Charlebois RL *et al.* Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res* 2006; **16**: 1099-108.
239. Matic I, Taddei F, Radman M. Interspecies recombination and mismatch repair. Generation of mosaic genes and genomes. *Methods Mol Biol* 2000; **152**: 149-57.
240. Denamur E, Lecointre G, Darlu P *et al.* Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 2000; **103**: 711-21.
241. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature* 1989; **342**: 396-401.
242. Alloing G, Martin B, Granadel C *et al.* Development of competence in Streptococcus pneumonaie: pheromone autoinduction and control of quorum sensing by the oligopeptide permease. *Mol Microbiol* 1998; **29**: 75-83.
243. Martin DP, Lemey P, Lott M *et al.* RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010; **26**: 2462-3.
244. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 2000; **16**: 562-3.
245. Martin DP, Posada D, Crandall KA *et al.* A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retroviruses* 2005; **21**: 98-102.
246. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci U S A* 2001; **98**: 13757-62.
247. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology* 1999; **265**: 218-25.
248. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 2000; **16**: 573-82.
249. Holmes EC, Worobey M, Rambaut A. Phylogenetic evidence for recombination in dengue virus. *Mol Biol Evol* 1999; **16**: 405-9.
250. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 2007; **176**: 1035-47.
251. Shakya T, Stogios PJ, Waglechner N *et al.* A small molecule discrimination map of the antibiotic resistance kinome. *Chem Biol* 2011; **18**: 1591-601.
252. Miki B, McHugh S. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J Biotechnol* 2004; **107**: 193-232.

253. Chen J, Jin M, Qiu ZG *et al.* A Survey of Drug Resistance bla Genes Originating from Synthetic Plasmid Vectors in Six Chinese Rivers. *Environ Sci Technol* 2012; **46**: 13448-54.
254. Shaw KJ, Rather PN, Hare RS *et al.* Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* 1993; **57**: 138-63.
255. WHO. Critically important antimicrobials for human medicine. 3rd revision 2011. *WHO Press, World Health Organization, Geneva http://wwwwhoint/foodborne_disease/resistance/cia/en/ (7 Oct 2013, date last accessed)* 2012.
256. Perlin MH, Lerner SA. High-level amikacin resistance in *Escherichia coli* due to phosphorylation and impaired aminoglycoside uptake. *Antimicrob Agents Chemother* 1986; **29**: 216-24.
257. Hall T. BioEdit version 7.0.0. Manual. http://www.mbio.ncsu.edu/bioedit/biodoc.pdf; last accessed: August 14th, 2014. 2007.
258. Altamia MA, Wood N, Fung JM *et al.* Genetic differentiation among isolates of Teredinibacter turnerae, a widely occurring intracellular endosymbiont of shipworms. *Mol Ecol* 2014; **23**: 1418-32.
259. Duron O. Arsenophonus insect symbionts are commonly infected with APSE, a bacteriophage involved in protective symbiosis. *FEMS Microbiol Ecol* 2014.
260. Freel KC, Millan-Aguinaga N, Jensen PR. Multilocus sequence typing reveals evidence of homologous recombination linked to antibiotic resistance in the genus Salinispora. *Appl Environ Microbiol* 2013; **79**: 5997-6005.
261. Hester SE, Park J, Goodfield LL *et al.* Horizontally acquired divergent O-antigen contributes to escape from cross-immunity in the classical bordetellae. *BMC Evol Biol* 2013; **13**: 209.
262. Thomas JC, Godfrey PA, Feldgarden M *et al.* Draft genome sequences of Staphylococcus aureus sequence type 34 (ST34) and ST42 hybrids. *J Bacteriol* 2012; **194**: 2740-1.
263. Alvarez-Perez S, de Vega C, Herrera CM. Multilocus sequence analysis of nectar pseudomonads reveals high genetic diversity and contrasting recombination patterns. *PLoS One* 2013; **8**: e75797.
264. Keymer DP, Boehm AB. Recombination shapes the structure of an environmental Vibrio cholerae population. *Appl Environ Microbiol* 2011; **77**: 537-44.
265. Smith SE, Showers-Corneli P, Dardenne CN *et al.* Comparative genomic and phylogenetic approaches to characterize the role of genetic recombination in mycobacterial evolution. *PLoS One* 2012; **7**: e50070.
266. Kosakovsky Pond SL, Posada D, Gravenor MB *et al.* Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Mol Biol Evol* 2006; **23**: 1891-901.
267. Yang FF, Sun YN, Li JX *et al.* Detection of aminoglycoside resistance genes in Riemerella anatipestifer isolated from ducks. *Vet Microbiol* 2012; **158**: 451-2.
268. Zhu B. Abundance dynamics and sequence variation of neomycin phosphotransferase gene (nptII) homologs in river water. *Aquat Microb Ecol* 2007; **48**: 131-40.
269. Hakenbeck R. Mosaic genes and their role in penicillin-resistant Streptococcus pneumoniae (minireview). *Electrophoresis* 1998; **19**: 597-601.
270. Oggioni MR, Dowson CG, Smith JM *et al.* The tetracycline resistance gene tet(M) exhibits mosaic structure. *Plasmid* 1996; **35**: 156-63.
271. Hanage WP, Fraser C, Spratt BG. The impact of homologous recombination on the generation of diversity in bacteria. *J Theor Biol* 2006; **239**: 210-9.
272. Barile S, Devirgiliis C, Perozzi G. Molecular characterization of a novel mosaic tet(S/M) gene encoding tetracycline resistance in foodborne strains of Streptococcus bovis. *Microbiology* 2012; **158**: 2353-62.
273. Stanton TB, Humphrey SB. Isolation of tetracycline-resistant Megasphaera elsdenii strains with novel mosaic gene combinations of tet(O) and tet(W) from swine. *Appl Environ Microbiol* 2003; **69**: 3874-82.
274. Hakenbeck R. Transformation in Streptococcus pneumoniae: mosaic genes and the regulation of competence. *Res Microbiol* 2000; **151**: 453-6.

275. Johansen BK, Wasteson Y, Granum PE *et al.* Mosaic structure of Shiga-toxin-2-encoding phages isolated from Escherichia coli O157 : H7 indicates frequent gene exchange between lambdoid phage genomes. *Microbiology-Sgm* 2001; **147**: 1929-36.
276. Nakamura K, Kohda T, Shibata Y *et al.* Unique Biological Activity of Botulinum D/C Mosaic Neurotoxin in Murine Species. *Infect Immun* 2012; **80**: 2886-93.
277. Bay RA, Bielawski JP. Recombination detection under evolutionary scenarios relevant to functional divergence. *J Mol Evol* 2011; **73**: 273-86.
278. della-Cioppa G, Bauer SC, Taylor ML *et al.* Targeting a herbicide–resistant enzyme from *Escherichia coli* to chloroplasts of higher plants. *Nat Biotechnol* 1987; **5**: 579 - 84.
279. Wang J, Zuo K, Lu W *et al.* Over-expression of the Gr5aroA gene from glyphosate-contaminated soil confers high tolerance to glyphosate in tobacco. *Mol Breed* 2014; **33**: 197-208.
280. Schloss PD, Westcott SL, Ryabin T *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; **75**: 7537-41.
281. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 1995; **59**: 143-69.
282. Lorenz MG, Wackernagel W. High Frequency of Natural Genetic Transformation of Pseudomonas stutzeri in Soil Extract Supplemented with a Carbon/Energy and Phosphorus Source. *Appl Environ Microbiol* 1991; **57**: 1246-51.
283. Feil E, Zhou J, Maynard Smith J *et al.* A comparison of the nucleotide sequences of the adk and recA genes of pathogenic and commensal Neisseria species: evidence for extensive interspecies recombination within adk. *J Mol Evol* 1996; **43**: 631-40.
284. Wang F, Zhang Y, Liu J *et al.* [Comparison of phylogeny analysis methods for rhizobia asolated from Albizia spp., Acacia spp. and Leucaena leucocephala]. *Acta Microbiol Sin* 2008; **48**: 1-7.
285. Hakenbeck R, Balmelle N, Weber B *et al.* Mosaic genes and mosaic chromosomes: Intra- and interspecies genomic variation of Streptococcus pneumoniae. *Infect Immun* 2001; **69**: 2477-86.
286. Moura H, Terilli RR, Woolfitt AR *et al.* Studies on botulinum neurotoxins type /C1 and mosaic/DC using Endopep-MS and proteomics. *FEMS Immunol Med Microbiol* 2011; **61**: 288-300.
287. Nakamura K, Kohda T, Seto Y *et al.* Improved detection methods by genetic and immunological techniques for botulinum C/D and D/C mosaic neurotoxins. *Vet Microbiol* 2013; **162**: 881-90.
288. Tsukamoto K, Arimitsu H, Ochi S *et al.* P19 embryonal carcinoma cells exhibit high sensitivity to botulinum type C and D/C mosaic neurotoxins. *Microbiol Immunol* 2012; **56**: 664-72.
289. Vidal D, Anza I, Taggart MA *et al.* Environmental factors influencing the prevalence of a Clostridium botulinum type C/D mosaic strain in nonpermanent Mediterranean wetlands. *Appl Environ Microbiol* 2013; **79**: 4264-71.
290. Woudstra C, Skarin H, Anniballi F *et al.* Validation of a real-time PCR based method for detection of Clostridium botulinum types C, D and their mosaic variants C-D and D-C in a multicenter collaborative trial. *Anaerobe* 2013; **22**: 31-7.
291. Zhang Y, Buchko GW, Qin L *et al.* Crystal structure of the receptor binding domain of the botulinum C-D mosaic neurotoxin reveals potential roles of lysines 1118 and 1136 in membrane interactions. *Biochem Biophys Res Commun* 2011; **404**: 407-12.
292. Moriishi K, Koura M, Abe N *et al.* Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms. *Biochim Biophys Acta* 1996; **1307**: 123-6.
293. Monstein HJ, Ryberg A, Karlsson A. Assessment of the mosaic structure in the Helicobacter pylori cagA gene 3'-region using an improved polymerase chain reaction-based assay. *Diagn Microbiol Infect Dis* 2012; **73**: 281-3.
294. Osteras M, Stanley J, Finan TM. Identification of Rhizobium-specific intergenic mosaic elements within an essential two-component regulatory system of Rhizobium species. *J Bacteriol* 1995; **177**: 5485-94.
295. Semedo T, Almeida Santos M, Martins P *et al.* Comparative Study Using Type Strains and Clinical and Food Isolates To Examine Hemolytic Activity and Occurrence of the cyl Operon in Enterococci. *J Clin Microbiol* 2003; **41**: 2569-76.

296. Bentley SD, Vernikos GS, Snyder LAS *et al.* Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18. *PLoS Genet* 2007; **3**: e23.
297. Radstrom P, Fermer C, Kristiansen BE *et al.* Transformational exchanges in the dihydropteroate synthase gene of Neisseria meningitidis: a novel mechanism for acquisition of sulfonamide resistance. *J Bacteriol* 1992; **174**: 6386-93.
298. Costa J, Tiago I, Da Costa MS *et al.* Molecular evolution of Legionella pneumophila dotA gene, the contribution of natural environmental strains. *Environ Microbiol* 2010; **12**: 2711-29.
299. Lacks SA, Mannarelli BM, Springhorn SS *et al.* Genetic basis of the complementary DpnI and DpnII restriction systems of S. pneumoniae: an intercellular cassette mechanism. *Cell* 1986; **46**: 993-1000.
300. Whatmore AM, Kapur V, Musser JM *et al.* Molecular population genetic analysis of the enn subdivision of group A streptococcal emm-like genes: horizontal gene transfer and restricted variation among enn genes. *Mol Microbiol* 1995; **15**: 1039-48.
301. Whatmore AM, Kehoe MA. Horizontal gene transfer in the evolution of group A streptococcal emm-like genes: gene mosaics and variation in Vir regulons. *Mol Microbiol* 1994; **11**: 363-74.
302. Kulick S, Moccia C, Didelot X *et al.* Mosaic DNA imports with interspersions of recipient sequence after natural transformation of Helicobacter pylori. *PLoS One* 2008; **3**: e3797.
303. Nelson K, Selander RK. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (gnd) in enteric bacteria. *Proc Natl Acad Sci U S A* 1994; **91**: 10227-31.
304. Stoltzfus A, Leslie JF, Milkman R. Molecular evolution of the Escherichia coli chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between trp and tonB. *Genetics* 1988; **120**: 345-58.
305. Davies RL, Campbell S, Whittam TS. Mosaic Structure and Molecular Evolution of the Leukotoxin Operon (lktCABD) in Mannheimia (Pasteurella) haemolytica, Mannheimia glucosida, and Pasteurella trehalosi. *J Bacteriol* 2002; **184**: 266-77.
306. Parker CT, Gilbert M, Yuki N *et al.* Characterization of Lipooligosaccharide-Biosynthetic Loci of Campylobacter jejuni Reveals New Lipooligosaccharide Classes: Evidence of Mosaic Organizations▽†. *J Bacteriol* 2008; **190**: 5681-9.
307. Bruckner R, Nuhn M, Reichmann P *et al.* Mosaic genes and mosaic chromosomes-genomic variation in Streptococcus pneumoniae. *Int J Med Microbiol* 2004; **294**: 157-68.
308. Filipe SR, Severina E, Tomasz A. Distribution of the mosaic structured murM genes among natural populations of Streptococcus pneumoniae. *J Bacteriol* 2000; **182**: 6798-805.
309. Gutierrez MC, Brisse S, Brosch R *et al.* Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Path* 2005; **1**: e5.
310. King SJ, Whatmore AM, Dowson CG. NanA, a neuraminidase from Streptococcus pneumoniae, shows high levels of sequence diversity, at least in part through recombination with Streptococcus oralis. *J Bacteriol* 2005; **187**: 5376-86.
311. Gut H, King SJ, Walsh MA. Structural and functional studies of Streptococcus pneumoniae neuraminidase B: An intramolecular trans-sialidase. *FEBS Lett* 2008; **582**: 3348-52.
312. Pettigrew MM, Fennie KP, York MP *et al.* Variation in the Presence of Neuraminidase Genes among Streptococcus pneumoniae Isolates with Identical Sequence Types. *Infect Immun* 2006; **74**: 3360-5.
313. Collins MD, East AK. Phylogeny and taxonomy of the food-borne pathogen Clostridium botulinum and its neurotoxins. *J Appl Microbiol* 1998; **84**: 5-17.
314. Snyder LAS, McGowan S, Rogers M *et al.* The Repertoire of Minimal Mobile Elements in the Neisseria Species and Evidence That These Are Involved in Horizontal Gene Transfer in Other Bacteria. *Mol Biol Evol* 2007; **24**: 2802-15.
315. Davies RL, Lee I. Sequence Diversity and Molecular Evolution of the Heat-Modifiable Outer Membrane Protein Gene (ompA) of Mannheimia(Pasteurella) haemolytica, Mannheimia glucosida, and Pasteurella trehalosi. *J Bacteriol* 2004; **186**: 5741-52.
316. Tang C, Zhang B, Yue H *et al.* Characteristics of the molecular diversity of the outer membrane protein A gene of Haemophilus parasuis. *Can J Vet Res* 2010; **74**: 233-6.

317. Omelchenko MV, Makarova KS, Wolf YI *et al.* Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol* 2003; **4**: R55.
318. Hao W. Unrecognized fine-scale recombination can mimic the effects of adaptive radiation. *Gene* 2013; **518**: 483-8.
319. Gonzalez I, Georgiou M, Alcaide F *et al.* Fluoroquinolone resistance mutations in the parC, parE, and gyrA genes of clinical isolates of viridans group streptococci. *Antimicrob Agents Chemother* 1998; **42**: 2792-8.
320. Du Plessis M, Smith AM, Klugman KP. Analysis of penicillin-binding protein lb and 2a genes from Streptococcus pneumoniae. *Microb Drug Resist* 2000; **6**: 127-31.
321. Ito M, Deguchi T, Mizutani KS *et al.* Emergence and spread of Neisseria gonorrhoeae clinical isolates harboring mosaic-like structure of penicillin-binding protein 2 in Central Japan. *Antimicrob Agents Chemother* 2005; **49**: 137-43.
322. Pandori M, Barry PM, Wu A *et al.* Mosaic penicillin-binding protein 2 in Neisseria gonorrhoeae isolates collected in 2008 in San Francisco, California. *Antimicrob Agents Chemother* 2009; **53**: 4032-4.
323. Takahata S, Senju N, Osaki Y *et al.* Amino acid substitutions in mosaic penicillin-binding protein 2 associated with reduced susceptibility to cefixime in clinical isolates of Neisseria gonorrhoeae. *Antimicrob Agents Chemother* 2006; **50**: 3638-45.
324. Tomberg J, Unemo M, Davies C *et al.* Molecular and Structural Analysis of Mosaic Variants of Penicillin-Binding Protein 2 Conferring Decreased Susceptibility to Expanded-Spectrum Cephalosporins in Neisseria gonorrhoeae: Role of Epistatic Mutations. *Biochemistry (Mosc)* 2010; **49**: 8062-70.
325. Spratt BG, Zhang QY, Jones DM *et al.* Recruitment of a penicillin-binding protein gene from Neisseria flavescens during the emergence of penicillin resistance in Neisseria meningitidis. *Proc Natl Acad Sci U S A* 1989; **86**: 8988-92.
326. Dowson CG, Coffey TJ, Kell C *et al.* Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae. *Mol Microbiol* 1993; **9**: 635-43.
327. Zerfass I, Hakenbeck R, Denapaite D. An important site in PBP2x of penicillin-resistant clinical isolates of Streptococcus pneumoniae: mutational analysis of Thr338. *Antimicrob Agents Chemother* 2009; **53**: 1107-15.
328. Sibold C, Henrichsen J, Konig A *et al.* Mosaic pbpX genes of major clones of penicillin-resistant Streptococcus pneumoniae have evolved from pbpX genes of a penicillin-sensitive Streptococcus oralis. *Mol Microbiol* 1994; **12**: 1013-23.
329. Ochiai S, Ishiko H, Yasuda M *et al.* Rapid detection of the mosaic structure of the Neisseria gonorrhoeae penA gene, which is associated with decreased susceptibilities to oral cephalosporins. *J Clin Microbiol* 2008; **46**: 1804-10.
330. Tapsall JW, Ray S, Limnios A. Characteristics and population dynamics of mosaic penA allele-containing Neisseria gonorrhoeae isolates collected in Sydney, Australia, in 2007-2008. *Antimicrob Agents Chemother* 2010; **54**: 554-6.
331. Unemo M, Golparian D, Nicholas R *et al.* High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure. *Antimicrob Agents Chemother* 2012; **56**: 1273-80.
332. Whiley D, Bates J, Limnios A *et al.* Use of a novel screening PCR indicates presence of Neisseria gonorrhoeae isolates with a mosaic penA gene sequence in Australia. *Pathology* 2007; **39**: 445-6.
333. Ameyama S, Onodera S, Takahata M *et al.* Mosaic-like structure of penicillin-binding protein 2 Gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime. *Antimicrob Agents Chemother* 2002; **46**: 3744-9.
334. Chen CC, Yen MY, Wong WW *et al.* Tracing subsequent dissemination of a cluster of gonococcal infections caused by an ST1407-related clone harbouring mosaic penA alleles in Taiwan. *J Antimicrob Chemother* 2013.

335. Huang CT, Yen MY, Wong WW *et al.* Characteristics and dissemination of mosaic penicillin-binding protein 2-harboring multidrug-resistant Neisseria gonorrhoeae isolates with reduced cephalosporin susceptibility in northern Taiwan. *Antimicrob Agents Chemother* 2010; **54**: 4893-5.
336. Huang CT, Niu J, Liao MH *et al.* A duplex PCR method to identify mosaic penA gene and predict reduced susceptibility to oral cephalosporins in Neisseria gonorrhoeae. *J Microbiol Methods* 2010; **83**: 257-9.
337. DuBose RF, Dykhuizen DE, Hartl DL. Genetic exchange among natural isolates of bacteria: recombination within the phoA gene of Escherichia coli. *Proc Natl Acad Sci U S A* 1988; **85**: 7036-40.
338. Groisman EA, Saier MH, Jr., Ochman H. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the Salmonella genome. *EMBO J* 1992; **11**: 1309-16.
339. Mattick JS. Type IV pili and twitching motility. *Annu Rev Microbiol* 2002; **56**: 289-314.
340. Myers GS, Rasko DA, Cheung JK *et al.* Skewed genomic variability in strains of the toxigenic bacterial pathogen, Clostridium perfringens. *Genome Res* 2006; **16**: 1031-40.
341. Hill SA, Davies JK. Pilin gene variation in Neisseria gonorrhoeae: reassessing the old paradigms. *FEMS Microbiol Rev* 2009; **33**: 521-30.
342. Diep DB, Straume D, Kjos M *et al.* An overview of the mosaic bacteriocin pln loci from Lactobacillus plantarum. *Peptides* 2009; **30**: 1562-74.
343. Fudyk TC, Maclean IW, Simonsen JN *et al.* Genetic Diversity and Mosaicism at the por Locus ofNeisseria gonorrhoeae. *J Bacteriol* 1999; **181**: 5591-9.
344. Pandya GA, McEllistrem MC, Venepally P *et al.* Monitoring the long-term molecular epidemiology of the pneumococcus and detection of potential 'vaccine escape' strains. *PLoS One* 2011; **6**: e15950.
345. McDaniel LS, McDaniel DO, Hollingshead SK *et al.* Comparison of the PspA sequence from Streptococcus pneumoniae EF5668 to the previously identified PspA sequence from strain Rx1 and ability of PspA from EF5668 to elicit protection against pneumococci of different capsular types. *Infect Immun* 1998; **66**: 4748-54.
346. Crain MJ, Waltman WD, 2nd, Turner JS *et al.* Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of Streptococcus pneumoniae. *Infect Immun* 1990; **58**: 3293-9.
347. Lai LB, Cho IM, Chen WY *et al.* Archaeal RNase P: A Mosaic of Its Bacterial and Eukaryal Relatives. In: Liu F, Altman S, eds. *Ribonuclease P*, 2010; 153-72.
348. Towers RJ, Fagan PK, Talay SR *et al.* Evolution of sfbI encoding streptococcal fibronectin-binding protein I: Horizontal genetic transfer and gene mosaic structure. *J Clin Microbiol* 2003; **41**: 5398-406.
349. Kapur V, Kanjilal S, Hamrick MR *et al.* Molecular population genetic analysis of the streptokinase gene of Streptococcus pyogenes: mosaic alleles generated by recombination. *Mol Microbiol* 1995; **16**: 509-19.
350. Harrison O, Maiden M, Rokbi B. Distribution of transferrin binding protein B gene (tbpB) variants among Neisseria species. *BMC Microbiol* 2008; **8**: 66.
351. Rokbi B, Mignon M, Caugant DA *et al.* Heterogeneity of tbpB, the transferrin-binding protein B gene, among serogroup B Neisseria meningitidis strains of the ET-5 complex. *Clin Diagn Lab Immunol* 1997; **4**: 522-9.
352. Lee I, Davies RL. Evidence for a common gene pool and frequent recombinational exchange of the tbpBA operon in Mannheimia haemolytica, Mannheimia glucosida and Bibersteinia trehalosi. *Microbiology* 2011; **157**: 123-35.
353. Potron A, Nordmann P, Rondinaud E *et al.* A mosaic transposon encoding OXA-48 and CTX-M-15: towards pan-resistance. *J Antimicrob Chemother* 2013; **68**: 476-7.
354. Kholodii G, Gorlenko Z, Mindlin S *et al.* Tn5041-like transposons: molecular diversity, evolutionary relationships and distribution of distinct variants in environmental bacteria. *Microbiology* 2002; **148**: 3569-82.
355. Szuplewska M, Bartosik D. Identification of a mosaic transposable element of Paracoccus marcusii composed of insertion sequence ISPmar4 (ISAs1 family) and an IS1247a-driven transposable module (TMo). *FEMS Microbiol Lett* 2009; **292**: 216-21.

356. Owen RJ, Xerry J, Peters TM *et al.* Surveillance and clinical relevance of vacA genotypes of Helicobacter pylori infecting dyspeptic patients in mid-Essex. *Commun Dis Public Health* 2002; **5**: 106-11.
357. Yamazaki S, Yamakawa A, Okuda T *et al.* Distinct Diversity of vacA, cagA, and cagE Genes of Helicobacter pylori Associated with Peptic Ulcer in Japan. *J Clin Microbiol* 2005; **43**: 3906-16.
358. Dykhuizen DE, Baranton G. The implications of a low rate of horizontal transfer in Borrelia. *Trends Microbiol* 2001; **9**: 344-50.
359. Baldo L, Lo N, Werren JH. Mosaic nature of the Wolbachia surface protein. *J Bacteriol* 2005; **187**: 5406-18.
360. Nielsen KM, Johnsen PJ, Bensasson D *et al.* Release and persistence of extracellular DNA in the environment. *Environ Biosafety Res* 2007; **6**: 37-53.
361. Hamilton HL, Dillard JP. Natural transformation of Neisseria gonorrhoeae: from DNA donation to homologous recombination. *Mol Microbiol* 2006; **59**: 376-85.
362. Kroll JS, Wilks KE, Farrant JL *et al.* Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens. *Proc Natl Acad Sci U S A* 1998; **95**: 12381-5.
363. Feavers IM, Heath AB, Bygraves JA *et al.* Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of Neisseria meningitidis. *Mol Microbiol* 1992; **6**: 489-95.
364. Doern GV, Heilmann KP, Huynh HK *et al.* Antimicrobial resistance among clinical isolates of Streptococcus pneumoniae in the United States during 1999--2000, including a comparison of resistance rates since 1994--1995. *Antimicrob Agents Chemother* 2001; **45**: 1721-9.
365. Hakenbeck R, Grebe T, Zahner D *et al.* beta-lactam resistance in Streptococcus pneumoniae: penicillin-binding proteins and non-penicillin-binding proteins. *Mol Microbiol* 1999; **33**: 673-8.
366. Majewski J, Cohan FM. DNA sequence similarity requirements for interspecific recombination in Bacillus. *Genetics* 1999; **153**: 1525-33.
367. Majewski J. Sexual isolation in bacteria. *FEMS Microbiol Lett* 2001; **199**: 161-9.
368. Murin CD, Segal K, Bryksin A *et al.* Expression vectors for Acinetobacter baylyi ADP1. *Appl Environ Microbiol* 2012; **78**: 280-3.
369. Metzgar D, Bacher JM, Pezo V *et al.* Acinetobacter sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res* 2004; **32**: 5780-90.
370. Vaneechoutte M, Young DM, Ornston LN *et al.* Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. *Appl Environ Microbiol* 2006; **72**: 932-6.
371. Rizzi A, Pontiroli A, Brusetti L *et al.* Strategy for in situ detection of natural transformation-based horizontal gene transfer events. *Appl Environ Microbiol* 2008; **74**: 1250-4.
372. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 2005; **3**: 711-21.
373. de Vries J, Wackernagel W. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci U S A* 2002; **99**: 2094-9.
374. Palmen R, Hellingwerf KJ. Uptake and processing of DNA by *Acinetobacter calcoaceticus*--a review. *Gene* 1997; **192**: 179-90.
375. CLSI. Performance standards for antimicrobial susceptibility testing; Twenty-second informational supplement. *Wayne, PA: Clinical and Laboratory Standards Institute* 2012; **CLSI document M100-S22**.
376. NCCLS. Performance standards for antimicrobial susceptibility testing; fourteenth informational supplement. *NCCLS, Wayne, PA* 2004; **NCCLS document M100-S14**.
377. Domingues S, Harms K, Fricke WF *et al.* Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS Pathog* 2012; **8**: e1002837.
378. Nielsen KM, Ray JL, Johnsen PJ. Horizontal gene transfer: uptake of extracellular DNA by bacteria. In: Shaechter M, ed. *Encyclopedia of Microbiology*. Oxford: Elsevier, 2009; 587-96.

379. Hulter N, Wackernagel W. Double illegitimate recombination events integrate DNA segments through two different mechanisms during natural transformation of *Acinetobacter baylyi*. *Mol Microbiol* 2008; **67**: 984-95.
380. Pote J, Rosse P, Rosselli W *et al.* Kinetics of mass and DNA decomposition in tomato leaves. *Chemosphere* 2005; **61**: 677-84.
381. Nielsen KM, Townsend JP. Environmental exposure, horizontal transfer and selection of transgenes in bacterial populations. *NATO Sci Ser Ser A Life Sci*, 2001; 145-58.
382. Hartl DL, Clark AG. *Principles of population genetics*. Sunderland, MA, USA Sinauer Associates Inc., 1997.
383. Lærke HN, Hedemann MS. The digestive system of the pig. In: Knudsen KEB, Kjeldsen NJ, Poulsen HD et al., eds. *Nutritional physiology of pigs*. Foulum, Denmark: Danish Pig Research Center, 2012; 1-27.
384. Helander HF, Fandriks L. Surface area of the digestive tract - revisited. *Scand J Gastroenterol* 2014; **49**: 681-9.
385. Le Goff G, Milgen J, Noblet J. Influence of dietary fibre on digestive utilization and rate of passage in growing pigs, finishing pigs and adult sows. *Animal science (Penicuik, Scotland)* 2002; **74**: 503-15.
386. Bayerische Landesanstalt für Landwirtschaft. Futterberechnung für Schweine. Vöttinger Straße 38, 85354 Freising-Weihenstephan: Institut für Tierernährung und Futterwirtschaft. 2014.
387. AMA. Futtermittel: Schweine. http://www.rund-ums-schwein.at/index.php?id=futtermittel. 2014.
388. Martin-Orue SM, O'Donnell AG, Arino J *et al.* Degradation of transgenic DNA from genetically modified soya and maize in human intestinal simulations. *Br J Nutr* 2002; **87**: 533-42.
389. Prakash S, Tomaro-Duchesneau C, Saha S *et al.* The gut microbiota and human health with an emphasis on the use of microencapsulated bacterial cells. *J Biomed Biotechnol* 2011; **2011**: 981214.
390. Stephen AM, Cummings JH. The microbial contribution to human fecal mass. *J Med Microbiol* 1980; **13**: 45-56.
391. Wiesner RS, Hendrixson DR, DiRita VJ. Natural transformation of Campylobacter jejuni requires components of a type II secretion system. *J Bacteriol* 2003; **185**: 5408-18.
392. Munroe DL, Prescott JF, Penner JL. Campylobacter jejuni and Campylobacter coli serotypes isolated from chickens, cattle, and pigs. *J Clin Microbiol* 1983; **18**: 877-81.
393. Jensen AN, Dalsgaard A, Baggesen DL *et al.* The occurrence and characterization of Campylobacter jejuni and C. coli in organic pigs and their outdoor environment. *Vet Microbiol* 2006; **116**: 96-105.
394. Young CR, Harvey R, Anderson R *et al.* Enteric colonisation following natural exposure to Campylobacter in pigs. *Res Vet Sci* 2000; **68**: 75-8.
395. Boes J, Nersting L, Nielsen EM *et al.* Prevalence and diversity of Campylobacter jejuni in pig herds on farms with and without cattle or poultry. *J Food Prot* 2005; **68**: 722-7.
396. Suerbaum S, Lohrengel M, Sonnevend A *et al.* Allelic diversity and recombination in Campylobacter jejuni. *J Bacteriol* 2001; **183**: 2553-9.
397. Nordgard L, Nguyen T, Midtvedt T *et al.* Lack of detectable DNA uptake by bacterial gut isolates grown in vitro and by Acinetobacter baylyi colonizing rodents in vivo. *Environ Biosafety Res* 2007; **6**: 149-60.
398. de Boer P, Wagenaar JA, Achterberg RP *et al.* Generation of Campylobacter jejuni genetic diversity in vivo. *Mol Microbiol* 2002; **44**: 351-9.
399. Statistik Austria. Jahresergebnis der Schlachtungsstatistik. Retrieved October 16, 2014, from http://www.statistik.at/web_de/statistiken/land_und_forstwirtschaft/viehbestand_tierische_erzeugung/schlachtungen/. 2014.
400. Cotta MA, Whitehead TR, Zeltwanger RL. Isolation, characterization and comparison of bacteria from swine faeces and manure storage pits. *Environ Microbiol* 2003; **5**: 737-45.
401. Ramirez MS, Don M, Merkier AK *et al.* Naturally competent Acinetobacter baumannii clinical isolate as a convenient model for genetic studies. *J Clin Microbiol* 2010; **48**: 1488-90.

402. Dijkshoorn L, Nemec A, Seifert H. An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii. *Nat Rev Microbiol* 2007; **5**: 939-51.
403. Simpson DJ, Fry JC, Rogers HJ *et al.* Transformation of Acinetobacter baylyi in non-sterile soil using recombinant plant nuclear DNA. *Environ Biosafety Res* 2007; **6**: 101-12.
404. Juni E, Janik A. Transformation of Acinetobacter calco-aceticus (Bacterium anitratum). *J Bacteriol* 1969; **98**: 281-8.
405. Binh CT, Heuer H, Kaupenjohann M *et al.* Piggery manure used for soil fertilization is a reservoir for transferable antibiotic resistance plasmids. *FEMS Microbiol Ecol* 2008; **66**: 25-37.
406. Heuer H, Kopmann C, Binh CT *et al.* Spreading antibiotic resistance through spread manure: characteristics of a novel plasmid type with low %G+C content. *Environ Microbiol* 2009; **11**: 937-49.
407. Zhu Y-G, Johnson TA, Su J-Q *et al.* Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proceedings of the National Academy of Sciences* 2013.
408. Jechalke S, Focks A, Rosendahl I *et al.* Structural and functional response of the soil bacterial community to application of manure from difloxacin-treated pigs. *FEMS Microbiol Ecol* 2014; **87**: 78-88.
409. Joy SR, Bartelt-Hunt SL, Snow DD *et al.* Fate and transport of antimicrobials and antimicrobial resistance genes in soil and runoff following land application of swine manure slurry. *Environ Sci Technol* 2013; **47**: 12081-8.
410. Harms K, Bauer J. Detection and occurrence of antibiotics and their metabolites in pig manure in Bavaria (Germany). In: Keen PL, Montforts MHMM, eds. *Antimicrobial resistance in the environment*. Hoboken, New Jersey, USA: Wiley & Sons Inc., 2012; 293 - 307.
411. Zhang F, Li Y, Yang M *et al.* Content of heavy metals in animal feeds and manures from farms of different scales in northeast China. *Int J Environ Res Public Health* 2012; **9**: 2658-68.
412. Keese P. Risks from GMOs due to horizontal gene transfer. *Environ Biosafety Res* 2008; **7**: 123-49.
413. Heuer H, Smalla K. Horizontal gene transfer between bacteria. *Environ Biosafety Res* 2007; **6**: 3-13.
414. Choi IG, Kim SH. Global extent of horizontal gene transfer. *Proc Natl Acad Sci U S A* 2007; **104**: 4489-94.
415. French GL. The continuing crisis in antibiotic resistance. *Int J Antimicrob Agents* 2010; **36 Suppl 3**: S3-7.
416. Howard SJ, Catchpole M, Watson J *et al.* Antibiotic resistance: global response needed. *Lancet Infect Dis* 2013; **13**: 1001-3.
417. Heuer H, Schmitt H, Smalla K. Antibiotic resistance gene spread due to manure application on agricultural fields. *Curr Opin Microbiol* 2011; **14**: 236-43.
418. You Y, Silbergeld EK. Learning from agriculture: understanding low-dose antimicrobials as drivers of resistome expansion. *Frontiers in Microbiology* 2014; **5**.
419. Kozhevin PA, Vinogradova KA, Bulgakova VG. The soil antibiotic resistome. *Moscow University Soil Science Bulletin* 2013; **68**: 53-9.
420. Wright GD. The antibiotic resistome. *Expert Opin Drug Discov* 2010; **5**: 779-88.
421. Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat Rev Microbiol* 2007; **5**: 175-86.
422. Nielsen KM. An assessment of factors affecting the likelyhood of horizontal transfer of recombinant plant DNA to bacterial recipients in the soil and phytosphere. *Collection of biosafety reviews* 2003; **1**: 98 - 151.
423. Boxall AB, Fogg LA, Blackwell PA *et al.* Veterinary medicines in the environment. *Rev Environ Contam Toxicol* 2004; **180**: 1-91.
424. Dibner JJ, Richards JD. Antibiotic growth promoters in agriculture: history and mode of action. *Poult Sci* 2005; **84**: 634-43.
425. WHO. Tackling antibiotic resistance from a food safety perspective in Europe. WHO Regional Office for Europe, 2011.
426. Balachandran P, Hollingshead SK, Paton JC *et al.* The autolytic enzyme LytA of Streptococcus pneumoniae is not responsible for releasing pneumolysin. *J Bacteriol* 2001; **183**: 3108-16.

427. Whatmore AM, Dowson CG. The autolysin-encoding gene (lytA) of Streptococcus pneumoniae displays restricted allelic variation despite localized recombination events with genes of pneumococcal bacteriophage encoding cell wall lytic enzymes. *Infect Immun* 1999; **67**: 4551-6.

7 Annex 1 – Mosaic Genes: Individual Characterization

Table 28. Collection of mosaic genes as available in the scientific literature.

Focus: 2000 -2013; plus additional references added due to personal experience and by recommendations of the external scientific advisor

Gene	Function	Organism	Reference	Year	Accession No.	Comments
adk	Adenylate kinase Housekeeping gene	*Neisseria gonorrhoeae, N. meningitidis, N. lactamica, N. polysaccharea, N. cinerea, N. mucosa, N. pharyngis var. flava, N. flavescens, N. animalis*	A comparison of the nucleotide sequences of the adk and recA genes of pathogenic and commensal Neisseria species: evidence for extensive interspecies recombination within adk[283]	1996	NC_003116 REGION: 991810..992457 No access to full text journal Partial coding sequences: U57712.1 U57708.1 U57709.1 U57711.1 U57710.1 U57713.1 U57901.1	Adk: highly variable recA: no variation • Adk and recA sequences were determined from the same isolates of Neisseria gonorrhoeae, N. meningitidis, N. lactamica, N. polysaccharea, N. cinerea, N. mucosa, N. pharyngis var. flava, N. flavescens, and N. animalis. • Patterns of sequence divergence observed at adk and recA were very different. • Adk data suggest there has been a history of interspecies recombination within the adk gene of the human Neisseria species which has obscured the phylogenetic relationships between the species. • supported by Sawyer's runs test, and the Index of Association (IA) between codons, which provided significant evidence for interspecies recombination between the adk genes from the human Neisseria species, but no evidence of interspecies recombination between the recA sequences.
aroE	Shikimate dehydrogenase Housekeeping gene	*Neisseria gonorrhoeae, N. meningitidis*	Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of Neisseria meningitidis and commensal Neisseria species[200]	1997	NC_003112 REGION: complement (364869..365678) U82700 U82834 – U82848	Recombinational exchange in housekeeping genes between: a) aroE and glnA gene in meningococci b) meningococcal aroE + glnA with commensal Neisseria homologs • major distortions in the Neisseria species tree • indicates frequent exchanges

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						• no hitch-hiking effect due to positive selection of neighbouring genes • highly localized fragment exchange
atpD	ATP synthase Housekeeping gene	*Rhizobium*	Comparison of phylogeny analysis methods for rhizobia isolated from Albizia spp., Acacia spp. and Leucaena leucocephala[284]	2008	No sequences available	F1 complex, β subunit; Chinese
blp	Bacteriocin	*Streptococcus pneumoniae, Enterococcus faecalis*	Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae[285]	2001	No sequences available	Bacteriocin cluster; check • 10% of the genes altered between individual isolates and the reference strain; • variability within clones below 2.1%. • Ten gene clusters covering 160 kb account for half of the variable genes. Most of them are associated with transposases and are assumed to be part of a flexible gene pool within the bacterial population; • mosaic genes encoding antibiotic resistance determinants and gene clusters related to bacteriocin production
BoNT C/D	Toxin	*Clostridium botulinum*	Studies on botulinum neurotoxins type /C1 and mosaic/DC using Endopep-MS and proteomics[276, 286, 287]	2011	No sequences available	• BoNT serotypes /C1 and /D affect birds and mammals • potentially lethal to humans[286] • Botulinum C/D and D/C mosaic neurotoxins (BoNT/CD and /DC) are related to avian and bovine botulism[287] • highest toxicity to mice among all BoNTs[276]
BoNT C/D	toxin	*Clostridium botulinum*	P19 embryonal carcinoma cells exhibit high sensitivity to botulinum type C and D/C mosaic neurotoxins[288]	2012	No sequences available	• Both BoNT/C and BoNT/DC bind to, and are internalized by, neurons derived from P19 cells. • intracellular substrates for BoNT/C and BoNT/DC were cleaved by treatment of the cells with the toxins in a ganglioside-dependent manner. • P19 neurons exhibited high sensitivity to BoNT/C and BoNT/DC, to the same

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						extent as cultured primary neurons.
BoNT C/D	toxin	*Clostridium botulinum*	Environmental factors influencing the prevalence of a Clostridium botulinum type C/D mosaic strain in nonpermanent Mediterranean wetlands[289]	2013	No sequences available	• Significant association between the number of dead birds recorded in each botulism outbreak and the mean temperature in July (always >26 degrees C). • Low concentrations of Cl(-) and high organic matter content in sediments were significantly associated with the presence of C. botulinum. • The prevalence of C. botulinum was 18.2% (n = 22 pools) in aquatic invertebrates (Chironomidae and Corixidae families) and 33.3% (n = 18 pools) in necrophagous invertebrates (Sarcophagidae and Calliphoridae families), including two pools of adult necrophagous flies collected around bird carcasses. • The presence of the bacteria in the adult fly form opens up new perspectives in the epidemiology of avian botulism, since these flies may be transporting C. botulinum from one carcass to another.
BoNT C/D	toxin	*Clostridium botulinum*	Validation of a real-time PCR based method for detection of Clostridium botulinum types C, D and their mosaic variants C-D and D-C in a multicenter collaborative trial[290]	2013	No sequences available	• mosaic variants C-D and D-C that are associated with avian and mammalian botulism • concordance among the eight laboratories of 99.4%-100% for both arrays. • robust and suitable tools for rapid detection of C. botulinum types C, D and mosaic types C-D and D-C
BoNT C/D	toxin	*Clostridium botulinum*	Crystal structure of the receptor binding domain of the botulinum C-D mosaic neurotoxin reveals potential roles of lysines 1118 and 1136 in membrane interactions[291]	2011	No sequences available	• botulinum neurotoxins (BoNTs) produced by different strains of the bacterium Clostridium botulinum are responsible for the disease botulism and include a group of immunologically distinct serotypes (A, B, E, and F) • most lethal natural proteins known for humans.

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						• Two BoNT serotypes, C and D, while rarely associated with human infection, are responsible for deadly botulism outbreaks afflicting animals. • Also associated with animal infections is the BoNT C-D mosaic protein (BoNT/CD), a BoNT subtype that is essentially a hybrid of the BoNT/C (approximately two-third) and BoNT/D (approximately one-third) serotypes. • The BoNT/CD-HCR structure is similar to the two sub-domain organization observed for other BoNT HCRs • Liposome-binding experiments demonstrate that BoNT/CD-HCR binds phosphatidylethanolamine liposomes more tightly than BoNT/D-HCR.
BoNT C/D	toxin	*Clostridium botulinum*	Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms[292]	1996	D38442	• Primary reference • BoNT: 1285 amino acids (molecular weight of 147,364) • The BoNT of Dsa (BoNT/Dsa) is composed of three regions on the basis of the homology to BoNT types C1 (BoNT/C1) and D (BoNT/D). • The N-terminal (Met-1 to Val-522) and the C-terminal regions (Trp-945 to Glu-1285) have high identity to corresponding regions of BoNT/D (96% identity) and BoNT/C1 (74% identity), respectively. The core region (Pro-523 to Lys-944) is common to three toxins (83% to 92% identity). • The results suggest that neurotoxins produced from Clostridium botulinum types C and D are composed in a mosaic-like fashion.
cagA	Pathogenicity factor virulence	*Helicobacter pylori*	Assessment of the mosaic structure in the Helicobacter pylori cagA gene 3'-region using an improved polymerase chain reaction-based assay[293]	2012	No sequences available	• CagA protein: highly immunogenic protein. • Mitogenic[357]

Gene	Function	Organism	Reference	Year	Accession No.	Comments
comC comCD	competence	*Streptococcus pneumoniae* *S. mitis* *Streptococcus sp*	Transformation in Streptococcus pneumoniae: mosaic genes and the regulation of competence[274]	2000	No sequences available	Genetic competence depends on production of the competence signaling peptide CSP, the processed product of comC, which is curiously part of a mosaic gene arrangement itself. Expression of comC is part of a complex regulatory network which has studied in this work.
chvI chvG	non coding	*Rhizobium meliloti* *Rhizobium* sp. *Strain NGR234, Rhizobium leguminosarum, Agrobacterium rhizogenes*	Identification of Rhizobium-specific intergenic mosaic elements within an essential two-component regulatory system of Rhizobium species[294]	1995	U32869.1 U32941.1	Not relevant; no protein coding region
cyl	cytolysin	*Enterococcus faecalis (S. pneumoniae)*	Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae[285, 295]	2001	L37110	Comparative Study Using Type Strains and Clinical and Food Isolates To Examine Hemolytic Activity and Occurrence of the cyl Operon in Enterococci
dca/pptA	competence-associated protein / pilin phosphorylcholine transferase	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2012	AM421808.1 NC_008767.1 AL157959 AE004969	Not relevant
dhps	Housekeeping gene	*Neisseria meningitidis*	Transformational exchanges in the dihydropteroate synthase gene of Neisseria meningitidis: a novel mechanism for acquisition of sulfonamide resistance[297]	2000	X68062.1 X68069.1 X68068.1 X68067.1 X68066.1 X68065.1 X68064.1 X68063.1	Genetic competence depends on production of the competence signaling peptide CSP, the processed product of comC, which is curiously part of a mosaic gene arrangement itself. Expression of comC is part of a complex regulatory network which has studied in this work.
dotA	virulence	*Legionella pneumophila*	Molecular evolution of Legionella pneumophila dotA gene, the contribution of natural environmental strains[298]	1995	FN658686.1 FN658685.1 FN658684.1 FN658681.1 FN658680.1 FN658683.1 FN658682.1 FN652664.1 FN652663.1	Not relevant; no protein coding region
dpnI, dpnII	Restriction/modification	*Streptococcus pneumoniae*	Genetic basis of the complementary DpnI and DpnII restriction systems of S.	2001	M14339.1 M14340.1	Comparative Study Using Type Strains and Clinical and Food Isolates To Examine Hemolytic Activity and Occurrence of the cyl Operon in Enterococci

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			pneumoniae: an intercellular cassette mechanism[299]			
emm	Immune response modulation	*Streptococus pyogenes*	Horizontal gene transfer and the evolution of resistance and virulence determinants in Streptococcus[57, 300]	1997	CP003121.1 NC_017053.1 CP003116.1 NC_017040.1	The emm gene family includes *emm, enn, mrp* and others. Members of the *emm* gene family encode cell sufrace M and M-like proteins with binding properties for a range of human immune response proteins. Traditionally, M-typing has been used for classification of *S.pyogenes* strains. However identical *emm* and *enn* alleles have been found in strains that are divergent in overall genetic relatedness. In addition mosaic enn and emm genes have been described.
emm-like	Immune response modulation	*S. pyogenes*	Horizontal gene transfer in the evolution of group A streptococcal emm-like genes: gene mosaics and variation in Vir regulons[301]	1994	CP003121.1 NC_017053.1 CP003116.1 NC_017040.1	The emm gene family includes *emm, enn, mrp* and others. Members of the *emm* gene family encode cell sufrace M and M-like proteins with binding properties for a range of human immune response proteins. Traditionally, M-typing has been used for classification of *S.pyogenes* strains. In this work a new emm-like gene is described, which has a mosaic structure and provided the first clear evidence that the horizontal transfer of emm-like sequences between distinct strains contributes to the evolution of group A streptococcal emm-like genes and Vir regulons.
fhaB	hemagglutinin	*Neisseria meningitidis*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. fhaB is an example of “Silent Gene Cassette mediated Variation”
frpB / fetA	iron-regulated outer membrane protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						diversification mechanisms in N. meningitidis. frpB / fetA in supplemental material
genome	Non protein coding	*Helicobacter pylori*	Mosaic DNA imports with interspersions of recipient sequence after natural transformation of Helicobacter pylori[302]	2008	No sequences available	Important. Short sequence imports
genome		*bacteria*	Interspecies recombination and mismatch repair. Generation of mosaic genes and genomes[239]	2000	No sequences available	Important! Connection mosaic gene formation and gene repair
glnA	Glutamine synthetase Housekeeeping gene	*Neisseria gonorrhoeae?*	Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of Neisseria meningitidis and commensal Neisseria species[200]	1997	U82849 – U82863	Visual inspection showed clear evidence of a history of intraspecies recombinational exchanges within the neighbouring meningococcal shikimate dehydrogenase (aroE) and glutamine synthetase (glnA) genes, which was supported by the non-congruence of the trees constructed from the sequences of these genes from different meningococcal strains, and by statistical tests for mosaic structure.
gnd	6-phosphogluconate dehydrogenase Housekeeping gene	*E. coli*	Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (gnd) in enteric bacteria[303]	1994	U14463.1 U14462.1 U14461.1 U14460.1 U14459.1 U14458.1 U14457.1 U14456.1 U14455.1 U14454.1	High frequency of recombination at gnd locus in E. coli (not in *Salmonella* enterica): • distortion in E. coli species tree • no exchange between S. enterica and E. Coli (but with Citrobacter + Klebsiella) • close linkage with rfb region (subject to diversifying selection
hmbR	hemoglobin receptor	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. hmbR in supplemental material
hpuA	hemoglobin-haptoglobin utilization protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed	2007	AM421808.1 NC_008767.1	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			through Comparative Analysis of Serogroup C Strain FAM18[296]		AL157959 AE004969	ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. hpuA in supplemental material
IgA protease porA	protease	*Neisseria sp* *Streprococcus pneumoniae*	Mosaic genes and their role in penicillin-resistant Streptococcus pneumoniae (minireview)[269]	1998	No sequences available	Old Minireview on the role of mosaic genes in penicillin resistant S. pneumoniae. The presence of mosaic genes encoding IgA protease, or PorA outer membrane porin in different clones of *Neisseria* are mentioned in the introduction, reference: Achtman, M., *Electrophoresis* 1997, *19,* 593-596.
Intergenic region Undefined ORFs	Undefined ORFs	*Escherichia coli*	Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between trp and tonB[304]	1988	No sequences available	3500-bp lying between the tryptophan operon and the tonB gene were sequenced in *Escherichia coli* strain K12. Analysis of the sequence yields six open reading frames that have properties characteristic of genes for proteins. The region was compared to several wild strains of E.coli. DNA rearrangements, that have end points within the open reading frames, were discovered.
lgtA	acto-N-neotetraose biosynthesis glycosyl transferase	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. lgtA in supplemental material
lgtC	lipopolysaccharide biosynthesis protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. lgtC in supplemental material
lgtD	lipopolysaccharide	*Neisseria*	Meningococcal Genetic	2007	AM421808.1	The authors have sequenced the genome of N.

Gene	Function	Organism	Reference	Year	Accession No.	Comments
	biosynthesis protein		Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]		NC_008767.1 AL157959 AE004969	meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. lgtD in supplemental material
lgtG	lipopolysaccharide glycosyl transferase	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. lgtG in supplemental material
lktA	Leukotoxin virulence factor	*Mannheimia (Pasteurella) haemolytica*	Sequence Diversity and Molecular Evolution of the Leukotoxin (lktA) Gene in Bovine and Ovine Strains of Mannheimia (Pasteurella) haemolytica[36]	2001	AF314525.1 - AF314500.1	virulence factor in the pathogenesis of both bovine and ovine pneumonic pasteurellosis
lktCABD operon	Leukotoxin virulence factor	*Mannheimia (Pasteurella) haemolytica, Mannheimia glucosida, Pasteurella trehalosi*	Mosaic structure and molecular evolution of the leukotoxin operon (lktCABD) in Mannheimia (Pasteurella) haemolytica, Mannheimia glucosida, and Pasteurella trehalosi[305]	2002	AF314525.1 - AF314500.1	virulence factor in the pathogenesis of both bovine and ovine pneumonic pasteurellosis nucleotide sequence comparison of lktA in 31 bovine and ovine strains representing the various evolutionary lineages and serotypes. Eight major allelic variants (1.4 to 15.7% nucleotide divergence) were identified; these have mosaic structures. Five allelic variants occur only in ovine strains and consist of recombinant segments derived from as many as four different sources.
LOS	Lipooligo-saccharide biosynthesis region	*Campylobacter jejunii*	Characterization of Lipooligosaccharide-Biosynthetic Loci of Campylobacter jejuni Reveals New Lipooligosaccharide Classes: Evidence of Mosaic Organizations[306]	2008	EU404106.1 EU404105.1 EU404111.1 EU410351.1 EU410350.1 EU404112.1 EU404110.1	The lipo-oligosaccharide (LOS) biosynthesis region is one of the more variable genomic regions between strains of Campylobacter jejuni. the authors sequenced the LOS biosynthesis loci of 15 strains . The authors identified 11 new classes of LOS loci that exhibited examples of deletions and insertions of genes and cassettes of genes found in

Gene	Function	Organism	Reference	Year	Accession No.	Comments
					EU404109.1 EU404108.1 EU404107.1 EU404104.1	other LOS classes or capsular biosynthesis loci leading to mosaic LOS loci.
lytA	Autolysin	*streptococci*	Horizontal gene transfer and the evolution of resistance and virulence determinants in Streptococcus[57]	1997	CP003121.1 NC_017053.1 CP003116.1 NC_017040.1	N-acetylmuramyl-L-alanine amidase • cell wall • binding to choline moiety of lipoteichoic acid • cell division, autolysis, AB induced lysis • virulence factor in combination with pneumolysin • maybe not involved in the release of pneumolysin[426] • mediates bile salts effect (deoxycholate) • composed of mosaics: divergence 20% recombination with bacteriophage encoded autolysin genes possible[427] **rather highly conserved gene**[427] • 0,11 – 3,20% sequence diversity between 9 allels[427] • Variation not randomly distributed: blocks • very localized recombination events • donors: bacteriophages
maf	adhesin	*N. lactamica*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. The maf genes are examples of “Silent Gene Cassette mediated Variation”
mafB	adhesin	*Neisseria meningitidis*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			FAM18[296]			meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. The maf genes are examples of "Silent Gene Cassette mediated Variation"
mefE	resistance	*Streptococcus pneumoniae*	Mosaic genes and mosaic chromosomes in S. pneumoniae[307]	2004	No sequences available	Genom comparison of the nonencapsulated laboratory strain R6, and of the serotype 4 TIGR strain isolated recently. Many regions bear signs of gene transfer events such as the presence of insertion sequences, transposable elements, and putative site-specific integrases/recombinases. Mef: putative macrolide efflux system, it is located in a Strain-specific gene clusters in the R6 genome.
murM	cell wall	*Streptococcus pneumoniae*	Distribution of the mosaic structured murM genes among natural populations of Streptococcus pneumoniae[308]	2000	AF281140.1 AF281139.1 AF281138.1 AF281137.1 AF281136.1 AF281135.1	The presence and sequence variation of the murM gene were studied in a large collection (814 strains) of genetically diverse Streptococcus pneumoniae isolates using nucleotide probes. Complete sequencing of murM from a group of penicillin-resistant isolates allowed the identification of a number of different alleles that differed in the length and exact position of the divergent sequences within the particular murM.
gyrA, gyrB	replication	*Mycobacterium tuberculosis*	Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis[309]	2005	AM283534.1 AM283533.1 AM283532.1 AM283531.1 AM283530.1	Mycobacterium tuberculosis has an extremely low level of genetic variation, which suggests that the entire population resulted from clonal expansion following an evolutionary bottleneck around 35,000 y ago. Here, the authors show that this population constitutes just the visible tip of a much broader progenitor species, whose extant representatives are human isolates of tubercle bacilli from East Africa. In these isolates, the authors detected incongruence among gene phylogenies as well as mosaic gene sequences, whose individual elements are retrieved in classical M. tuberculosis.
nadA	putative adhesin / invasin	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. nadA in supplemental material
nanA	Neuraminidase virulence	*Streptococcus pneumoniae*	Horizontal gene transfer and the evolution of resistance and virulence determinants in Streptococcus[57, 310]	1997	CP003121.1 NC_017053.1 CP003116.1 NC_017040.1	• Cleaves terminal sialic acid residues • Reveal attachment sites for pneumococcal adhesion • Important for bacterial nutrition • 2 size classes of nanA genes: 3' direct repeats • DR similar to a region in IgA protease (S. sanguis) • 5' region differs in 35% from the homologous region of the donor • Direct correlation between NANA and coma and advrse outcome of menigitis[310]
nanB	Neuraminidase virulence	*Streptococcus pneumoniae*	Structural and functional studies of Streptococcus pneumoniae neuraminidase B: An intramolecular trans-sialidase[311]		No sequences available	Not relevant: paper is not about mosaicism (keywort mosaic missing in full text). Here, the authors report the first structure of a neuraminidase from S. pneumoniae: the crystal structure of NanB.
nanC	Neuraminidase virulence	*Streptococcus pneumoniae*	Variation in the presence of neuraminidase genes among Streptococcus pneumoniae isolates with identical sequence types[312]		NZ_ABAM02000001.1 DQ351287.1 NC_011083.1 NC_011080.1 NC_011094.1 NC_011149.1	
NIME	neisserial intergenic mosaic elements	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. Our analysis provides evidence for the hypothesis that the noncoding repeat arrays in neisserial genomes (neisserial intergenic mosaic elements) provide a crucial mechanism for the generation of surface antigen variants.

Gene	Function	Organism	Reference	Year	Accession No.	Comments
NTNH	non-toxic-non-haemagglutinins virulence	*Clostridium botulinum*	Phylogeny and taxonomy of the food-borne pathogen Clostridium botulinum and its neurotoxins[313]	1998	No sequences available	Full text missing
opa	Membrane protein	*Neisseria*	The Repertoire of Minimal Mobile Elements in the Neisseria Species and Evidence That These Are Involved in Horizontal Gene Transfer in Other Bacteria[314]	2007	DQ115769.1 DQ115760.1 DQ115774.1 DQ115771.1 DQ115764.1 DQ117942.1 DQ115777.1 DQ115776.1 DQ115775.1 DQ115773.1 DQ115772.1 DQ115770.1 DQ115768.1 DQ115767.1 DQ115766.1 DQ115765.1 DQ115763.1 DQ115762.1 DQ115759.1 DQ115758.1 DQ115757.1	In the Neisseria spp., natural competence for transformation and homologous recombination generate antigenic variants through creation of mosaic genes (such as opa's) and through recombination with silent cassettes (such as pilE/pilS) and gene-complement diversity through the horizontal exchange of whole genes or groups of genes, in minimal mobile elements (MMEs). The paper is not very relevant, because it is about the MMEs. Mosaic opa genes are only mentioned in the introduction without reference.
opc	class 5 outer membrane protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. opc in supplemental material
ompA	Outer membrane protein	*Mannheimia (Pasteurella) haemolytica, Mannheimia glucosida, Pasteurella trehalosi*	Sequence Diversity and Molecular Evolution of the Heat-Modifiable Outer Membrane Protein Gene (ompA) of Mannheimia (Pasteurella) haemolytica,	2004	AY244653.1,AY244654.1, AY244655.1,AY244656.1, AY244657.1,AY244658.1, AY244659.1,AY244660.1, AY244661.1,AY244662.1, AY244663.1,AY244664.1,	The OmpA (or heat-modifiable) protein is a major structural component of the outer membranes of gram-negative bacteria. The genetic diversity and molecular evolution of OmpA were investigated in 31 Mannheimia (Pasteurella) haemolytica, 6 Mannheimia glucosida, and 4 Pasteurella trehalosi

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			Mannheimia glucosida, and Pasteurella trehalosi[315]		AY244665.1,AY244666.1, AY582755.1,AY582756.1, AY582757.1,AY582758.1	strains. Class III alleles have mosaic structures and were derived by horizontal DNA transfer and intragenic recombination. It is proposed that the OmpA protein of M. haemolytica acts as a ligand and is involved in binding to specific host cell receptor molecules in cattle and sheep.
ompA	Outer membrane Protein	*Haemophiulus parasuis*	Characteristics of the molecular diversity of the outer membrane protein A gene of Haemophilus parasuis[316]	2010	FJ667948.1,FJ667949.1,FJ667950.1,FJ667951.1,FJ667952.1,FJ667953.1,FJ667954.1,FJ667955.1,FJ667956.1,FJ667957.1,FJ667958.1,FJ667959.1,FJ667960.1,FJ667961.1,FJ667962.1,FJ667963.1,FJ667964.1,FJ667965.1,FJ667966.1,FJ667967.1,FJ667968.1,FJ667969.1,FJ667970.1,FJ667971.1,FJ667972.1,FJ667973.1,FJ667974.1,FJ667975.1,FJ667976.1,FJ667977.1,FJ667978.1,FJ667979.1,FJ667980.1,FJ667981.1,FJ667982.1,FJ667983.1,FJ667984.1,FJ667985.1,FJ667986.1,FJ667987.1,FJ667988.1,FJ667989.1,FJ667990.1,FJ667991.1,FJ667992.1,FJ667993.1,FJ667994.1,FJ667995.1,FJ667996.1,FJ667997.1,FJ667998.1,FJ667999.1,FJ668000.1,FJ668001.1,FJ668002.1,FJ668003.1,FJ668004.1,FJ668005.1,FJ668006.1,FJ668007.1,FJ668008.1,FJ668009.1,FJ668010.1,FJ668011.1,FJ668012.1,FJ668013.1,FJ668014.1,FJ668015.1	In this study, the structural characteristics, sequence types, and genetic diversity of ompA were investigated in 15 H. parasuis reference strains of different serovars and 20 field isolates. Keyword “mosaic” missing in the full text.
operon		*several*	Evolution of mosaic operons by horizontal gene transfer and gene displacement in	2003	No sequences available	Not relevant: higher level mosaics (operons)

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			situ[317]			
oprF	Membrane protein	*Pseudomonas*	Unrecognized fine-scale recombination can mimic the effects of adaptive radiation[318]	2013	No sequences available	Recently, Bodilis et al. (2011) observed artificial phylogenetic incongruence using the Pseudomonas surface protein gene oprF,. In this study, an alternative hypothesis, namely fine-scale recombination, was tested on the same dataset. The results reveal that regions in oprF are of different evolutionary origins, and the mosaic gene structure resulted in confounding phylogenetic signals.
parC, parE, gyrA	Topoisomerase 4	*Streptococcus mitis* *S. pneumoniae*	Fluoroquinolone resistance mutations in the parC, parE, and gyrA genes of clinical isolates of viridans group streptococci[319]	1998	AF170998.1,AF170997.1, AF170996.1,AF079208.1, AF079207.1,AF079206.1, AF079205.1,AF079204.1, AF079203.1,AF079202.1, AF079201.1,AF079200.1, AF079199.1,AF079198.1, AF079197.1,AF079196.1, AF079195.1,AF079194.1, AF079193.1,AF079192.1, AF079191.1,AF079190.1, AF079189.1,AF079188.1, AF079187.1,AF079214.1, AF079213.1,AF079212.1, AF079211.1,AF079210.1, AF079209.1,AF170994.1, AF170993.1,AF079220.1 AF079219.1 AF079218.1 AF079217.1 AF079216.1 AF079215.1	Not relevant: paper about mutations not about rearrangements
pbp1b	transpeptidase resistance	*Streptococcus pneumoniae*	Analysis of penicillin-binding protein lb and 2a genes from Streptococcus pneumoniae[320]	2000	NC_009674.1 NC_004722.1	In clinical isolates of Streptococcus pneumoniae were Penicillin-binding protein (pbp) 1b and 2a genes were analysed by DNA fingerprinting methods. These results show that the pbp1b and 2a genes examined here do not display the typical mosaic gene patterns observed in the pbp2x, 2b, and 1a genes of penicillin-resistant pneumococci. Transformation experiments with pbp1b and 2a genes isolated from two resistant strains failed to

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						transform pneumococcal strains to increased levels of penicillin resistance, suggesting that PBPs 1B and 2A may not play a role in the development of penicillin resistance in some pneumococci.
pbp2	transpeptidase resistance	*Neisseria gonorrhoeae*	Emergence and spread of Neisseria gonorrhoeae clinical isolates harboring mosaic-like structure of penicillin-binding protein 2 in Central Japan[321]	2005	No sequences available	DNA based screening of clinical isolates for mosaic PBP2.
pbp2	Transpeptidase resistance	*Neisseria gonorrhoeae*	Mosaic penicillin-binding protein 2 in Neisseria gonorrhoeae isolates collected in 2008 in San Francisco, California[322]	2009	No sequences available	In this study the qPCR method of Occhiai et al (2008) was applied to screen 54 isolates for presence of the mosaic penA gene. 5 were positive. In another 100 islates from 2002-2006 none were positive. -> temporal development!
pbp2	transpeptidase resistance	*Neisseria gonorrhoeae*	Amino acid substitutions in mosaic penicillin-binding protein 2 associated with reduced susceptibility to cefixime in clinical isolates of Neisseria gonorrhoeae[323]	2006	No sequences available	The complete sequence of ponA, penA, and por genes, encoding, respectively, PBP1, PBP2, and porin, were determined for 58 strains isolated in 2002 from Japan. In order to identify the mutations responsible for the reduced susceptibility to cefixime in isolates with mosaic PBP2, penA genes with various mutations were transferred to a susceptible strain by genetic transformation. 3 mutations (G545S, I312M, and V316T) in mosaic PBP2 were identified as the amino acid substitutions responsible for reduced susceptibility to cefixime in N. gonorrhoeae.
pbp2	transpeptidase resistance	*Neisseria gonorrhoeae*	Molecular and Structural Analysis of Mosaic Variants of Penicillin-Binding Protein 2 Conferring Decreased Susceptibility to Expanded-Spectrum Cephalosporins in Neisseria gonorrhoeae: Role of Epistatic Mutations[324]	2010	No sequences available	Three of the similar to 60 mutations present in mosaic alleles of penA, G545S, I312M, and V316T, have been reported to be responsible for increased resistance, especially to cefixime (Takahata, S., et al. , 2006). The authors show that these three mutations display epistasis, in that their capacity to increase resistance to p-lactam antibiotics is dependent on the presence of other mutations in the mosaic alleles.
pbp2	Transpeptidase resistance	*Neisseria meningitidis*	Recruitment of a penicillin-binding protein gene from Neisseria flavescens during the emergence of penicillin resistance in Neisseria	1989	M26645.1 M26644.1	N. flavescens: low affinity PBP Pre-antibiotic era isolates + recent

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			meningitidis [325]			
pbp2B	transpeptidase Resistance	*Streptococcus pneumoniae*	Evolution of penicillin resistance in Streptococcus pneumoniae; the role of Streptococcus mitis in the formation of a low affinity PBP2B in S. pneumoniae[326]	1993	No sequences available	S. mitis: no low affinity PBP But mutation of single AS leads to resistance (Thr445Ala) • Blocks of sequences up to 21% divergence to sensitive PBP2 genes • Some PBPs contain a single block from S. mitis • Others consist of a complex organisation involving several recombination events and several donors
pbp2X	transpeptidase resistance	*Streptococcus pneumoniae*	An important site in PBP2x of penicillin-resistant clinical isolates of Streptococcus pneumoniae: mutational analysis of Thr338[327]	2009	No sequences availabe	Site-directed experiments are presented. The data support the view that PBP2x and PBP1a interact with each other on some level and that alterations of both PBPs in resistant clinical isolates have evolved to ensure cooperation between both proteins.
pbpX (PBP2X)	transpeptidase resistance	*Streptococcus pneumoniae*	Mosaic pbpX genes of major clones of penicillin-resistant Streptococcus pneumoniae have evolved from pbpX genes of a penicillin-sensitive Streptococcus oralis[328]	1994	X78217.1 X78215.1 X78216.1	1.The authors have now identified a gene homologous to the pneumococcal PBP2x gene (pbpX) in a penicillin-sensitive Streptococcus oralis isolate M3. 2.Transformation experiments: With low concentrations of cefotaxime, transformants of the sensitive S. pneumoniae R6 strain could be selected containing pbpX genes from either S. mitis NCTC 10712 or S. oralis M3, demonstrating that genetic exchange can already occur between beta-lactam-sensitive species.
pcn	bacteriocin	*Streptococcus pneumoniae*	Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae[285]	2001	No sequences available	Microarray study. Pcn only shown in one figure and mentioned in the abstract. No info, if pcn is a mosaic gene. The authors investigated genomic variation in 20 S. pneumoniae isolates representing major antibiotic-resistant clones and 10 different capsular serotypes.
penA	resistance	*Neisseria gonorrhoeae*	Rapid detection of the mosaic structure of the Neisseria gonorrhoeae penA Gene, which is associated with decreased susceptibilities to oral cephalosporins[329]	2008	No sequences available	The aim of this paper was to develop a qPCR detection method for mosaic penA. The method was used to screen 621 clinical isolates and had a sensitivity of >=10^1

Gene	Function	Organism	Reference	Year	Accession No.	Comments
penA	resistance	*Neisseria gonorrhoeae*	Characteristics and population dynamics of mosaic penA allele-containing Neisseria gonorrhoeae isolates collected in Sydney, Australia, in 2007-2008[330]	2010	No sequences available	1800 Neisseria gonorrhoeae isolates collected in Sydney in 2007 and 2008 were examined for mosaic penA alleles that mediated cephalosporin resistance. In 2008, there were substantial increases in proportions (from 1.5 to 10.3%) of mosaic-containing gonococci and major shifts in genotypic patterns, with 10 new genotypes representing 74 of the 85 mosaic-containing isolates , and genotypes detected between 2001 and 2005 having disappeared.
penA	resistance	*Neisseria gonorrhoeae*	High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure[331]	2012	JQ073701.1	Here, the authors confirm and characterize a second strain (F89) with high-level cefixime and ceftriaxone resistance which was isolated in France and most likely caused a treatment failure with cefixime. A novel penA mosaic allele (penA-CI), which was penA-XXXIV with an additional A501P alteration in penicillin-binding protein 2, was the primary determinant for high-level ESC resistance, as determined by transformation into a set of recipient strains.
penA	resistance	*Neisseria gonorrhoeae*	Use of a novel screening PCR indicates presence of Neisseria gonorrhoeae isolates with a mosaic penA gene sequence in Australia[332]	2007	No sequences available	Only title available.
penA (pbp2)	resistance	*Neisseria gonorrhoeae*	penicillin-binding protein 2 Gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime[333]	2002	AB071984.1 AB536877.1	Neisseria gonorrhoeae strains with reduced susceptibility to cefixime (MICs, 0.25 to 0.5 micro g/ml) were isolated from male urethritis patients in Tokyo, Japan, in 2000 and 2001. Some regions in the transpeptidase-encoding domain in their penA gene were similar to those in the penA genes of Neisseria perflava (N. sicca), Neisseria cinerea, Neisseria flavescens, and Neisseria meningitidis. These results showed that a mosaic-like structure in the penA gene conferred reductions in the levels of susceptibility of N. gonorrhoeae to cephems and penicillin in a manner similar to that found for N. meningitidis and Streptococcus pneumoniae.
penA (pbp2)	resistance	*Neisseria gonorrhoeae*	Tracing subsequent dissemination of a cluster of gonococcal infections caused	2013	JQ782218.1	The authors characterized the 47 isolates of a strain cluster of ST4378. Forty-six of the 47 isolates had a mosaic penA allele type XXXIV and one had a

Gene	Function	Organism	Reference	Year	Accession No.	Comments
			by an ST1407-related clone harbouring mosaic penA alleles in Taiwan[334]			new allele type XL, which appeared to be a recombinant of mosaic penA type XXXIV and non-mosaic penA type II.
penA (pbp2)	resistance	*Neisseria gonorrhoeae*	Characteristics and dissemination of mosaic penicillin-binding protein 2-harboring multidrug-resistant Neisseria gonorrhoeae isolates with reduced cephalosporin susceptibility in northern Taiwan[335]	2010	No sequences available	Among 254 Neisseria gonorrhoeae isolates from a sexually transmitted infection (STI) clinic in northern Taiwan, 69 isolates were found to contain the mosaic penA (MA) gene and were associated with elevated cefixime and ceftriaxone MICs.
penA (pbp2)	resistance	*Neisseria gonorrhoeae*	A duplex PCR method to identify mosaic penA gene and predict reduced susceptibility to oral cephalosporins in Neisseria gonorrhoeae.[336]	2010	No sequences available	Full text missing
pglE	pilin glycosylation protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. pglE in supplemental material
pglG	pilin glycosylation protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. pglG in supplemental material
pglH	pilin glycosylation protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						diversification mechanisms in N. meningitidis. pglH in supplemental material
pgtA / pglA	pilin glycosylation protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. pgtA/pglA in supplemental material
phoA	Alkaline phosphatase	*Escherichia coli*	Genetic exchange among natural isolates of bacteria: recombination within the phoA gene of Escherichia coli[337]	1988	ACNS01000001.1,AE017355.1,AEVV03000001.1,AEVW03000001.1,AM270990.1,AQPG01000012.1,AQRD01000007.1,AVPM01000001.1,CM002129.1,CP000802.1,CP001022.1,CP003289.1,CP003297.1,CP003301.1,CP004848.1,CP004849.1,CP004851.1,CP004852.1,CP004853.1,CP004855.1,CP006937.1,CP006980.1,CP006981.1,CP006982.1,CP006983.1,CP006984.1,CP006985.1,CP007390.1,CP007391.1,CP007393.1,CP007394.1,CP007659.1,CP007690.1,M29663.1,M29664.1,M29665.1,M29666.1,M29667.1,M29668.1,M29669.1,M29670.1,NC_002745.2,NC_005957.1,NC_006274.1,NC_009800.1,NC_009801.1,NC_010556.1,NC_021710.1,NC_021712.1,NC_021713.1,NC_021714.1,NC_021717.1,NC_023045.1,NC_023066.1,NZ_AAYU01000001.1,NZ_ACNS01000001.	A 1871-nucleotide region including the phoA gene was cloned and sequenced from eight naturally occurring strains of Escherichia coli. Maximum parsimony analysis revealed six equally parsimonious trees with a consistency index of 0.80. Of the 42 informative sites, 22 were inconsistent with each of the maximum parsimony trees. The implication is that different segments of the phoA gene have different phylogenetic histories.

Gene	Function	Organism	Reference	Year	Accession No.	Comments
					1,NZ_AEVV03000001.1,NZ_AEVW03000001.1,NZ_AFXI01000001.1,NZ_AFXJ01000001.1,NZ_AFXK01000001.1,NZ_AVPM01000001.1,NZ_CM001858.1,NZ_CP006980.1,NZ_CP006981.1,NZ_CP006982.1,NZ_CP006983.1,NZ_CP006985.1	
phoN	phosphatase	*Salmonella*	Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome[338]	1992	X63599.1 AM270994.1	Not relevant. Differences in GC content and other characteristics compared to the S. typhimurium chromosome suggest that the phoN gene, and perhaps the entire loop structure, was acquired by lateral transmission in a plasmid-mediated event.
pilA	pilin biogenesis protein	*Gram negative bacteria, Pseudomonas*	Type IV pili and twitching motility[339]	2002	12770 hits	High variability of C-terminal ends
pilA	pilin biogenesis protein	*Clostridium perfringens*	Skewed genomic variability in strains of the toxigenic bacterial pathogen, Clostridium perfringens[340]	2006	CP000246.1,CP000312.1,CP000313.1,CP000314.1,CP000315.1,NC_008261.1,NC_008262.1,NC_008263.1,NC_008264.1,NC_008265.1,NR_121697.1,NR_121812.1,NR_121927.1	High variability of C-terminal ends
pilC	pilin biogenesis protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	X63599.1 AM270994.1	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. PilC is an example of “Repeat Arrays and Flanking Genes”
pilE	Surface protein	*Neisseria gonorrhoeae*	Pilin gene variation in Neisseria gonorrhoeae: reassessing the old paradigms[341]	2009	No sequences available	In Neisseria gonorrhoeae various surface antigens can change either by antigenic variation using RecA-dependent recombination schemes (e.g. PilE antigenic variation) or, alternatively, through phase variation (on/off switching) in a RecA-

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						independent fashion. PilE antigenic variation occurs following genetic recombination of the PilE expression locus (*pilE*) with one or more silent *pil* loci (*pilS*). All *pil* genes contain variable *pil* gene segments (cassettes) interspersed with short conserved regions.
pilS	Surface protein	*Neisseria meningitidis*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. The pilS genes are examples of "Silent Gene Cassette mediated Variation"
pln	toxin	*Lactobacillus plantarum*	Bacteriocin loci in Lactobacillus plantarum[342]	2009	1963 hits	Not relevant. Mosaic locus containing (entire) genes of different origin). (verify in detail, if there is also intra genic recombination)
por	outer membrane protein porin	*Neisseria gonorrhoeae*	Genetic Diversity and Mosaicism at the por Locus of Neisseria gonorrhoeae[343]	1999	AF090792.1,AF090793.1, AF090794.1,AF090795.1, AF090796.1,AF090797.1, AF090798.1,AF090799.1, AF090800.1,AF090801.1, AF090802.1,AF090803.1, AF090804.1,AF090805.1, AF090806.1,AF090807.1, AF090808.1,AF090809.1, AF090810.1,AF090811.1	The por genes of the predominant serovars of *Neisseria gonorrhoeae* circulating in a high-frequency transmitter core group located in Nairobi, Kenya, were examined for nucleotide sequence polymorphism. The level of por gene diversity did not differ significantly between core group-derived gonococcal strains and gonococcal strains originating elsewhere. However, por mosaicism appeared to be more frequent among core group-derived strains.
porA	outer membrane protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. The porA locus is an example of "Repeat Arrays and Flanking Genes"
pspA	Surface protein,	*Streptococcus pneumoniae*	Pneumococcal surface protein.	2000	No sequences available	Pneumococcal surface protein A (PspA) is a

Gene	Function	Organism	Reference	Year	Accession No.	Comments
	immune system evasion		Diversity of PspA: mosaic genes and evidence for past recombination in Streptococcus pneumoniae[32]			serologically variable protein of Streptococcus pneumoniae. Twenty-four diverse alleles of the pspA gene were sequenced. A highly mosaic gene structure was observed in which individual mosaic sequence blocks in PspAs diverged from each other by over 20% in many cases. This level of divergence exceeds that observed for blocks in the penicillin-binding proteins of S. pneumoniae.
pspA	Surface protein, immune system evasion	*Streptococcus pneumoniae*	Monitoring the long-term molecular epidemiology of the pneumococcus and detection of potential 'vaccine escape' strains[344]	2011	No sequences available	Not directly relevant. Development of a resequencing array consisting of 5 conserved and 6 variable genes to characterize 72 Streptococcus pneumoniae isolates. the other variable genes (encoding cell surface proteins and surface anchor proteins are described as "variable" with many SNPs, but no mosaicisms are reported.
pspA	Surface protein, immune system evasion	*Streptococcus pneumoniae*	Comparison of the PspA sequence from Streptococcus pneumoniae EF5668 to the previously identified PspA sequence from strain Rx1 and ability of PspA from EF5668 to elicit protection against pneumococci of different capsular types[345]	1998	U89711.1	Comparison of two pspA genes . Sequencing of the gene encoding PspA/EF5668 revealed 71% identity with that of PspA/Rx1.
pspA	Surface protein, immune system evasion	*Streptococcus pneumoniae*	Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of Streptococcus pneumoniae[346]		No sequences available	pspA The protein was originally defined by the two mouse monoclonal antibodies Xi64 and Xi126, which together recognized PspA on 14% of pneumococcal isolates. In the present study the authors demonstrated that PspA is produced by all pneumococci, confirming our hypothesis that there are variants of PspA which are not detected by Xi64 and Xi126. When the 53 strains reactive with the monoclonal antibodies were analyzed by capsular type as well as by serologic type and molecular weight of PspA, the authors observed 50 different clonotypes of pneumococci.
pspA	Surface protein, immune system evasion	*Staphylococcus aureus* *Clostridium perfringens*	Skewed genomic variability in strains of the toxigenic bacterial pathogen, Clostridium perfringens.[340]	1998	CP000246.1,CP000312.1, CP000313.1,CP000314.1, CP000315.1,NC_008261.1 ,NC_008262.1,NC_00826	Comparison of two pspA genes . Sequencing of the gene encoding PspA/EF5668 revealed 71% identity with that of PspA/Rx1.

Gene	Function	Organism	Reference	Year	Accession No.	Comments
					3.1,NC_008264.1,NC_008265.1,NR_121697.1,NR_121812.1,NR_121927.1	
RNaseP	RNAseP	*Pyrococcus-horikoshii ot3*	Archaeal RNase P: A Mosaic of Its Bacterial and Eukaryal Relatives[347]	2010	CP002906.1 NC_017195.1	Only Abstract available
siaD	Capsule biosynthesis protein	*Neisseria*	Meningococcal Genetic Variation Mechanisms Viewed through Comparative Analysis of Serogroup C Strain FAM18[296]	2007	AM421808.1 NC_008767.1 AL157959 AE004969	The authors have sequenced the genome of N. meningitidis strain FAM18, a representative of the ST-11/ET-37 complex, providing the first genome sequence for the disease-causing serogroup C meningococci. Genome comparison to other strains was carried out. Here the authors provide a detailed view of novel genetic diversification mechanisms in N. meningitidis. siaD in supplemental material.
sfbl	Virulence factor	*Streptococcus pyogenes*	Evolution of sfbI encoding streptococcal fibronectin-binding protein I: horizontal genetic transfer and gene mosaic structure[348]	2003	AJ347791.1,AJ347792.1,AJ347793.1,AJ347794.1,AJ347795.1,AJ347796.1,AJ347797.1,AJ347798.1,AJ347799.1,AJ347800.1,AJ347801.1,AJ347802.1,AJ347803.1,AJ347804.1,AJ347805.1,AJ347806.1,AJ347807.1,AJ347808.1,AJ347809.1,AJ347810.1	Streptococcal fibronectin-binding protein is involved in colonization and invasion of epithelial cells and tissues by Streptococcus pyogenes. The sfbI genes from 54 strains were sequenced. Thirty-four distinct alleles were identified. The amino-terminal aromatic amino acid-rich domain is the most variable region and is apparently generated by intergenic recombination of horizontally acquired DNA cassettes, resulting in a genetic mosaic in this region.
ska	Streptokinase	*Streptococus pyogenes*	Horizontal gene transfer and the evolution of resistance and virulence determinants in Streptococcus[57, 349]	1997	CP003116.1 CP003121.1 NC_017040.1 NC_017053.1	Streptokinase activates plasminogen to yield plasmin which is believed to enhance pathogenicity by promoting lysis of blood clots and thereby facilitate the spread of infection. Analysis of a segment of the streptokinase gene (*ska*) encoding two variable regions from 47 *Strep. pyogenes* isolates demonstrated that this gene also possesses a mosaic structure (Kapur *et al.* 1995).
skn	Streptokinase	*Streptococcus pyogenes*	Streptokinase: Molecular population genetic analysis of the streptokinase gene of Streptococcus pyogenes: mosaic alleles generated by recombination[349]	1995	U25853.1,U25854.1,U25855.1,U25856.1,U25857.1,U25858.1,U25859.1,U25860.1,U25861.1,U25862.1,U25863.1,U25864.1,U25865.1,U25866.1,U25867.1,U25868.1,U25869.1,U25870.1	Kapur et al. 1995 only abstracta available. “skn” (Kapur et al., 1995) = „ska“ (Dowson et al.,1995).

Gene	Function	Organism	Reference	Year	Accession No.	Comments
tbpB	Transferrin binding protein B	*N. sicca, N. mucosa, N. flava, N. subflava, N. cinerea, N. flavescens, N. polysaccharea*	Distribution of transferrin binding protein B gene (tbpB) variants among Neisseria species[350]	2008	AJ704730.2,AJ704731.2,AJ704732.2,AJ704733.2,AJ704734.2,AJ704735.2,AJ704736.2,AJ704737.2,AJ704738.2,AJ704739.2,AJ704740.2,AJ704746.2,AJ704748.2,AJ704761.2,AJ704762.2,AM849571.1,AM849572.1,AM849573.1,AM849574.1,AM849575.1	Transferrin binding protein B (tbpB), an outer membrane lipoprotein, is required for the acquisition of iron from human transferrin. As a result, the diversity of the tbpB gene was investigated in a defined collection of Neisseria species. A mosaic gene structure was present in the tbpB genes indicating that recombination occurred frequently among *tbpBB* genes from *N. sicca, N. flava, N. subflava, N. mucosa, N. flavescens, N. cinerea* and *N. polysaccharea* (Fig. 2a).
tbpB	Transferrin binding protein B	*Neisseria minigitidis*	Heterogeneity of tbpB, the transferrin-binding protein B gene, among serogroup B Neisseria meningitidis strains of the ET-5 complex[351]	1997	Y09617.1 Y09618.1 Y09619.1 Y09977.1	ET-5 complex strains of Neisseria meningitidis were traced intercontinentally and have been causing hyperendemic meningitis on a worldwide scale. In an attempt to develop a fully broad cross-reactive transferrin-binding protein B (TbpB)-based vaccine, the authors undertook to assess the extent of variability of TbpB proteins among strains of ET-5 complex. The paper indicates an older paper on tbpB mosaicism.
tbpBA operon	Transferrin binding protein(s)	*Mannheimia haemolytica, Mannheimia glucosida, Bibersteinia trehalosi*	Evidence for a common gene pool and frequent recombinational exchange of the tbpBA operon in Mannheimia haemolytica, Mannheimia glucosida and Bibersteinia trehalosi[352]	2011	AY850230.1,AY850231.1,AY850232.1,AY850233.1,AY850234.1,AY850235.1,AY850236.1,AY850237.1,AY850238.1,AY850239.1,AY850240.1,AY850241.1,AY850242.1,AY850243.1,AY850244.1,AY850245.1,AY850246.1,AY850247.1,AY850248.1,AY850249.1	The tbpBA operon was sequenced in 42 representative isolates of Mannheimia haemolytica (32), Mannheimia glucosida (6) and Bibersteinia trehalosi (4). The tbpBA phylogeny has been severely disrupted by numerous small- and large-scale intragenic recombination events. In addition, assortative (entire gene) recombination events, involving either the entire tbpBA operon or the individual tbpB and tbpA genes, have occurred. Our findings indicate that a common gene pool exists for tbpBA in M. haemolytica, M. glucosida and B. trehalosi. Variation on tbpBA is supposed to allow adaptation to iron acquisition in different animal hosts (shep, cattle…)
tet(M)	Resistance	*Enterococcus faecalis, Streptococcus pneumoniae, Staphylococcus aureus, Ureaplasma urealyticum, and Neisseria*	The tetracycline resistance gene tet(M) exhibits mosaic structure[270]	1996	*S. pneumoniae* Tn1545 (parent allele 1), *S. aureus* (parent allele 2) apparent in this study, initial origin discussed as unclear	Excellent basic paper on characteristics and discovery of mosaic genes. tetM is in contrast to chromosomal mosaic genes exchanged accross different genera.
tet(O)	Resistance	*Mixed fecal samples*	Mosaic tetracycline resistance	2007	AM180355.1,BA000016.3	very good graphics

Gene	Function	Organism	Reference	Year	Accession No.	Comments
tet(W)			genes are widespread in human and animal fecal samples[235]		,CM001477.1,CP000246.1 ,CP002465.1,CP002688.1, GG668518.1,GL385912.1, GL397071.1,GL397103.1, GL397128.1,GL397173.1, GL397214.1,JH594533.1, NC_003076.8,NC_003366 .1,NC_008261.1,NC_0090 89.1,NC_017618.1,NZ_C M001477.1	PCR detection system for mosaic genes nomenclature in tet genes informs about mosaicism novel isolation of a bacterium containing the original tet32gene
tet(O) tet(W)	Resistance	*Megasphaera elsdenii*	Isolation of tetracycline-resistant Megasphaera elsdenii strains with novel mosaic gene combinations of tet(O) and tet(W) from swine[273]	2003	DQ647324.2 DQ679926.1 EF065523.1 EF065524.1	Anaerobic bacteria insensitive to chlortetracycline were isolated from cecal contents and cecal tissues of swine fed or not fed chlortetracycline. Eight of 84 isolates from seven different animals were identified as Megasphaera elsdenii strains. Sequence analysis revealed that the M. elsdenii genes represent two different mosaic genes formed by interclass (double-crossover) recombination events involving tet(O) and tet(W). These findings suggest a role for commensal bacteria not only in the preservation and dissemination of antibiotic resistance in the intestinal tract but also in the evolution of resistance.
tet(O/W/32/O /W/O) tet(W/32/O) tet(O/W)	resistance	*Bifidobacterium thermophilum, Lactobacillus johnsonii*	Mosaic tetracycline resistance genes and their flanking regions in Bifidobacterium thermophilum and Lactobacillus johnsonii[237]	2008	AM889122.1 - AM889118.1 AM710605.1 - AM710601.1 DQ525023.1	For the first time, mosaic tetracycline resistance genes were identified in Lactobacillus johnsonii and in Bifidobacterium thermophilum strains. The L. johnsonii strain investigated contains a complex hybrid gene, tet(O/W/32/O/W/O), whereas the five bifidobacterial strains possess two different mosaic tet genes: i.e., tet(W/32/O) and tet(O/W). As reported by others, the crossover points of the mosaic tet gene segments were found at similar positions within the genes, suggesting a hot spot for recombination.
tet(S/M)	resistance	*Streptococcus bovis*	Molecular characterization of a novel mosaic gene encoding tetracycline resistance in foodborne strains of Streptococcus bovis[272]	2012	HM367711.1	Molecular characterization of a tetracycline (Tet)-resistance gene in 39 foodborne isolates of S. bovis. The gene was identified as a novel tet(S/M) fusion. Nucleotide sequencing of the surrounding genomic region of 16.2 kb revealed large blocks of homology with the genomes of Streptococcus

Gene	Function	Organism	Reference	Year	Accession No.	Comments
						infantarius and Lactococcus lactis.
transposon	resistance	*E. coli ESS1* *Enterobacter cloacae*	A mosaic transposon encoding OXA-48 and CTX-M-15: towards pan-resistance[353]	2013	JX423832.1	Not relevant. Higher level mosaic (transposon)
transposon		*Paracoccus marcusii*	Identification of a mosaic transposable element of Paracoccus marcusii composed of insertion sequence ISPmar4 (ISAs1 family) and an IS1247a-driven transposable module (TMo)[355]	2009	FJ422380.1 GU906878.1	Mosaic transposon containing a mosaic gene: Besides IS1247a sequences, the TMo also contains the 3'-end region of a putative alpha/beta hydrolase gene, whose expression might be activated from the P(2) promoter of IS1247a.
transposon	transposon		Tn5041-like transposons: molecular diversity, evolutionary relationships and distribution of distinct variants in environmental bacteria[354]	2002	X98999.3 Y19000.2 Y19001.1 Y18999.2 AJ318529.1	Not relevant: higher level mosaic s (transposons)
vacA	toxin	*Helicobacter pylori*	Surveillance and clinical relevance of vacA genotypes of Helicobacter pylori infecting dyspeptic patients in mid-Essex[356, 357]	2002	No sequences available	The Helicobacter pylori vacuolating cytotoxin is a putative pathogenicity factor encoded by vacA, a mosaic gene with a global distribution. The vacA type prevalence and diversity of H. pylori isolated from antral gastric biopsies of 360 dyspeptic patients in mid-Essex, and of 79 patients from other locations, were investigated Only abstract available.
vlsE	Surface protein	*Borrelia burgdorferi*	The implications of a low rate of horizontal transfer in Borrelia[358]	2001	No sequences available	Full text missing
wsp	Surface protein	*Wolbachia*	Mosaic nature of the wolbachia surface protein[359]	2005	AJ833931.1 AJ833930.1	The authors conduct a detailed analysis of the patterns of variation and recombination within the Wolbachia surface protein (wsp), utilizing an extensive set of published and new sequences from five main supergroups of Wolbachia. Comparison of shared polymorphisms reveals a complex mosaic structure of the gene. Exchanges occurred both within and between the arthropod supergroups.

8 Annex 2 – Transformation, Competence, Hotspots of Recombination

Tables 1, 2 and 3 (Literature analysis A + B + C) are part of a single Excel database.

Literature analysis A: Mosaic gene formation

Table 29: Literature analysis A: Mosaic gene formation

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Alfano, J.R.	The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants	PNAS	2000	Hrp pathogenicity island (PAI)	AF232003, AF232004, AF232005, AF232006 (it doesn't include the complete PAI)	chromosome	3	pathogenicity	type III protein secretion system that appears to translocate Avr and Hop effector proteins into plant cells
Ameyama, S.	Mosaic-like structure of penicillin-binding protein 2 gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime	Antimicrobial Agents and Chemotherapy	2002	penA	AB071984	chromosome	-	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Atherton, J.C.	Vacuolating Cytotoxin (vacA) Alleles of Helicobacter pylori Comprise Two Geographically Widespread Types, m1 and m2, and Have Evolved Through Limited Recombination	Current Microbiology	1999	vacA	-	-	2	virulence	vacuolating citotoxin
Aziz, R.K.	Mosaic Prophages with Horizontally Acquired Genes Account for the Emergence and Diversification of the Globally Disseminated M1T1 Clone ofStreptococcus pyogenes	Journal of Bacteriology	2005	prophage M1T1.X, M1T1.Y, M1T1.Z	-	prophage	-	virulence	-

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Baldo, L.	Mosaic Nature of the Wolbachia Surface Protein	Journal of Bacteriology	2005	wsp	complete genomes of Wolbachia: AE017259, AY095156, AY331986, AY331128, AY330308, AY177735, AF521151, AY101200, AF521165, AF020070, AF217725, AF481165, AB109581, AF237885, AF071910, AY331114, AY566421, AB109614, AJ130716, AB109580, AF521156, AB077201, AF481164, AY552552, AB094372, AY188687, AJ271198, AJ276608, AJ276605, AJ276598, AF521152, AJ276615, AB024571, AY330316, AJ252062, AJ25218, AJ252180, AJ252061, AJ833930, AJ833931	chromosome	several	virulence	Surface antigen; function unknown, but strong suggestion of involvement in the interaction with the host cytoplasm
Barile, S.	Molecular characterization of a novel mosaic tet(S/M) gene encoding tetracycline resistance in foodborne strains of Streptococcus bovis	Microbiology	2012	tet S/M	HM367711	chromosome	2	resistance	protein synthesis inhibitor
Blackman Northwood, J.	Characterization of Macrolide Efflux PumpmefSubclasses Detected in Clinical Isolates ofStreptococcus pyogenesIsolated between 1999 and 2005	Antimicrobial Agents and Chemotherapy	2009	mef(A/E) variant 1, variant 2 and variant 3	EU870856-8	-	3	resistance	macrolide efflux pump
Boc, A.	Towards an accurate identification of mosaic genes and partial horizontal gene transfers	Nucleic Acids Research	2011						

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Bourkal'tseva, M.V.	Bacteriophage phi297, a new species of *Pseudomonas aeruginosa* temperate phages with a mosaic genome: potencial use in phage therapy	Russian Journal of Genetics	2011	phi297 bacteriophage	HQ711984	bacteriophage	several	halo-forming temperate phage	degradation of bacterial exopolysaccharides
Bowler, L.D.	Interspecies recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicilin resistance in N. meningitidis: natural events and laboratory simulation	Journal of Bacteriology	1994	penA	-	chromosome	2	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Bruckner, R.	Mosaic genes and mosaic chromosomes–genomic variation in Streptococcus pneumoniae	International Journal of Medical Microbiology	2004						
Bull, P.C.	Plasmodium falciparumantigenic variation. Mapping mosaic vargene sequences onto a network of shared, highly polymorphic sequence blocks	Molecular Microbiology	2008	var				virulence	
Canchaya, C.	Mosaic-Like Sequences Containing Transposon, Phage, and Plasmid Elements amongListeria monocytogenesPlasmids	Applied and Environmental Microbiology	2010	plasmid pLM33	GU244485	plasmid	-	-	-

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Chen, C.C.	Tracing subsequent dissemination of a cluster of gonococcal infections caused by an ST1407-related clone harbouring mosaic penA alleles in Taiwan	Journal of Antimicrobial Chemotherapy	2013	penA-XXXIV	-	chromosome	-	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Costa, J.	Molecular evolution ofLegionella pneumophila dotA gene, the contribution of natural environmental strains	Environmental Microbiology	2010	dotA	FN652504 to FN652664; FN658680-FN658689	chromosome	3 to 7	virulence	integral cytoplasmic membrane protein; invasion of the eukaryotic cell
Cuvelier, M.L.	Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton	PNAS	2010						
Del Grosso, M.	Genetic Resistance Elements CarryingmefSubclasses Other than mef(A) inStreptococcus pyogenes	Antimicrobial Agents and Chemotherapy	2011	mef(A/E)	JF501521	prophage ϕm46.1-like	2	resistance	efflux-mediated erythromycin resistance
Diep, D.B.	An overview of the mosaic bacteriocinplnloci from Lactobacillus plantarum	Peptides	2009	pln loci					

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Du Plessis, M.	Analysis of Penicillin-Binding Protein 1b and 2a Genes from Streptococcus pneumoniae	Microbial Drug Resistance	2000						
Eda, M.	Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses)	Genes and Genetic Systems	2010						
Fecskeova, L.	Mosaic structure of the small cryptic plasmid pKST23 from Escherichia coli	Folia Microbiologica	2012	plasmid pKST23	JN253636	plasmid	2	-	-
Filipe, S.R.	Distribution of the Mosaic Structured murM Genes among Natural Populations of Streptococcus pneumoniae	Journal of Bacteriology	2000	murM	AF281135, AF281136, AF281137, AF281139, AF281140	chromosome	4 t o7	cell wall biogenesis	biosynthesis of the branched structured muropeptide components; expression of β-lactamase resistance
Fudyk, T.C.	Genetic Diversity and Mosaicism at the por Locus of Neisseria gonorrhoeae	Journal of Bacteriology	1999	por IA and IB	AF090794, AF090796, AF090801, AF090802, AF090800, AF090803, AF090793, AF090797, AF090804, AF090795, AF090805, AF090798, AF090806, AF090792, AF090799, AF090807, AF090822, AF090823, AF090824, AF090808, AF090811, AF090813, AF090816, AF090817, AF090818, AF090821, AF090819, AF090810, AF090809, AF090812, AF090814, AF090820, AF090815	chromosome	-	cell wall biogenesis	porin protein; major outer membrane protein

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Geysen, D.	An unusual mosaic structure of the PIM gene of Theileria parvaand its relationship to allelic diversity	Molecular and Biochemical Parasitology	2004	PIM gene					polymorphic immunodominant molecule; intrinsic membrane protein
Groisman, E.A.	Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome	The EMBO Journal	1992	phoN region	X63599	chromosome	-	-	-
Gutierrez, M.C.	Ancient Origin and Gene Mosaicism of the Progenitor ofMycobacterium tuberculosis	Plos Pathogens	2005	gyrB, gyrA, hsp65, rpoB, katG, sodA	AJ749904 to AJ749948 (genes from smooth tubercle bacilli)	chromosome	2 for gyrB and gyrA from M. prototuberculosis; not defined for the others	housekeeping	
Hakenbeck, R.	Mosaic genes and their role in penicillin-resistant Streptococcus pneumoniae	Electrophoresis	1998						
Hakenbeck, R.	Transformation in Streptococcus pneumoniae: mosaic genes and the regulation of competence	Research in Microbiology	2000						

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Hakenbeck, R.	Mosaic Genes and Mosaic Chromosomes: Intra- and Interspecies Genomic Variation ofStreptococcus pneumoniae	Infection and Immunity	2001	pbp2x and/or dhfr	-	chromosome	2 to 11 for pbp2x	resistance	penicillin binding protein 2x, bacterial cell wall synthesis (pbp2x); dihydrofolate reductase (dhfr)
Hao, W.	Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion	PNAS	2010						
Hao, W.	Unrecognized fine-scale recombination can mimic the effects of adaptive radiation	Gene	2013	oprF	(https://sites.google.com/site/haowlab/home/oprf)	chromosome	-	cell wall biogenesis	outer membrane protein, porin; environmental sensor
Havarstein, L.S.	Natural Competence in the GenusStreptococcus: Evidence that Streptococci Can Change Pherotype by Interspecies Recombinational Exchanges	Journal of Bacteriology	1997	comCD	AJ000867, AJ000868, AJ000870	chromosome	3	competence	competence regulation; competence-stimulating peptide precusrsor (comC), histidine kinase (comD)
Herman, A.	Identification of Multiple Gephyrin Variants in Different Organs of the Adult Rat	Biochemical and Biophysical Research Communications	2001	gephyrin gene					

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Heymans, R.	Clonally RelatedNeisseria gonorrhoeaeIsolates with Decreased Susceptibility to the Extended-Spectrum Cephalosporin Cefotaxime in Amsterdam, the Netherlands	Antimicrobial Agents and Chemotherapy	2012	penA-XXXIV	-	chromosome	-	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Hollingshead, S.K.	Diversity of PspA: Mosaic Genes and Evidence for Past Recombination inStreptococcus pneumoniae	Infection and Immunity	2000	pspA	AF071802 to AF071827	chromosome	-	virulence	pneumococcal surface protein A; antibody mediated protection
Huang, C.T.	Characteristics and Dissemination of Mosaic Penicillin-Binding Protein 2-Harboring Multidrug-ResistantNeisseria gonorrhoeae Isolates with Reduced Cephalosporin Susceptibility in Northern Taiwan	Antimicrobial Agents and Chemotherapy	2010	penA	-	chromosome	-	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Kwon, D.H.	High-Level -Lactam Resistance Associated with Acquired Multidrug Resistance inHelicobacter pylori	Antimicrobial Agents and Chemotherapy	2003	pbp-1A	AY241260	chromosome	2	resistance	penicillin binding protein 1A bacterial cell wall synthesis
Lujan, R.	Penicillin-Resistant Isolates of Neisseria lactamica Produce Altered Forms of Penicillin-Binding Protein 2 That Arose by Interspecies Horizontal Gene Transfer	Antimicrobial Agents and Chemotherapy	1991	penA	-	chromosome	3	resistance	penicillin binding protein 2 bacterial cell wall synthesis

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
McGraw, E.A.	Molecular Evolution and Mosaic Structure of α, β, and γ Intimins of Pathogenic Escherichia coli	Molecular Biology and Evolution	1999	eae	AF081182-7	chromosome (in a pathogenicity island)	5	virulence	outer membrane protein; intimate attachment of the bacterial cell to eukayotic cells
Novais, C.	A tet(S/M) hybrid from CTn6000 and CTn916 recombination	Microbiology	2012	tet S/M	AY534326	-	2	resistance	protein synthesis inhibitor
Ohnishi, M.	Spread of a Chromosomal Cefixime-Resistant penA Gene among Different Neisseria gonorrhoeae Lineages	Antimicrobial Agents and Chemotherapy	2010	penA-X, penA-XXVI, penA-XXX, penA-XXXI and penA-XXXII	AB511942-4	chromosome	-	resistance	penicillin binding protein 2 bacterial cell wall synthesis
Radstrom, P.	Transformational Exchanges in the Dihydropteroate Synthase Gene of Neisseria meningitidis: a Novel Mechanism for Acquisition of Sulfonamide Resistance	Journal of Bacteriology	1992	dhps	X68068	chromosome	3	resistance	dihydropteroate synthase
Reichmann, P.	A Global Gene Pool for High-Level Cephalosporin Resistance in Commensal Streptococcus Species and Streptococcus pneumoniae	The Journal of Infectious Diseases	1997	pbp2x, pbp2b, pbp1a	Y10532-6	chromosome	2 to 8	resistance	penicillin binding protein 1 or 2 bacterial cell wall synthesis

1. Author	Title	Journal	Year	Mosaic Gene	Mosaic Gene Accession No. (GenBank)	Localisation of Genetic Element	Mosaic Gene Number of gene fragments	Mosaic Gene Protein / Function (general)	Mosaic Gene Protein / Function (details)
Takeda, M.	Characterization of the Neurotoxin Produced by Isolates Associated with Avian Botulism	Avian Diseases	2005	BoNT C/D	-	-	2	toxin	botulinum neurotoxin; single chain peptide activated by cleavage into light and heavy chain connected by a disulfide bond L-chain: Zn-finger protease H-chain: internalisation
Vidal, D.	Environmental factors influencing the prevalence of a Clostridium botulinum type C/D mosaic strain in nonpermanent Mediterranean wetlands	Applied and Environmental Microbiology	2013	BoNT C/D	-	-	2	toxin	-
Lindberg, A.	Real-time PCR for Clostridium botulinum type C neurotoxin (BoNTC) gene, also covering a chimeric C/D sequence--application on outbreaks of botulism in poultry	Veterinary Microbiology	2010	BoNT C/D	FN436021 FN436022	phage	2	toxin	-
Potron, A.	A mosaic transposon encoding OXA-48 and CTX-M-15: towards pan-resistance	Journal of Antimicrobial Chemotherapy	2013	mosaic transposon; combination of OXA-48 with CTX-15-M	JX423832	transposon	-	resistance	ß-lactamases
Unemo, M.	High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure	Antimicrobial Agents & Chemotherapy	2012	penA-CI, penA-XXXIV derivate	JQ073701	chromosome		resistance	penicillin binding protein 2 bacterial cell wall synthesis

Literature analysis B: Host strain characteristics

Table 30: Literature analysis B: Host strain characteristics

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Alfano, J.R.	The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants	PNAS	2000	*Pseudomonas syringae pv. syringae, Pseudomonas syringae pv. tomato*	-	fruits and vegetables's surfaces	wide pool of pathogenic bacteria	necrosis	plants	-
Ameyama, S.	Mosaic-like structure of penicillin-binding protein 2 gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime	Antimicrobial Agents and Chemotherapy	2002	*Neisseria gonorrhoeae*	reduced susceptibility to cefixime; natural competent	male urethra	Neisseria perflava, Neisseria cinerea, Neisseria flavescens, Neisseria meningitidis	gonorrhea	humans	-
Atherton, J.C.	Vacuolating Cytotoxin (vacA) Alleles of Helicobacter pylori Comprise Two Geographically Widespread Types, m1 and m2, and Have Evolved Through Limited Recombination	Current Microbiology	1999	*Helicobacter pylori*	natural competent	human stomach	Helicobacter pylori strains	human peptic ulceration	humans	major risk factor for the development of distal gastric adenocarcinoma and gastric lymphoma
Aziz, R.K.	Mosaic Prophages with Horizontally Acquired Genes Account for the Emergence and Diversification of the Globally Disseminated M1T1 Clone ofStreptococcus pyogenes	Journal of Bacteriology	2005	*Streptococcus pyogenes M1T1*	-	-	-	invasive and noninvasive infections (pharyngitis, rheumatic heart disease, streptococcal toxic shock syndrome, necrotizing-fasciitis)	humans	-
Baldo, L.	Mosaic Nature of the Wolbachia Surface Protein	Journal of Bacteriology	2005	*Wolbachia*	-	arthropodes and nematodes	Wolbachia genus	reproduction manipulation	arthropodes and nematodes	-

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Barile, S.	Molecular characterization of a novel mosaic tet(S/M) gene encoding tetracycline resistance in foodborne strains of Streptococcus bovis	Microbiology	2012	*Streptococcus bovis*	tetracycline resistance; natural competent	gastrointestinal tract of humans and animals; environment, fermented foods	-	-	humans and animals	opportunistic pathogen in immune-compromised individuals
Blackman Northwood, J.	Characterization of Macrolide Efflux PumpmefSubclasses Detected in Clinical Isolates ofStreptococcus pyogenesIsolated between 1999 and 2005	Antimicrobial Agents and Chemotherapy	2009	*Streptococcus pyogenes ST39, ST421, ST422, ST423*	macrolide resistance	-	-	respiratory tract infections	humans	-
Boc, A.	Towards an accurate identification of mosaic genes and partial horizontal gene transfers	Nucleic Acids Research	2011							
Bourkal'tseva, M.V.	Bacteriophage phi297, a new species of *Pseudomonas aeruginosa* temperate phages with a mosaic genome: potencial use in phage therapy	Russian Journal of Genetics	2011	*Pseudomonas aeruginosa*	-	-	phage D3 and F116	-	-	-
Bowler, L.D.	Interspecies recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicilin resistance in N. meningitidis: natural events and laboratory simulation	Journal of Bacteriology	1994	*Neisseria meningitidis*	penicillin resistance; natural competent	human naso- and oropharynx	Neisseria cinerea (in one isolate) and Neisseria flavescens (in another isolate)	meningitis	humans	highly clinically relevant
Bruckner, R.	Mosaic genes and mosaic chromosomes–genomic variation in Streptococcus pneumoniae	International Journal of Medical Microbiology	2004							
Bull, P.C.	Plasmodium falciparumantigenic variation. Mapping mosaic vargene sequences onto a network of shared, highly polymorphic sequence blocks	Molecular Microbiology	2008							

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Canchaya, C.	Mosaic-Like Sequences Containing Transposon, Phage, and Plasmid Elements amongListeria monocytogenesPlasmids	Applied and Environmental Microbiology	2010	*Listeria monocytogenes*	-	cheese	-	gastroenteritis, meningitis, encephalitis, materno-fetal and perinatal infections	humans	-
Chen, C.C.	Tracing subsequent dissemination of a cluster of gonococcal infections caused by an ST1407-related clone harbouring mosaic penA alleles in Taiwan	Journal of Antimicrobial Chemotherapy	2013	*Neisseria gonorrhoeae*	resistance to penicillin G (83%), ciprofloxacin (100%), cefpodoxime (26%), cefixime (13%) and ceftriaxone (4%); natural competent	-	-	gonorrhea	human males	-
Costa, J.	Molecular evolution ofLegionella pneumophila dotA gene, the contribution of natural environmental strains	Environmental Microbiology	2010	*Legionella pneumophila*	natural competent	fresh water environments; man-made acquatic environments; human lungs	different ssubspecies/strains of Legionella pneumophila	Legionnaires' disease (severe pneumonia)	humans	-
Cuvelier, M.L.	Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton	PNAS	2010							
Del Grosso, M.	Genetic Resistance Elements CarryingmefSubclasses Other than mef(A) inStreptococcus pyogenes	Antimicrobial Agents and Chemotherapy	2011	*Streptococcus pyogenes*	resistance to erythromycin and tetracycline	-		-	-	-
Diep, D.B.	An overview of the mosaic bacteriocinplnloci from Lactobacillus plantarum	Peptides	2009							
Du Plessis, M.	Analysis of Penicillin-Binding	Microbial Drug	2000							

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
	Protein 1b and 2a Genes from Streptococcus pneumoniae	Resistance								
Eda, M.	Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses)	Genes and Genetic Systems	2010							
Fecskeova, L.	Mosaic structure of the small cryptic plasmid pKST23 from Escherichia coli	Folia Microbiologica	2012	*Escherichia coli*	-	rumen of sheep	-	-	sheep	-
Filipe, S.R.	Distribution of the Mosaic Structured murM Genes among Natural Populations of Streptococcus pneumoniae	Journal of Bacteriology	2000	*Streptococcus pneumoniae*	resistance to penicillin; natural competent	-	nonpneumococcal species	-	-	-
Fudyk, T.C.	Genetic Diversity and Mosaicism at the por Locus of Neisseria gonorrhoeae	Journal of Bacteriology	1999	*Neisseria gonorrhoeae*	natural competent	women cervix	Neisseria gonorrhoeae strains	-	humans	
Geysen, D.	An unusual mosaic structure of the PIM gene of Theileria parvaand its relationship to allelic diversity	Molecular and Biochemical Parasitology	2004							
Groisman, E.A.	Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome	The EMBO Journal	1992	*Salmonella typhimurium*	-	-	-	salmonelosis	humans	-
Gutierrez, M.C.	Ancient Origin and Gene Mosaicism of the Progenitor ofMycobacterium tuberculosis	Plos Pathogens	2005	*Mycobacterium prototuberculosis and Mycobacterium tuberculosis complex (MTBC)*	-	human lungs, lymph	different groups of smooth tubercle bacilli (for mosaics in MTBC)	tuberculosis	mammals	
Hakenbeck, R.	Mosaic genes and their role in penicillin-resistant Streptococcus pneumoniae	Electrophoresis	1998							
Hakenbeck, R.	Transformation in Streptococcus pneumoniae: mosaic genes and the regulation of competence	Research in Microbiology	2000							

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Hakenbeck, R.	Mosaic Genes and Mosaic Chromosomes: Intra- and Interspecies Genomic Variation ofStreptococcus pneumoniae	Infection and Immunity	2001	*Streptococcus pneumoniae*	penicillin resistance and/or trimethoprim resistance; natural competent	-	-	otitis media, sinusitis, pneumonia, septicemia, meningitis	humans	-
Hao, W.	Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion	PNAS	2010							
Hao, W.	Unrecognized fine-scale recombination can mimic the effects of adaptive radiation	Gene	2013	*Pseudomonas spp.*	-	-	-	-	-	-
Havarstein, L.S.	Natural Competence in the GenusStreptococcus: Evidence that Streptococci Can Change Pherotype by Interspecies Recombinational Exchanges	Journal of Bacteriology	1997	*Streptococcus constellatus, Streptococcus milleri, Streptococcus gordonii*	natural competent	-	-	-	-	-
Herman, A.	Identification of Multiple Gephyrin Variants in Different Organs of the Adult Rat	Biochemical and Biophysical Research Communications	2001							
Heymans, R.	Clonally RelatedNeisseria gonorrhoeaeIsolates with Decreased Susceptibility to the Extended-Spectrum Cephalosporin Cefotaxime in Amsterdam, the Netherlands	Antimicrobial Agents and Chemotherapy	2012	*Neisseria gonorrhoeae*	reduced suscpetibility/resistance to cefotaxime; resistance to penicillin G, tetracycline and ciprofloxacin; natural competent	human urethra, cervix, rectum and pharynx	-	gonorrhea	humans	
Hollingshead, S.K.	Diversity of PspA: Mosaic Genes and Evidence for Past Recombination inStreptococcus pneumoniae	Infection and Immunity	2000	*Streptococcus pneumoniae*	natural competent	human nasopharynx	unknown	pneumoniae and meningitis	humans	reponsible for worldwide deaths of children

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Huang, C.T.	Characteristics and Dissemination of Mosaic Penicillin-Binding Protein 2-Harboring Multidrug-ResistantNeisseria gonorrhoeae Isolates with Reduced Cephalosporin Susceptibility in Northern Taiwan	Antimicrobial Agents and Chemotherapy	2010	*Neisseria gonorrhoeae*	reduced susceptibility to cefixime, ceftriaxone (16%); resistance to penicillin (71%) and ciprofloxacin (100%); natural competent	-	-	gonorrhea	humans	one of the most globally prevalent sexually transmitted infections
Kwon, D.H.	High-Level -Lactam Resistance Associated with Acquired Multidrug Resistance inHelicobacter pylori	Antimicrobial Agents and Chemotherapy	2003	*Helicobacter pylori*	resistance to β-lactams; low- to moderate-level resistance to ciprofloxacin, chloramphenicol, metronidazole, rifampin and tetracycline. Naturally competent	human gastric tract	-	gastric infections	humans	infections are difficult to treat
Lujan, R.	Penicillin-Resistant Isolates of Neisseria lactamica Produce Altered Forms of Penicillin-Binding Protein 2 That Arose by Interspecies Horizontal Gene Transfer	Antimicrobial Agents and Chemotherapy	1991	*Neisseria lactamica*	resistance to penicillin; natural competent	human nasopharynx	Neisseria flavescens	commensal	humans	-
McGraw, E.A.	Molecular Evolution and Mosaic Structure of α, β, and γ Intimins of Pathogenic Escherichia coli	Molecular Biology and Evolution	1999	*Escherichia coli*	-	human gastrointestinal tract	-	diarrhea, dysentery	humans	-
Novais, C.	A tet(S/M) hybrid from CTn6000 and CTn916 recombination	Microbiology	2012	*Streptococcus intermedius*	natural competent	-	-	-	-	-

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
Ohnishi, M.	Spread of a Chromosomal Cefixime-Resistant penA Gene among Different Neisseria gonorrhoeae Lineages	Antimicrobial Agents and Chemotherapy	2010	*Neisseria gonorrhoeae ST7363, ST1596, ST1588, ST1590, ST7358, ST1901*	reduced susceptibility to cefixime; natural competent	human urogenital tract	-	gonorrhea	humans	one of the most common sexually transmissiblee infective disease
Radstrom, P.	Transformational Exchanges in the Dihydropteroate Synthase Gene of Neisseria meningitidis: a Novel Mechanism for Acquisition of Sulfonamide Resistance	Journal of Bacteriology	1992	*Neisseria meningitidis*	resistance to sulfonamides; natural competent	-	(homologous segments in Neisseria gonorrhoeae and Neisseria lactamica)	-	-	-
Reichmann, P.	A Global Gene Pool for High-Level Cephalosporin Resistance in Commensal Streptococcus Species and Streptococcus pneumoniae	The Journal of Infectious Diseases	1997	*Streptococcus mitis, S. oralis, S. pneumoniae*	penicillin and cephalosporin-resistant; natural competent	-	-	-	-	-
Takeda, M.	Characterization of the Neurotoxin Produced by Isolates Associated with Avian Botulism	Avian Diseases	2005	*Clostridium botulinum*	gram-positive, spore-forming, obligatory anaerobic	ubiquitous, rivers, lakes, soil, animal intestines		Type C botulism; majority of animal botulism	birds (ducks, waterfowls, broiler chicken, etc…)	-
Vidal, D.	Environmental factors influencing the prevalence of a Clostridium botulinum type C/D mosaic strain in nonpermanent Mediterranean wetlands	Applied and Environmental Microbiology	2013	*Clostridium botulinum*	-	wetlands; low Cl-, high organic content of sediments; July temperatures above 26°C, temperatures above digestive tract of birds		Type C botulism	birds (ducks, waterfowls, broiler chicken, etc…); vectors: Chironomidae, Corixidae,Sarcophagidae, Calliphoridae	most important avian disase (mortality)
Lindberg, A.	Real-time PCR for Clostridium botulinum type C neurotoxin (BoNTC) gene, also covering a chimeric C/D sequence--application	Veterinary Microbiology	2010	*Clostridium botulinum*	-	soil, aquatic environments		Type C botulism	poultry flocks, water fowl	-

1. Author	Title	Journal	Year	Bacterial Host	Phenotype of Host Strain (including info on competence)	Habitat of Host Strain	Putative donor of mosaic fragments	Disease	Affected organisms	Relevance
	on outbreaks of botulism in poultry									
Potron, A.	A mosaic transposon encoding OXA-48 and CTX-M-15: towards pan-resistance	Journal of Antimicrobial Chemotherapy	2013	*Enterobacteriaceae*	resistance to penicillins, penicillin/b-lactamase inhibitor combinations and broadspectrum cephalosporins, decreased susceptibility to carbapenems	mammalian gastrointestinal tract		-	humans	highly clinical relevant
Unemo, M.	High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure	Antimicrobial Agents & Chemotherapy	2012	*Neisseria gonorrhoeae ST1091*	high level resistance to extended spectrum cepaholsprins (ceftriaxon, cefixime)	human urogenital tract		gonorrhea	humans	second most prevalent bacterial sexually transmitted disease worldwide

Literature analysis C: Gene transfer/transformation/recombination

Table 31: Literature analysis C: Gene transfer/transformation/recombination

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Alfano, J.R.	The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants	PNAS	2000	The Hrp Pai has evidence of horizontal acquisition. Genes in the EEL part (Exchangeable Effector Locus) has fragments of mobile genetic elements such as phages, insertion sequences and plasmids. This part (EEL) is unstable and might be continually changing by acquisition and loss of sequences realted to mobile genetic elements.	-	-		not exactly a mosaic gene, but a mobile genetic element with a tripartite mosaic structure
Ameyama, S.	Mosaic-like structure of penicillin-binding protein 2 gene (penA) in clinical isolates of Neisseria gonorrhoeae with reduced susceptibility to cefixime	Antimicrobial Agents and Chemotherapy	2002	natural transformation (experimentally shown)	transfer of the penA mosaic-like gene (as PCR product) from N. gonorrhoeae NG-3 to N. gonorrhoeae ATCC 19424	homologous recombination		The formation of the mosaic gene is not experimentally shown, but natural transforamtion is suggested to be responsible
Atherton, J.C.	Vacuolating Cytotoxin (vacA) Alleles of Helicobacter pylori Comprise Two Geographically Widespread Types, m1 and m2, and Have Evolved Through Limited Recombination	Current Microbiology	1999	-	-	sequence analysis suggests recombination between m1 and m2, two m1 or two m2 alleles led to the formation of mosaic genes		
Aziz, R.K.	Mosaic Prophages with Horizontally Acquired Genes Account for the Emergence and Diversification of the Globally Disseminated M1T1 Clone ofStreptococcus pyogenes	Journal of Bacteriology	2005	-	-	-		not exactly a mosaic gene, but a mobile genetic element with a mosaic structure
Baldo, L.	Mosaic Nature of the Wolbachia Surface Protein	Journal of Bacteriology	2005	horizontal gene transfer is suggested to be involved	-	intergenic and intragenic (recombination between motifs within the HVRs (hypervariable regions)) recombination is		

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
						suggested		
Barile, S.	Molecular characterization of a novel mosaic tet(S/M) gene encoding tetracycline resistance in foodborne strains of Streptococcus bovis	Microbiology	2012	the tetS/M is inserted in a region flanked by two ISs sites and a IS1216 transposase, suggestive of the transposase-mediated circularization/excision of the region . This 4.3 Kb region is embedded into alarger region with 16 Kb, containing several ISs, suggestive of a mobilizable region	-	-	Patterson, 2007 ; van Hoek, 2008	
Blackman Northwood, J.	Characterization of Macrolide Efflux PumpmefSubclasses Detected in Clinical Isolates ofStreptococcus pyogenesIsolated between 1999 and 2005	Antimicrobial Agents and Chemotherapy	2009	-	-	-		First description of the mefA/E mosaic
Boc, A.	Towards an accurate identification of mosaic genes and partial horizontal gene transfers	Nucleic Acids Research	2011				Hollingshead, 2000	This paper is related with the development of a method for detecting partial HGT events and related intragenic recombination giving rise to the formation of mosaic genes
Bourkal'tseva, M.V.	Bacteriophage phi297, a new species of *Pseudomonas aeruginosa* temperate phages with a mosaic genome: potencial use in phage therapy	Russian Journal of Genetics	2011	-	-	-		not exactly a mosaic gene, but a mobile genetic element with a mosaic structure

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Bowler, L.D.	Interspecies recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicilin resistance in N. meningitidis: natural events and laboratory simulation	Journal of Bacteriology	1994	natural transformation performed	transformation of the penicillin-susceptible N. meningitidis C311 with penicillin-resistant N. meningitidis K196, N. cinera LNP2060 and N. flavescens NCTC8263. Transfer of the mosaic penA gene was confirmed in all transformants obtained with N. meningitidis K196, 50% with N. flavescens NCTC8263, and 0% with N. cinera LNP2060.	-		
Bruckner, R.	Mosaic genes and mosaic chromosomes–genomic variation in Streptococcus pneumoniae	International Journal of Medical Microbiology	2004					Review paper. Not about a mosaic gene, but gene clusters
Bull, P.C.	Plasmodium falciparumantigenic variation. Mapping mosaic vargene sequences onto a network of shared, highly polymorphic sequence blocks	Molecular Microbiology	2008					The mosaic gene is present in a protozoan parasite, Plasmodium falciparum
Canchaya, C.	Mosaic-Like Sequences Containing Transposon, Phage, and Plasmid Elements amongListeria monocytogenesPlasmids	Applied and Environmental Microbiology	2010	parts of the plasmid are suggested to have been acquired by HGT. The plasmid contains genes encoding proteins involved in genetic mobilization, such as transposases and integrases	-	-		not exactly a mosaic gene, but a mobile genetic element with a mosaic structure
Chen, C.C.	Tracing subsequent dissemination of a cluster of gonococcal infections caused by an ST1407-related clone harbouring mosaic penA alleles in Taiwan	Journal of Antimicrobial Chemotherapy	2013	-	-	-	Pandori 2009	
Costa, J.	Molecular evolution ofLegionella pneumophila dotA gene, the contribution of natural environmental strains	Environmental Microbiology	2010	-	-	-	Baldo, 2005	

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Cuvelier, M.L.	Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton	PNAS	2010					This paper is related with a mosaic gene in eukaryotic phytoplankton cells
Del Grosso, M.	Genetic Resistance Elements CarryingmefSubclasses Other than mef(A) inStreptococcus pyogenes	Antimicrobial Agents and Chemotherapy	2011	ϕm46.1 is transferable by conjugation	-	the mosaic gene seems to be the result of recombination events with the Tn1207.1	Blackman Northwood, 2009	
Diep, D.B.	An overview of the mosaic bacteriocinplnloci from Lactobacillus plantarum	Peptides	2009					Review paper. Not about a mosaic gene, but loci
Du Plessis, M.	Analysis of Penicillin-Binding Protein 1b and 2a Genes from Streptococcus pneumoniae	Microbial Drug Resistance	2000					This reference should be excluded: the analysed genes pbp1b and 2a do not display a mosaic gene pattern
Eda, M.	Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses)	Genes and Genetic Systems	2010					This paper is related with a mosaic gene in mitochondrial DNA present in eukaryotic cells from albatrosses
Fecskeova, L.	Mosaic structure of the small cryptic plasmid pKST23 from Escherichia coli	Folia Microbiologica	2012	-	-	-		not exactly a mosaic gene, but a mobile genetic element (plasmid) with a mosaic structure
Filipe, S.R.	Distribution of the Mosaic Structured murM Genes among Natural Populations of Streptococcus pneumoniae	Journal of Bacteriology	2000	-	-	-		
Fudyk, T.C.	Genetic Diversity and Mosaicism at the por Locus of Neisseria gonorrhoeae	Journal of Bacteriology	1999	acquisition of the por gene by transformation during coinfection by different gonococcal strains is hypothesized	-	localized recombination	Smith, 1991	
Geysen, D.	An unusual mosaic structure of the PIM gene of Theileria parvaand its	Molecular and Biochemical Parasitology	2004					The mosaic gene is present in a protozoan parasite, Theileria parva

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
	relationship to allelic diversity							
Groisman, E.A.	Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome	The EMBO Journal	1992	acquisition of the region containing phoN is suggested to have been acquired by conjugation	not involved	-		Not a mosaic gene. A region, containing the phoN gene, of the *Salmonella* chromosome has a mosaic structure
Gutierrez, M.C.	Ancient Origin and Gene Mosaicism of the Progenitor ofMycobacterium tuberculosis	Plos Pathogens	2005	Multiple remote HGT events between different groups of smooth tubercle bacilli and MTBC likely resulted in mosaicism in MTBC	-	intragenic recomination (mosaics in gyrA and gyrB of M. prototuberculosis)		
Hakenbeck, R.	Mosaic genes and their role in penicillin-resistant Streptococcus pneumoniae	Electrophoresis	1998				Reichmann, 1997; Oggioni, 1996	Review paper
Hakenbeck, R.	Transformation in Streptococcus pneumoniae: mosaic genes and the regulation of competence	Research in Microbiology	2000				Haverstein, 1997	Review paper
Hakenbeck, R.	Mosaic Genes and Mosaic Chromosomes: Intra- and Interspecies Genomic Variation ofStreptococcus pneumoniae	Infection and Immunity	2001	the mosaic genes arose by HGT	-	-	Haverstein, 1997; Oggioni, 1996	
Hao, W.	Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion	PNAS	2010					This paper is related with a mosaic gene in mitochondrial DNA
Hao, W.	Unrecognized fine-scale recombination can mimic the effects of adaptive radiation	Gene	2013	the sequence of the oprF genes shows evidence of HGT	-	intragenic recombination	Baldo, 2005	
Havarstein, L.S.	Natural Competence in the GenusStreptococcus: Evidence that Streptococci Can Change Pherotype by Interspecies Recombinational Exchanges	Journal of Bacteriology	1997	interspecies recombinational exchanges by natural transformation are suggested	-	-		
Herman, A.	Identification of Multiple Gephyrin Variants in Different Organs of the Adult Rat	Biochemical and Biophysical Research Communications	2001					This paper is related with a mosaic gene in eukaryotic cells, from a rat
Heymans, R.	Clonally RelatedNeisseria	Antimicrobial	2012	-	-	-	Ohnishi,	

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
	gonorrhoeaeIsolates with Decreased Susceptibility to the Extended-Spectrum Cephalosporin Cefotaxime in Amsterdam, the Netherlands	Agents and Chemotherapy					2011	
Hollingshead, S.K.	Diversity of PspA: Mosaic Genes and Evidence for Past Recombination inStreptococcus pneumoniae	Infection and Immunity	2000	intraspecies horizontal transfer is suggested to be involved	-	intragenic recomination	Atherton 1999; Smith, 1991; McGraw 1999; Reichmann 1997	
Huang, C.T.	Characteristics and Dissemination of Mosaic Penicillin-Binding Protein 2-Harboring Multidrug-ResistantNeisseria gonorrhoeae Isolates with Reduced Cephalosporin Susceptibility in Northern Taiwan	Antimicrobial Agents and Chemotherapy	2010	-	-	-	Ameyama, 2002; Ito, 2005; Ohnishi, 2010	
Kwon, D.H.	High-Level -Lactam Resistance Associated with Acquired Multidrug Resistance inHelicobacter pylori	Antimicrobial Agents and Chemotherapy	2003	natural transformation performed	transfer of the mosaic block present in the pbp-1A gene during tranformation of susceptible H. pylori ATCC 700392 and ATCC 43629 with naked DNA from resistant H. pylori IH-1 (TF = 10^{-4} to 10^{-3}).	-		
Lujan, R.	Penicillin-Resistant Isolates of Neisseria lactamica Produce Altered Forms of Penicillin-Binding Protein 2 That Arose by Interspecies Horizontal Gene Transfer	Antimicrobial Agents and Chemotherapy	1991	natural transformation is suggested to be involved in the mosaic formation	-	homologous recombination		
McGraw, E.A.	Molecular Evolution and Mosaic Structure of α, β, and γ Intimins of Pathogenic Escherichia coli	Molecular Biology and Evolution	1999	-	-	recombination occurring particularly in the 3' end of the gene		

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Novais, C.	A tet(S/M) hybrid from CTn6000 and CTn916 recombination	Microbiology	2012	-	-	the mosaic gene has probably resulted from recombination between CTn916 and CTn6000		
Ohnishi, M.	Spread of a Chromosomal Cefixime-Resistant penA Gene among Different Neisseria gonorrhoeae Lineages	Antimicrobial Agents and Chemotherapy	2010	Natural transformation assays performed	transfer of the whole mosaic gene or generation of a mosaic gene during transformation of N. gonorrhoeae ST1901 to N. gonorrhoeae ST7363	-		
Radstrom, P.	Transformational Exchanges in the Dihydropteroate Synthase Gene of Neisseria meningitidis: a Novel Mechanism for Acquisition of Sulfonamide Resistance	Journal of Bacteriology	1992	Evidence of HGT detected during sequence analysis. (Natural transformation assays performed - but not related with the mosaic gene transfer)		homologous recombination after acquisition of DNA fragments from unknown sources by N. meninigitidis by HGT is suggested to have been involved in the formation of the mosaic gene		
Reichmann, P.	A Global Gene Pool for High-Level Cephalosporin Resistance in Commensal Streptococcus Species and Streptococcus pneumoniae	The Journal of Infectious Diseases	1997	Natural transformation assays performed	transfer of PBP 2x and PBP 1a from S. mitis to S. pneumoniae in one or two transformation steps; transfer of pbp2x from S. oralis and S. mitis to S. pneumoniae in one-step transformation; transfer of more than 1 mosaic gene from S. mitis or S. pneumoniae to S. mitis; transfer of pbp2b (and pbp2x in some transformants) from S. mitis and S. oralis to S. pneumoniae and S. mitis	-		Different classes of the pbp2x were present. The susceptible S. mitis NTCT 10712 also contains a mosaic pbp2x gene
Takeda, M.	Characterization of the Neurotoxin Produced by Isolates Associated with Avian Botulism	Avian Diseases	2005	-	-	-	Vidal, 2013 Lindberg, 2010	

	Bibliography				Gene transfer			
1. Author	**Title**	**Journal**	**Year**	**Horizontal gene transfer** (transfer between environments; transfer between cells via transformation/conjugation/ transduction)	**Transformation** (info on involved components)	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Vidal, D.	Environmental factors influencing the prevalence of a Clostridium botulinum type C/D mosaic strain in nonpermanent Mediterranean wetlands	Applied and Environmental Microbiology	2013	vector borne transfer (flies, larvae)	-	-	Takeda, 2005 Lindberg, 2010	
Lindberg, A.	Real-time PCR for Clostridium botulinum type C neurotoxin (BoNTC) gene, also covering a chimeric C/D sequence--application on outbreaks of botulism in poultry	Veterinary Microbiology	2010	-	-	-	Vidal, 2013 Takeda, 2005	
Potron, A.	A mosaic transposon encoding OXA-48 and CTX-M-15: towards pan-resistance	Journal of Antimicrobial Chemotherapy	2013	conjugation (tests performed)	-	-		not exactly a mosaic gene
Unemo, M.	High-level cefixime- and ceftriaxone-resistant Neisseria gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment failure	Antimicrobial Agents & Chemotherapy	2012					

Additional literature on competence induction and transformation

Tables 9 -11 (Literature analysis D, E, F) are part of a **single** Excel database.

Literature analysis D: Competence, transformation

Table 32. Literature analysis D: Competence, transformation

1. Author	Title	Journal	Year	Competent bacterial species	Normal Habitat	Origin of the strain	Recipient Strain(s)	Donor Strain/DNA
Juni, E.	Transformation Assay for Identification of Psychrotrophic Achromobacters	Applied and Environmental Microbiology	1980	Achromobacter spp.		fish and fisheries, poultry, gamma-ray-treated Vienna sausage and ground beef, pork sausage, seawater, and a few unknown sources.	strain 14 (Hyx-7). 109 Achromobacter strains	109 Achromobacter strains
Domingues, S.	Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species	PloS Pathogens	2012	Acinetobacter baylyi		Soil	BD413	Acinetobacter baumannii 064 (kanR); Acinetobacter baumannii 65FFC (CtxR); Acinetobacter baylyi SD1 (spectR); Acinetobacter baylyi SD2 (kanR); Acinetobacter baylyi SD3 (ctxR); Acinetobacter baylyi SD4 (spectR); Acinetobacter baylyi SD5 (spectR); Acinetobacter baylyi SD6 (spectR); Citrobacter freundii C16R385 (spectR); Enterobacter cloacae C2R371 (spectR); *Escherichia coli* C10R379 (trimR); *Escherichia coli* K71-77 (genR); Escherichia fergusonii AS041A2 (trimR); Klebsiella pneumoniae K66-45 (genR); *Pseudomonas aeruginosa* K34-73 (genR); *Pseudomonas aeruginosa* SM(spectR); *Salmonella* rissen 486(spectR); *Salmonella* typhimurium 490 (spectR)
Hendrickx, L.	Natural Genetic Transformation in Monoculture Acinetobacter sp. Strain BD413 Biofilms	Applied and Environmental Microbiology	2003	Acinetobacter baylyi		Soil	BD413	gfp-carrying plasmid pGAR1 (Mob^+, Tra^-, TetR)
Lorenz, M.G.	Plasmid transformation of naturally competent Acinetobacter calcoaceticus in non-sterile soil extract and groundwater	Archives of Microbiology	1992	Acinetobacter baylyi		soil	BD413 (trpE27)	prototrophic A. baylyi BD4
Nielsen, K.M.	Natural Transformation of Acinetobacter sp. Strain BD413 with Cell Lysates of Acinetobacter sp.,Pseudomonas fluorescens, and Burkholderia cepacia in Soil Microcosms	Applied and Environmental Microbiology	2000	Acinetobacter baylyi		Soil	BD413 (pFG4) (KanS)	Acinetobacter spp., Pseudomonas fluorescens R2f and Burkholderia cepacia P2 (KanR due to a *npt*II gene)
Rizzi, A.	Strategy for In Situ Detection of Natural Transformation-Based Horizontal Gene Transfer Events	Applied and Environmental Microbiology	2008	Acinetobacter baylyi		Soil	BD413 (rbcL- PaadA::gfp)(RifR, KanR, SpectS, GFP negative)	E. coli XL-1 Blue (pCLT); A. brasilense SPF94 (pRK290TS); A. tumefaciens LBA 4404 (pRK290TS); S. meliloti RF1 (pRK290TS); B. subtilis TS116; transplastomic leaf and root; wild type leaf
Vaneechoutte, M.	Naturally Transformable Acinetobacter sp. Strain ADP1 Belongs to the Newly Described Species Acinetobacter baylyi	Applied and Environmental Microbiology	2006	Acinetobacter baylyi		Soil (BD413), river (93A2) and activated sludge (B2T, C5 and A7)	BD413(=ADP1), 93A2, $B2^T$, C5 and A7	pZR80
Williams, H.G.	Natural Transformation in River Epilithon	Applied and Environmental Microbiology	1996	Acinetobacter baylyi		Soil	A. baylyi HGW1521(pQM17) (His- Met+ RifR SpR HgR); A. baylyi HGW1510 (His+ Met+ RifR SpR)	A. baylyi HGW98(pQM17) (His+ Met- HgR); A. baylyi BD413
Harding, C.M.	Acinetobacter baumannii Strain M2 Produces Type IV Pili Which Play a Role in Natural Transformation and Twitching Motility but Not Surface-Associated Motility	mBio	2013	Acinetobacter baumannii		Clinical	M2	pGEM-blsA::strAB (strR)

1. Author	Title	Journal	Year	Competent bacterial species	Normal Habitat	Origin of the strain	Recipient Strain(s)	Donor Strain/DNA
Ramirez, M.S.	Naturally Competent Acinetobacter baumannii Clinical Isolate as a Convenient Model for Genetic Studies	Journal of Clinical Microbiology	2010	Acinetobacter baumannii		Clinical	A118	pJHCMW1 (KanR AmkR); pMET1 (KanR AmkR); pAADA1KN (KanR); pAADB (KanR); pVK102 (KanR)
Ramirez, M.S.	Acinetobacter baumanniiis Able to Gain and Maintain a Plasmid Harbouring In35 Found inEnterobacteriaceaeIsolates From Argentina	Current Microbiology	2012	Acinetobacter baumannii		Clinical	A118	pDCMSR1 (from Proteus mirabilis Prm9) (CtxR)
Bosse, J.T.	Harnessing natural transformation in Actinobacillus pleuropneumoniae: a simple method for allelic replacements	FEMS Microbiology Letters	2004	Actinobacillus pleuropneumoniae		-	Shope 4074 (serotype 1) and L20 (serotype 5b)	ΔapxIAΔapxIIA aroA::Km sodC::Km pJSK333 (sodC::Km)
Bosse, J.T.	Natural competence in strains of Actinobacillus pleuropneumoniae	FEMS Microbiology Letters	2009	Actinobacillus pleuropneumoniae		-	Shope 4074 (serotype 1), 1421 (serotype 3), M62 (serotype 4), L20 (serotype 5b), 405 (serotype 8), HS143 (serotype 15). Other serotypes (10 more strains) were transformed (but didn't produce transformants)	A. pleuropneumoniae 4074 sodC::kan
Redfield, R.J.	Evolution of competence and DNA uptake specificity in the Pasteurellaceae	BMC Evolutionary Biology	2006	Actinobacillus pleuropneumoniae; Haemophilus influenzae		-	A. pleuropneumoniae HS143; H. influenzae KW20	A. pleuropneumoniae sodC::kan; H. influenzae MAP7; A. pleuropneumoniae HS143; H. influenzae KW20, B. subtilis, H. parasuis DNA
Fujise, O.	Clonal distribution of natural competence in Actinobacillus actinomycetemcomitans	Oral Microbiology and Immunology	2004	Actinobacillus actinomycetemcomitans		clinical strains; laboratory strains	67 A. actinomycetemcomitans strains belonging to 7 different serotypes (only 18 were shown to be natural competent)	recombinant PCR product from A. actinomycetemcomitans (pilA'-SpeR-pilC') and chromosomal DNA from NalR RifR mutant of A. actinomycetemcomitans strain D17S
Tonjum, T.	Identification of Haemophilus aphrophilus and Actinobacillus actinomycetemcomitans by DNA-DNA Hybridization and Genetic Transformation	Journal of Clinical Microbiology	1990	Actinobacillus actinomycetemcomitans; Haemophilus aphrophilus	oral cavity	clinical	H. aphrophilus 9574/86, E.H./79, 3203/86, 40503/88; A. actinomycetemcomitans 47/87, 1489/79 (other strains were tested but were not competent)	H. aphrophilus NCTC 5906, NCTC 5886; H. paraphrophilus NCTC 10557; A. actinomycetemcoitans NCTC 9710; H. influenza NCTC 8143; H. aphrophilus E.H/79, 3203/86, 9574/86, 46/87, 40503/88; A. actinomycetemcoitans 1489/79, 11305/79, 47/87, 5775/87; H. aphrophilus/H. paraphrophilus IIMOA1/82, IIMOA2/82, 800/89
Wang, Y.	Natural Transformation and DNA Uptake Signal Sequences in Actinobacillus actinomycetemcomitans	Journal of Bacteriology	2002	Actinobacillus actinomycetemcomitans	oral cavity	clinical	A. actinomycetemcomitans D7S (serotype a) (other 16 bacteria were tested but were not naturally transformable)	Transformation experiments: A. actinomycetemcomitans mutant D17S (NalR RifR); transformant of D7S (RifR); pPK1 competing DNA: A. actinomycetemcomitans strains D7S and HK1651; E. coli strains DH5 and E44; H. influenzae Rd; pcr-USS-pilA; pcr-pilA
Wang, Y.	Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans	Gene	2003	Actinobacillus actinomycetemcomitans	oral cavity	clinical	A. actinomycetemcomitans strain D7S	-
Buzby, J.S.	Plasmid Transformation in Agmenellum quadruplicatum PR-6: Construction of Biphasic Plasmids and Characterization of Their Transformation Properties	Journal of Bacteriology	1983	Agmenellum quadriplicatum			PR-6 (strS)	pAQ1; biphasic chimeric plasmids pAQE1 to pAQE11; E. coli pAQE1-mixture (mixture of monomeric and multimeric forms of the pAQE1 plasmid from E. coli); pAQBX1 (total plasmid DNA from PR-6 transformant obtained with pAQE1-mixture)
Essich, E.	Chromosomal Transformation in the Cyanobacterium Agmenellum quadruplicatum	Journal of Bacteriology	1990	Agmenellum quadriplicatum		marine	PR-6 (strS)	PR-6 mutant (strR); PR-6 mutant (rifR); E. coli W3110 (for competition experiment)
Stevens, S.E.	Transformation in Agmenellum quadruplicatum	PNAS	1980	Agmenellum quadriplicatum	-	-	PR-6 (strS)	PR-6 mutant (strR)
Stevens, S.E.	Heterospecific Transformation among Cyanobacteria	Journal of Bacteriology	1986	Agmenellum quadriplicatum; Eucapsis sp.		brine or near-shore seawater	Agmenellum quadriplicatum PR-6 (strS); Agmenellum quadriplicatum	Anacystis nidulans TX-20 (strR); Coccochloris elabens 17-A (strR); Coccochloris elabens Di (strR); Agmenellum quadruplicatum PR-6 (strR); Agmenellum quadruplicatum BG-1

1. Author	Title	Journal	Year	Competent bacterial species	Normal Habitat	Origin of the strain	Recipient Strain(s)	Donor Strain/DNA
							BG-1 (strS); Eucapsis sp. (or Synechocystis) PCC6906 (strS)	(strR); Aphanocapsa (Synechocystis) PCC 6714 (strR); Eucapsis (Synechocystis) PCC 6909 (strR); Eucapsis (Synechocystis) PCC 6902 (strR).
Demaneche, S.	Natural Transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in Soil	Applied and Environmental Microbiology	2001	Agrobacterium tumefaciens Pseudomonas fluorescens		soil	A. tumefaciens GM19023 (RifR StrR) fluorescens LP59JG (RifR GentR)	E. coli DH10B; P. fluorescens AK15; pKT230 (KanR StrepR); pPZP111 (kanR CmR)
Daniell, H.	Transformation of the cyanobacterium Anacystis nidulans 6301 with the *Escherichia coli* plasmid pBR322	PNAS	1986	Anacystis nidulans			6301 (or UTEX625)	pBR322 (ampR tetR)
Golden, S.S.	Optimal Conditions for Genetic Transformation of the Cyanobacterium Anacystis nidulans R2	Journal of Bacteriology	1984	Anacystis nidulans			R2	R2A6 (pCH1) (ampR); R2D2 (DCMU-R); E. coli (pSG111) (ampR cmR)
Shestakov, S.V.	Evidence for Genetic Transformation in Blue-Green Alga Anacystis nidulans	Molecular and General Genetics	1970	Anacystis nidulans			A. nidulans strain 602 (eryS strepS) and mutants: R-20 (eryR), R-32 (strepR), Fil-M2 (eryS strepS)	R-20 (eryR), R-32 (strepR)
David, M.	Transformation of Azotobacter vinelandii with Plasmids RP4 (IncP-i Group) and RSF1010 (IncQ Group)	Journal of Bacteriology	1981	Azotobacter vinelandii Escherichia coli	soil (A. vinelandii)		A. vinelandii UW; A. vinelandii MS(=ATCC 12837); E. coli SK1590	RP4 (from UW, MS and E. coli C60), RSF1010
Glick, B.R.	Transformation of Azotobacter vinelandii with Plasmid DNA	Journal of Bacteriology	1985	Azotobacter vinelandii	soil		ATCC 12837	pRK2501 (tetR kanR), RSF1010 (sulR strepR), pGSS15 (tetR ampR)
Page, W.J.	Physiological Factors Affecting Transformation of Azotobacter vinelandii	Journal of Bacteriology	1976	Azotobacter vinelandii	soil		A. vinelandii ATCC 12837; A. vinelandii ATCC 12837 mutants nif-5, ura-21, ade-15, hyp-18; A. vinelandii UW; A. vinelandii UW1 (nif-)	A. vinelandii ATCC 12837 mutant: rif^r113 (rifR).
Maier, R. J.	Transfer from Rhizobium japonicum to Azotobacter vinelandii of Genes Required for Nodulation	Journal of Bacteriology	1978	Azotobacter vinelandii			A. vinelandii UW10 (Nif-)	Rhizobium japonicum 61A76 (Nif+)
Lu, N.	Adsorption of Extracellular Chromosomal DNA and Its Effects on Natural Transformation of Azotobacter vinelandii	Applied and Environmental Microbiology	2010	Azotobacter vinelandii	soil		DJ77	A. vinelandii DJ
Coukoulis, H.	Transformation in Bacillus amyloliquefaciens	Journal of Bacteriology	1971	Bacillus amyloliquefaciens			N-10 (arginine auxotroph)	B. amyloliquefaciens strain F, K, N, N-10T, P, SB, T, VA. B. subtilis starin 168-w, W-23
Gwinn, D.D.	Transformation of Bacillus licheniformis	Journal of Bacteriology	1964	Bacillus licheniformis			B. licheniformis 9945A mutants: -M28, -M30, -M33 (others were used but did not produce transformants: 9945A-M1, -M2, -M3, -M5, -M8). 10716 mutants (were not transformed)	B. licheniformis ATCC 9945A B. licheniformis 10716
Jensen, K.K.	Protoplast Transfor ma tion of Bacillus licheniformis MC 14	Journal of General Microbiology	1989	Bacillus licheniformis			MC14 (tetS)	pBC16 (tetR)
Zawadzki, P.	The Log-Linear Relationship Between Sexual Isolation and Sequence Divergence in Bacillus Transformation is Robust	Genetics	1995	Bacillus mojavensis Bacillus subtilis			B. subtilis YB886, RO-J-2, RS-C-4, RS-E-2. B. mojavensis IM-A-312, IM-C-45, IM-F-1.	B. subtilis 168 group (1A2, RO-A-4, RO-NN-1); B. subtilis W23 group (2A2 and RO-E-2). B. mojavensis RO-C-2, RO-H-1, RO-QQ-2. B. atrophaeus NRS-213, B. amyloliquefaciens ATCC 23350, B. licheniformis ATCC 14580.
Mareckova, H.	Transformation in Rhizobium japonicum	Archiv für Mikrobiologie	1969	Bradyrhizobium japonicum		soybean	211 (strepS)	211 mutant (strepR)
Raina, J.L.	Deoxyribonucleate Binding and Transformation in Rhizobium japonicum	Journal of Bacteriology	1972	Bradyrhizobium japonicum		soybean root nodules	str-s (strepS)	str-r1 (strepR)
Singh, R.K.	Transfer of Nitrate Reductase Genes of the Cyanobacterium Nostoc muscorum into Rhizobium japonicum	Journal of General Microbiology	1983	Bradyrhizobium japonicum			CB1809 mutant: NR-6 SmR NeoR	Nostoc muscorum ATCC 27893
Graham, J.B.	Genetic Exchange in Bacillus subtilis in Soil	Molecular and General Genetics	1978	Bacillus subtilis			168R (RifR EryR SpcR LinR); SB-/R (His- Trp- AmtR AzlR)	168R (RifR EryR SpcR LinR); SB-/R (His- Trp- AmtR AzlR)
Hauser, P.M.	A rapid and simple method for Bacillus subtilis transformation on solid media	Microbiology	1994	Bacillus subtilis			M22 leuA8, L5256 ilvA1, 168 trpC2, 2A1 thr	L1440, GSY1127, 2A2
Vojcic, L.	An efficient transformation method for Bacillus subtilis DB104	Applied Microbiology and Biotechnology	2012	Bacillus subtilis			DB104	pC194, pHY300PLK, pHY300Car, pHCMC04
Wang, X.	Across Genus Plasmid Transformation Between Bacillus subtilis and *Escherichia coli* and the Effect of *Escherichia coli* on the transforming Ability of Free Plasmid DNA	Current Microbiology	2007	Bacillus subtilis			DB104; 168	E. coli HB101 (pAPR8-1), E. coli HB101 (pMK4), E. coli $TG1^m$ (pAPR8-1), TG1 (pAPR8-1)

1. Author	Title	Journal	Year	Competent bacterial species	Normal Habitat	Origin of the strain	Recipient Strain(s)	Donor Strain/DNA
Kim., J.S.	Natural Transformation-Mediated Transfer of Erythromycin Resistance in Campylobacter coli Strains from Turkeys and Swine	Applied and Environmental Microbiology	2006	Campylobacter coli		farm animals (turkeys and swine)	961, 3325, 1536, 931g, 3237, 1651, 1702rnd, 1787, 44nec, 6034, 37-2nec, 7474, 7580, 614-3m, 426, 4963, 2113, 3175, P5, WP66, 5980, WP126, 1684, WP14, WP19, WP145 (all eryS)	C. coli 1705, 1800r, 2901, 1686, 1420, 1702, 2562, 2774 (all eryR)
Richardson, P.T.	Integration of Heterologous Plasmid DNA into Multiple Sites on the Genome ofCampylobacter coli following Natural Transformation	Journal of Bacteriology	1997	Campylobacter coli			UA585	pSP105, pCK1, pCK12, pCK2, pCK14, pCK17, pCK4
Wang, Y.	Natural Transformation in Campylobacter Species	Journal of Bacteriology	1990	Campylobacter coli Campylobacter jejuni			C. coli UA417, UA420, UA585, UA724, BM2509; UA585 (pUOA15), UA585 (pUOA13). C. jejuni UA67, UA466, UA649, UA580, UA697, C31; UA649 (pUOA649).	C. coli mutant UA417R (NalR StrR); C. jejuni mutant UA466R (NalR StrR); E. coli JM107; pILL550A (C. jejuni), pUOA17 (E. coli), pUOA17 (C. coli), pUOA15 (C. coli), pUA466 (C. jejuni).
de Boer, P.	Generation of Campylobacter jejuni genetic diversity in vivo	Molecular Microbiology	2002	Campylobacter jejuni		chicken	C. jejuni mutants 2412hipO::CmR, 2412htrA::KmR, 2535hipO::CmR, 2535htrA::KmR	C. jejuni mutants 2412hipO::CmR, 2412htrA::KmR, 2535hipO::CmR, 2535htrA::KmR
Vegge, C.S.	Natural Transformation of Campylobacter jejuni Occurs Beyond Limits of Growth	PLoS ONE	2012	Campylobacter jejuni		clinical	C. jejuni NCTC11168	C. jejuni NCTC11168 Δtlp1::Cam^{R}, NCTC11168 rpsL^{Sm}. C. jejuni DVI-SC11 (strpR), DVI-SC181 (strepR). Helicobacter pylori ATCC700392, Arcobacter butzleri ATCC4916
Wilson, D.L.	Variation of the natural transformation frequency of Campylobacter jejuniin liquid shake culture	Microbiology	2003	Campylobacter jejuni			C. jejuni 3130, 33292, 33560, 1-176	C. jejuni 3130 CR2161, 33292 CR2162, 33560 CR2162, 81-176 CR2161, 43429 CR2161, 43470 CR2161 (all cipR). C. coli 18493. E. coli DH5alpha
Frigaard, N.U.	Chromosomal Gene Inactivation in the Green Sulfur Bacterium Chlorobium tepidum by Natural Transformation	Applied and Environmental Microbiology	2001	Chlorobium tepidum			WT2321	pTN1G4 (AmpR GenR), pTN2G1 (AmpR GenR), pTN3G11 (AmpR GenR), pTN1S3 (AmpR StrepR and specR), pTN1CE1 (AmpR CmR EryR). Synechococcus sp. Strain PCC 7002
Mattimore, V.	Genetic Characterization of Forty Ionizing Radiation-Sensitive Strains of Deinococcus radiodurans: Linkage Information from transformation	Journal of Bacteriology	1995	Deinococcus radiodurans			43 Ionizing radiation sensitive (IRS) strains - mutants of 302 (40 were natural transformable)	D. radiodurans R1 mutant: LS18 (strepR); pPG9, pPG11, pPG12. IRS strains
Tirgari, S.	Transformation in Micrococcus radiodurans : Measurement of Various Parameters and Evidence for Multiple, Independently segregating Genomes per Cell	Journal of General Microbiology	1980	Deinococcus radiodurans			D. radiodurans (wild-type)	D. radiodurans Krase
Kennan, R.M.	The Type IV Fimbrial Subunit Gene (fimA)of Dichelobacter nodosus Is Essential for Virulence, Protease Secretion, and Natural Competence	Journal of Bacteriology	2001	Dichelobacter nodosus		clinical (ovine)	VCS1703A; JIR3727 and JIR3728 (fimA mutants)	pJIR1895 (tetR); pJIR1836 (eryR)
Villar, M.T.	Role of the Eikenella corrodens pilALocus in Pilus Function and Phase Variation	Journal of Bacteriology	2001	Eikenella corrodens	oral cavity and gastrointestinal tract in humans	clinical	VA1; VA1-S1; T99 (mutant)	pEC233, pEC237
Kristensen, B.M.	Natural Transformation of Gallibacterium anatis	Applied and Environmental Microbiology	2012	Gallibacterium anatis			12656-12, F149, 10672/6, 10T4, 21K2, 24T10, Avicor, 4895, 07990 (F149, 4895 and 07990 couldn't be transformed by plasmid pBA1100)	12656-12 mutant NalR; 12656-12 gtzA::KmR; 12656-12 gtxBD::KmR; pBA1100
Dargis, M.	Modification in Penicillin-Binding Proteins during In Vivo Development of Genetic Competence of Haemophilus influenzae Is Associated with a Rapid Change in the Physiological State of Cells	Infection and Immunity	1992	Haemophilus influenzae			Rd	H. influenzae T-1,3
Enne, V.I.	Sulfonamide Resistance in Haemophilus influenzae Mediated by Acquisition of sul2 or a Short Insertion in Chromosomal folP	Antimicrobial Agents and Chemotherapy	2002	Haemophilus influenzae			Rd RM118; R162	H. influenzae A12 and A18 (PFGE type HI001); H. influenzae Z26, Z43, Z46 and Z49 (PFGE type HI002)
Takahata, S.	Horizontal Gene Transfer of ftsI, Encoding Penicillin-Binding Protein 3, in Haemophilus influenzae	Antimicrobial Agents and Chemotherapy	2007	Haemophilus influenzae			H. influenzae Rd mutant: Rd^{RIF}	H. influenzae MSC06647, MSC06651, MSC06663
Gromkova, R.C.	Genetic Transformation in Haemophilus parainfluenzae Clinical Isolates	Current Microbiology	1998	Haemophilus parainfluenzae		clinical (human)	73 H. parainfluenzae belonging to biotype I (n= 30, 4 were transformable), II (n=30, 15 were transformable) and III (n=13 -> none	H. parainfluenzae mutant StrRNalR; nontransformable H. parainfluenzae

1. Author	Title	Journal	Year	Competent bacterial species	Normal Habitat	Origin of the strain	Recipient Strain(s)	Donor Strain/DNA
							were transformable)	
Sisco, K.L.	Sequence-specific DNA uptake in Haemophilus transformation	PNAS	1979	Haemophilus parainfluenzae Haemophilus influenzae			H. influenzae Rd strain KW 21 H. parainfluenzae	pKS17
Biggas, A.	Development of a genetic manipulation system for Haemophilus parasuis	Veterinary Micrpbiology	2005	Haemophilus parasuis		pig	HP100	HP101 (Hp100 mutant - strepR); pHRP309; pUA1058 (didn't produce transformants)
Baltrus, D.A.	Multiple phases of competence occur during the Helicobacter pylori growth cycle	FEMS Microbiology Letters	2005	Helicobacter pylori			G27, J99, 26695; G27traG::aphA3 (KanR)	pcagA::cat, pSNV017, cag::cat, H. pylori G27SR (strepR); 26695 pcagA::cat; G27cagA::cat
Jiesong, H.	Helicobacter pylori acquistion of metronidazole resistance by natural transformation in vitro	World Journal of Gastroenterology	1998	Helicobacter pylori		clinical	NCTC 11637, H1, H9, H13, H41, H43, H46, H50, H53, H68 (metS). (H11, H29 and H62 were also tested but were not transformed)	H. pylori H38 (metR)
Hofreuter, D.D.	Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system	Molecular Microbiology	2001	Helicobacter pylori			P1, P12, 26695	H. pylori P1 derivative (strepR); pDH29 (eryR)
Noto, J.M.	Genetic Manipulation of a Naturally Competent Bacterium, Helicobacter pylori	Methods in Molecular Biology	2012	Helicobacter pylori	human stomach	-	-	-
Israel, D.A.	Characteristics of Helicobacter pylori natural transformation	FEMS Microbiology Letters	2000	Helicobacter pylori			HPK5 HPK1 84-183 26695	HPK5 (StrepR KanR) HPK1 (StrepR KanR) CH4 (StrepR) 84-183 (StrepR KanR) 549/91 (StrepR) HPK5 (StrepR KanR; purifed DNA)
Scocca, J. J.	Specificity in deoxyribonucleic acid uptake by transformable Haemophilus influenzae	Journal of Bacteriology	1974	Haemophilus influenzae				
Moccia, C.	The nucleotide excision repair (NER) system of Helicobacter pylori: role in mutation prevention and chromosomal import patterns after natural transformation	BMC Microbiology	2012	Helicobacter pylori				
Kulick, S.	Mosaic DNA Imports with Interspersions of Recipient Sequence after Natural Transformation of Helicobacter pylori	PLoS ONE	2008	Helicobacter pylori				

Literature Analysis E: Competence, transformation

Table 33: Literature analysis E: Competence, transformation

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Juni, E.	Transformation Assay for Identification of Psychrotrophic Achromobacters	Applied and Environmental Microbiology	1980	crude DNA	yes - Achromobacter strains	yes - Acinetobacter calcoaceticus (NCTC 7976), Branhamella catarrhalis (ATCC 23246), Flavobacterium meningosepticum (ATCC 13253), Group IIj, (one of the unnamed groups of gram-negative bacteria classified according to phenotypic characteristics by the Center for Disease Control) (CDC 2706 and 5839), Moraxella atlantae (CDC 5118), Moraxella bovis (ATCC 10900), Moraxella lacunata (ATCC 17952 and 17970), Moraxella nonliquefaciens (ATCC 17953 and 19975), Moraxella osloensis (ATCC 10973), Moraxella phenylpyruvica (ATCC 23333 and CDC 9158), Moraxella urethralis (ATCC 17960), Neisseria caviae (ATCC 14659), and Neisseria ovis (ATCC 17575).	-	-	-	-	-	Solid; in vitro	-	-	-
Domingues, S.	Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species	PloS Pathogens	2012	chromosomal; cell lysates	yes (A. baylyi SD1, A. baylyi SD2, A. baylyi SD3, A. baylyi SD4, A. baylyi SD5, A. baylyi SD6)	yes (A. baumannii 064, A. baumannii 65FFC, C.freundii C16R385, E. cloacae C2R371, E. coli C10R379, E. coli K71-77, E. fergusonii AS041A2, K. pneumoniae K66-45, P. aeruginosa K34-73, P. aeruginosa SM, S. rissen 486, S. typhimurium 490)	**>4.5E-04** (A. baylyi SD1 DNA); **5.2E-04** (A. baylyi SD2 DNA); **<3.3E-05** (A. baylyi SD3 DNA); **<2.5E-06** (A. baylyi SD4 DNA); **<1.2E-05** (A. baylyi SD5 DNA); **<9.6E-05** (A. baylyi SD6 DNA); **<2.1E-08** (A. baumannii 064 DNA); **< 1.6E-07** (A. baumannii 65FFC DNA); **1.2E-08** (C. freundii C16R385 DNA); **1.9E-08** (E. cloacae C2R371 DNA); **0** (E. coli C10R379 DNA); **2.1E-07** (E. coli K71-77 DNA); **0** (E. fergusonii AS041A2 DNA); **4.4E-09** (K. pneumoniae K6-45 DNA); **7.0E-08** (P. aeruginosa K34-73 DNA); **<2.8E-08** (P. aeruginosa SM DNA); **<1.8E-07** (S. rissen 486 DNA); **<1E-07** (S. typhimurium 490 DNA); **<8.9E-08** (A. baumannii 064 lysate); **<1.2E-07** (A. baumannii 65FFC lysate); **<1.4E-07** (P. aeruginosa SM lysate); **<2.1E-07** (S. rissen 486 lysate); **4.4E-08** (S. typhimurium 490 lysate)	1.3E-08 (A. baumannii 064 DNA); 7.5E-08 (S. typhimurium 490 DNA)	-	at the begin of the starting culture	DNA saturation (10 µg of naked DNA were used; lysates contained approx 10 ng of DNA)	solid (filter transformation); in vitro	-	-	-

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Hendrickx, L.	Natural Genetic Transformation in Monoculture Acinetobacter sp. Strain BD413 Biofilms	Applied and Environmental Microbiology	2003	plasmid	no	yes (E. coli plasmid pGAR1)	-	exposure to 0.2 µg/ml: 1 day old biofilm: 1.9E-02 (15 min DNA exposure), 1.3E-02 (45 min DNA exposure); 3 day old biofilm: 7.7 E-04 (15 min DNA exposure), 7.4E-04 (45 min DNA exposure). 0 to 1.5 µg/ml of DNA: 10^{-5} to 10^{-3} (increased frequencies with increased amount of DNA). Starvation of cells, 0.1 µg/ml DNA: 9.1E-05 and 1.9E-05	log phase	1 day after biofilm formation	(different concentration from 1E-09 to 1.5 µg/ml of plasmid DNA were used); saturation was not reached	biofilm; in situ	-	a fraction	-
Lorenz, M.G.	Plasmid transformation of naturally competent Acinetobacter calcoaceticus in non-sterile soil extract and groundwater	Archives of Microbiology	1992	chromosomal, plasmid, phage	yes chromosomal DNA BD413, BD413 (trp-), BD4	yes - E. coli AB1157 chr DNA, phage P22, salmon testes DNA, plasmids RSF1010 (A. baylyi BD413) and pKT210 (A. baylyi BD413 and E. coli DH5)	-	Transformation of A. baylyi BD413 by plasmid pKT210 = **10^{-6}.** Transformation of BD413 by pKT210 (A. baylyi and E. coli) and RSF1010 (A. baylyi) = **10^{-4}.** Transformation of BD413 by pKT210 in groudwater and soil extract = **10^{-5}**	-	-	chr DNA saturating level = 0.45 µg/ml. Plasmid DNA saturating level = 0.25 µg/ml	liquid; in vitro	no - competition between homologous transforming DNA (trp+) and homologous nontransforming DNA (trp-) or heterologous DNA from eukaryotic, prokaryotic or phages or with plasmid DNA. BD413 does not discriminate between homologous and heterologous DNA or between chromosomal or plasmid DNA	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Nielsen, K.M.	Natural Transformation of Acinetobacter sp. Strain BD413 with Cell Lysates of Acinetobacter sp.,Pseudomonas fluorescens, and Burkholderia cepacia in Soil Microcosms	Applied and Environmental Microbiology	2000	chromosomal and cell lysates	yes - homologous (isogenic A. baylyi BD413) or heterologous (Pseudomonas fluorescens R2f and Burkholderia cepacia P2) DNA containing a region with a homologous region (KTG cassette)		Filter transformation with 100 µl cell lysates: **3.0E-05** (A. baylyi BD413), **3.5E-06** (P. fluorescens R2f), **6.3E-06** (B. cepacia P2). Filter transformation with 10-fold concentrated cell lysate: **5.8E-04** (A. baylyi BD413). Filter transformation with sterile-filtered cell suspensions: **1.2E-05** (A. baylyi BD413), **1.3E-07** (P. fluorescens R2f), **6.9E-07** (B. cepacia P2). Transformation in sterile soil microcosms: **7.4E-06** (A. baylyi BD413), **3.1E-07** (P. fluorescens R2f), **1.8E-06** (B. cepacia P2). Transformation in nonsterile soil microcosms: **1.1E-06** (A. baylyi BD413), **<1.4E-08** (transformants not detected with P. fluorescens R2f), **<1.8E-08** (transformants not detected with B. cepacia P2). Transformation in sterile soil after incubation of A. baylyi BD413 cell lysates: **9.9E-06** (0 days), **5.6E-07** (1 day), **1.2E-07** (2 days), **8.8E-08** (3 days), **4.5E-08** (4 days). Transformation in sterile soil after incubation of P. fluorescens R2f cell lysates: **2.2E-06** (0 days), **3.1E-07** (1 day), **3.5E-08** (2 days), **4.9E-08** (3 days), **<5.0E-09** (4 days, transformants not detected). Transformation in sterile soil after incubation of B. cepacia P2 cell lysates: **5.3E-07** (0 days), **3.6E-07** (1 day), **2.5E-07** (2 days), **4.8E-08** (3 days), **<5.6E-09** (4 days, transformants not detected). Transformation in nonsterile soil after incubation of A. baylyi BD413 cell lysates: **3.0E-06** (0 hrs), **1.7E-06** (1 hr), **6.2E-07** (2 hrs), **2.1E-07** (4 hrs), **2.1E-07** (8 hrs), **<2.2E-08** (24 hrs). Transformation in nonsterile soil after incubation of P.fluorescens R2f cell lysates: **2.0E-07** (0 hrs), **1.7E-07** (1 hr), **9.3E-08** (2 hrs), **4.9E-08** (4 hrs), **<4.3E-08** (8 hrs, ransformants not detected), transformants not detected (24 hrs). Transformation in nonsterile soil after incubation of B. cepacia P2 cell lysates: **4.1E-07** (0 hrs), **3.9E-07** (1 hr), **2.2E-07** (2 hrs), **1.3E-07** (4 hrs), **<1.5E-08** (8 hrs, transformants not detected),	-	-	-	(7.8 to 19.2 µg/ml of DNA present in the cell lysates were used; 0.1, 1, 10, and 50 µg of purified DNA were used)	solid (filter transformation; soil microcosms); in vitro and in situ	non specific	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
							transformants not detected (24 hrs).								
Rizzi, A.	Strategy for In Situ Detection of Natural Transformation-Based Horizontal Gene Transfer Events	Applied and Environmental Microbiology	2008	circular plasmid (pCLT); linear plasmid (pCLT); total DNA (A. brasilense SPF94 (pRK290TS); A. tumefaciens LBA 4404 (pRK290TS); S. meliloti RF1 (pRK290TS)); chromosome (B. subtilis TS116); cell lysate (E. coli XL-1 Blue (pCLT), A. brasilense SPF94 (pRK290TS), A. tumefaciens LBA 4404 (pRK290TS), S. meliloti RF1 (pRK290TS)); total DNA (transplastomic leaf tissue, transplastomic root tissue); homogenate (transplastomic leaf, wild type leaf)	yes - heterologous DNA containing a region with few homologous genes (rbcL-aadA-accD)	-	**1.8E-05** (circular plasmid pCLT); **1.4E-05** (linear plasmid pCLT); **6.1E-07** (total DNA A. brasilense SPF94 pRK290TS); **2.7E-07** (A. tumefaciens LBA 4404 pRK290TS); **3.1E-07** (S. meliloti RF1 pRK290TS); **3.3E-08** (chromosome B. subtilis TS116); **9.5E-06** (cell lysate E. coli XL-1 Blue pCLT); **4.3E-07** (cell lysate A. brasilense SPF94 pRK290TS); **1.5E.07** (cell lysate A. tumefaciens LBA 4404 pRK290TS); **1.7E-07** (cell lysate S. meliloti RF1 pRK290TS)); **2.4E-08** (total DNA transplastomic leaf tissue); **3.4E-09** (total DNA transplastomic root tissue); **8.4E-10** (transplastomic leaf homogenate); **<1E-11** (wild type leaf homogenate); in situ transformation frequency = **6.3E-03** (plasmid or cell lysate pCLT)	-	-	at the begin of the starting culture	(1 µg of plasmid DNA, 3 µg of total DNA, 50 µl of cell lysates, 5 µg of total DNA from transplastomic leaf or transplastomic root tissue, and 500 mg of transplastomic or wild type leaf homogenate were used)	solid (filter transformation; tobacco root and leaf); in vitro and in situ	-	-	-
Vaneechoutte, M.	Naturally Transformable Acinetobacter sp. Strain ADP1 Belongs to the Newly Described Species Acinetobacter baylyi	Applied and Environmental Microbiology	2006	Plasmid linear	yes -plasmid containing na ADP1 lipA allele in which a kanamycin resistance cassette was inserted	-	**7.7E03** (ADP1); **3.2E03** (93A2); **1.7E03** ($B2^T$); **1.0E03** (C5); **1.0E03** (A7)	-	-	-	(0.1 µg was used)	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Williams, H.G.	Natural Transformation in River Epilithon	Applied and Environmental Microbiology	1996	cell lysates or whole cells	yes	no	Transformation of HGW1521(pQM17) by: lysates of BD413 in vitro (PCA) = **7.13E-04**; BD413 whole cells in vitro (PCA) = **9.55E-05**; Lysates of BD413 in situ (River Taff) = **7.94E-07**; BD413 whole cells in situ (River Taff) = **7.76E-06**. Transformation of HGW1521 by BD413 lysates in: sterile microcosm River Taff = **1.1E-04**; sterile microcosm River Hillsborough = **9.66E-04**; sterile microcosm River Weeki Wachee = **5.50E-04**; nonsterile microcosm River Taff = **2.45E-05**; nonsterile microcosm River Hillsborough = **4.62E-05**; nonsterile microcosm River Weeki Wachee = **1E-03**; in situ River = **7.88E-04**; in situ River Hillsborough = **1.02E-03**; in situ River Weeki Wachee = **2.2E-06**. Transformation of HGW1521(pQM17) by BD413 lysates in situ in River Hillsborough at 22°C = **1.04E-02**. Transformation of HGW1521(pQM17) by saturating levels of BD413 lysates ≈ **1E-03.** In situ Transformation of exponential-phase culture of HGW1521(pQM17) by BD413 lysates in the Hillsborough River = **7.94E-05 to 1.04E-02**; In situ Transformation of stationary-phase culture of HGW1521(pQM17) by BD413 lysates in the Hillsborough River = **1E-07 to 2.45E-06**. Transformation of HGW1521(pQM17) incorporated into the epilithon in Hillsborough River by: BD413 lysates = **1E-04**; BD413 whole cells = not above the background frequency. In vitro (PCA) transformation of HGW1510 by: cell lysates of HGW98(pQM17) = **1.11E-06**; whole cells of HGW98(pQM17) = **4.64E-07**. In situ transformation (River Hillsborough) of HGW1510 by: HGW98(pQM17) lysates = **3.07E-07**; HGW98(pQM17) whole cells = **1.26E-07**.	-	In situ transformation: exponential-phase	-	-	-	in vitro and in situ	-	-	-
Hardin g, C.M.	Acinetobacter baumannii Strain M2 Produces Type IV Pili Which Play a Role in Natural	mBio	2013	plasmid, linear	yes - plasmid containing a homologous gene (blsA was amplified from A.		approx 0.5 x 10^{-6}	-	-	-	(1 µg was used)	liquid; in vitro	-	-	DNA uptake by type IV pili

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
	Transformation and Twitching Motility but Not Surface-Associated Motility				baumannii M2)										
Ramirez, M.S.	Naturally Competent Acinetobacter baumannii Clinical Isolate as a Convenient Model for Genetic Studies	Journal of Clinical Microbiology	2010	plasmid	no	yes (different plasmids)	-	(11.3 2.4) x 10^2 (pJHCMW1) (7.4 0.3) x 10^2 (pMET1) (7.9 0.8) x 10^2 (pAADA1KN) (13.7 2.9) x 10^2 (pAADB) (26.9 7.4) x 10^2 (pKV102)	-	at the begin of the starting culture	(100 ng DNA was used)	liquid; in vitro	-	1.8% ± 0.9% of the cells (uptake of fluorophore-labeled 10-mer phosphorothioate oligodeoxynucleotide analog)	-
Ramirez, M.S.	Acinetobacter baumanniiis Able to Gain and Maintain a Plasmid Harbouring In35 Found inEnterobacteriaceaeIsolates From Argentina	Current Microbiology	2012	plasmid	no	yes (plasmid from P. mirabilis)	-	-	-	at the begin of the starting culture	(100 ng DNA was used)	liquid; in vitro	-	-	-
Bosse, J.T.	Harnessing natural transformation in Actinobacillus pleuropneumoniae: a simple method for allelic replacements	FEMS Microbiology Letters	2004	Chromosomal (ΔapxIAΔapxIIA, aroA::Km, sodC::Km) and linear plasmid (pJSK333)	yes- serotype 1 knockout strains	E. coli plasmid (with homologous gene - sodC)	10^{-8}	serotype 1 recipient: **361±61** (donor ΔapxIAΔapxIIA); **144±17** (donor aroA::Km); **405±71** (donor sod::Km); **152±20** (donor pJSK333). Serotype 5 recipient: **30±8** (donor ΔapxIAΔapxIIA); **117±28** (donor sod::Km); **140±34** (donor pJSK333).	-	-	saturating level of chromosomal DNA: 0.125 µg/10^9 (1 µg of DNA per ml can be generally used)	solid and liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Bosse, J.T.	Natural competence in strains of Actinobacillus pleuropneumoniae	FEMS Microbiology Letters	2009	chromosomal	yes	-	**10^{-8}** (strains 4074, 1421, M62, L20 and 405) **10^{-4}** (HS143)	-	beginning of stationary phase or after transfer to MIV (stravation medium) only for one strain, HS143	-	-	solid and liquid; in vitro	-	-	-
Redfield, R.J.	Evolution of competence and DNA uptake specificity in the Pasteurellaceae	BMC Evolutionary Biology	2006	chromosomal	yes	yes	-	-	-	-	-	in vitro	yes: discrimination in transformation-competition experiments. H. influenzae took up preferentially H. influenzae DNA and not B. subtilis or H. parasuis DNA. A. pleuropneumoniae took up its own DNA in preference to H. influenzae and B. subtilis DNA; H. parasuis DNA also competed with the A. pleuropneumoniae DNA.	-	A. pleuropneumoniae: Apl-type uptake signal sequence (USS). H. influenzae: Hin-type USS
Fujise, O.	Clonal distribution of natural competence in Actinobacillus actinomycetemcomitans	Oral Microbiology and Immunology	2004	cloned DNA and chromosomal DNA	yes	no	**5E-03 to 4E-06** (median **1.5E-04**) (No significant variations were seen in the transformation frequencies with respect to the sources of the donor DNA)	-	-	-	-	Solid; in vitro	-	-	uptake of DNA containing specific USS

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Tonjum, T.	Identification of Haemophilus aphrophilus and Actinobacillus actinomycetemcomitans by DNA-DNA Hybridization and Genetic Transformation	Journal of Clinical Microbiology	1990	total genomic DNA	yes	yes	Recipient H. aphrophilus 9574/86: **5.1E-04** to **2E-03**. A. actnomycetemcomitans 47/87 and 1489/79: **6.2E-05** to **1.3E-03**	-	-	-	(0.05 ml DNA 200 µg/ml were used)	in vitro	yes: each strain preferentially uptakes DNA from the same species (table with transformation ratios)	-	-
Wang, Y.	Natural Transformation and DNA Uptake Signal Sequences in Actinobacillus actinomycetemcomitans	Journal of Bacteriology	2002	Transformation experiments: genomic and plasmid (supercoiled) DNA competing DNA: genomic DNA (A. actinomycetemcomitans strains D7S and HK1651;E. coli strains DH5 and E44; H. influenzae Rd); DNA fragments produced by PCR (pcr-USS-pilA (874 bp) and pcr-pilA (868 bp)).	yes	no	1.1E-03 (chr DNA) 10^{-8} (plasmid)	represented in a graphic	-	2 hrs after incubation of recipient cells	1 µg	Solid; in vitro	yes: competition for uptake between A. actinomycetemcomitans and H. influenzae DNA and pcr-USS-pilA, but E. coli DNA and pcr-pilA does not compete	-	uptake of DNA with a 9 bp USS
Wang, Y.	Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans	Gene	2003	-	(yes?)	(no?)	10^{-3}	-	-	-	(0.1 to 0.5 µg were used)	Solid; in vitro	-	-	involvement of type IV pilus in natural transformation

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Buzby, J.S.	Plasmid Transformation in Agmenellum quadruplicatum PR-6: Construction of Biphasic Plasmids and Characterization of Their Transformation Properties	Journal of Bacteriology	1983	plasmid	no	yes	**7.1E-09** (donor pAQE1); **3.6E-08** (donor pAQE2); **4.8E-05** (donor pAQE3); **7.3E-05** (donor pAQE4); **3.4E-06** (donor pAQE5); **2.2E-04** (donor pAQE6); **<5.9E-09** (donor pAQE9); **1.3E-08** (donor pAQE10); **7.9E-07** (donor pAQE11); **6.6E-08 transformants per CFU/µg DNA** (donor pAQE1-mixture E. coli); **5.3E-06 transformants per CFU per normalized quantity of DNA** (donor pAQBX1 PR-6)	-	-	-	20 µg	liquid; in vitro	-	-	-
Essich, E.	Chromosomal Transformation in the Cyanobacterium Agmenellum quadruplicatum	Journal of Bacteriology	1990		yes	yes	Donor strR: **5.1E-04** (30°C); **1E-04** (39°C). Donor rifR: **6.4E-04** (30°C); **8.6E-05** (39°C). Donor rifR with 1 µg/ml DNA: **2E-03** (with modified transformation procedure); **3.7E-05** and **1.2E-04** (with standard procedure)	-	exponential growth	-	1 µg (in some experiments 10/20/23 µg /ml were used)/ 4E07 cells	liquid; in vitro	no: no discrimination between A. quadruplicatum and E. coli DNA in competition experiments	-	-
Stevens, S.E.	Transformation in Agmenellum quadruplicatum	PNAS	1980	purified DNA	yes	no	5E-04 to 3.7E-03	-	exponential growth	-	1 µg/ 4E07 cells	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Stevens, S.E.	Heterospecific Transformation among Cyanobacteria	Journal of Bacteriology	1986	purified DNA	yes	yes	recipient A. quadruplicatum PR-6: **3.5E-06** (donor A. nidulans TX20); **2.0E-03** (donor Coccochloris elabens 17-A); **6.7E-04** (donor Coccochloris elabens Di); **4.3E-04** (donor A. quadruplicatum PR-6); **5.8E-04** (donor A. quadruplicatum BG-1); **< 3.2E-08** (not detected donor Aphanocapsa PCC 6714); **1.7E-04** (donor Eucapsia PCC 6909); **6.8E-06** (donor Eucapsia PCC 6902). Recipient A. quadruplicatum BG-1: **5.1E-04** (donor A. quadruplicatum PR-6); **6.1E-04** (A. quadruplicatum BG-1). Recipient Eucapsia PCC 6906: **4.5E-04** (donor A. quadruplicatum PR-6); **1.1E-04** (donor Eucapsia PCC 6906).	-	-	-	1 µg/ 4E07 cells	liquid; in vitro	(from the tranformation frequencies obtained with homologous and heterologous DNA, transformation does not seem to significantly discriminate DNA from the different tested origins, from different species and genera)	-	-
Demaneche, S.	Natural Transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in Soil	Applied and Environmental Microbiology	2001	bacterial suspension (E. coli DH10B; P. fluorescens AK15) plasmid (pKT230, pPZP111)	no	yes	Sterile soil microcosms - Recipient P. fluorescens LP59JG: **5.8E-08** (donor E. coli DH10B); **8.3E-08** (donor pKT230). Sterile soil microcosms -Recipient A. tumefaciens GM19023: **2.7E-09** (donor P. fluorescens AK15); **1.3E-08** (donor pPZP111). In vitro recipient P. fluorescens LP59JG: **<1.3E-07** (not detected; donor E. coli DH10B); **<2.6E-09** (not detected; donor pKT230). In vitro recipient A. tumefaciens GM19023: **6.0E-09** (donor pPZP111).	-	-	-		solid, liquid, soil microcosms; in vitro, in situ	-	-	-
Daniell, H.	Transformation of the cyanobacterium Anacystis nidulans 6301 with the *Escherichia coli* plasmid pBR322	PNAS	1986	plasmid	no	yes	-	10^{-5} (the efficiency of transformation was enhanced 50-fold using permeaplasts compared to cells as recipients	maximum transformation was observed after 28 hrs of contact between the donor DNA and the recipient cells	-	(1 µg DNA was used)	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Golden, S.S.	Optimal Conditions for Genetic Transformation of the Cyanobacterium Anacystis nidulans R2	Journal of Bacteriology	1984	plasmid; chromosomal (R2D2)	yes	yes	-	**2.2E6** (donor pCH1, transformation under dark conditions); **7.6E06** (donor pCH1, dark 15-18 h); **2.9E07** (donor pCH1, dark 15-18 h, iron absence); **1.2E04** (donor pSG111, dark 15-18 h); **1.4E03** (donor R2D2 high molecular chr DNA, dark 15-18 h); **1.1E03** (donor R2D2 XhoI cleaved chr DNA, dark 15-18 h)	all phases of culture growth	-	750 ng/5E08 cells (with plasmid pCH1)	liquid; in vitro	-	1.4% cells in the absence of iron (highest level of transformamtion); 0.15% cells (transformation in the dark, 1 µg donor pCH1)	-
Shestakov, S.V.	Evidence for Genetic Transformation in Blue-Green Alga Anacystis nidulans	Molecular and General Genetics	1970	purified DNA	yes	no	-	-	end of the log phase	-	20-30 µg/ml	Solid; in vitro	-	-	-
David, M.	Transformation of Azotobacter vinelandii with Plasmids RP4 (IncP-i Group) and RSF1010 (IncQ Group)	Journal of Bacteriology	1981	plasmid	no	yes	Recipient E. coli SK1590: **5E-05** (donor C600-RP4); **1.7E-05** (donor UW-RP4); **5E-05** (donor MS-RP4). Recipient UW: **<10^{-10}** (donor UW-RP4); **<10^{-10}**, **1.3E-08**, **3.5E.07**, **5E-05**, **2E-06** (donor UW-RP4 with 0, 30, 100, 200 or 300 mM CaCl2, respectively); **10^{-8}** (donor C600-RP4); **<10^{-10}** (donor MS-RP4); **3E-06** (donor UW-RSF1010). Recipient MS: **9E-06** (donor MS-RP4); **<10^{-10}** (donor UW-RP4); **10^{-09}** (donor C600-RP4).	Recipient UW and MS: **3E05** (donor 5 µg/ml UW-RP4 and MS-RP4)	-	-	-	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Glick, B.R.	Transformation of Azotobacter vinelandii with Plasmid DNA	Journal of Bacteriology	1985	plasmid	no	yes (E. coli plasmids)	defined as the number of transformants per microgram of plasmid DNA per total number of viable cells: **5.7±1.8E-02** (donor pRK2501)**; 3E-04 to 4E-03** (donor RSF1010); **1.9±0.6E-02** (donor pGSS15).	-	all phases of culture growth	-	no saturation (pRK2501 added up to 51 µg)	liquid; in vitro	-	44% (with 51 µg of pRK2501)	-
Page, W.J.	Physiological Factors Affecting Transformation of Azotobacter vinelandii	Journal of Bacteriology	1976	crude lysate DNA	yes	no	Recipient A. vinelandii ATCC 12837: **1.5E-05** (10 mM phospate Burk medium); **1.3E-06** (5 mM phospate Burk medium); **5E-07** (2.5 mM phospate Burk medium); **≈2.5E-06** (0.81 mM magnesium and 0.29 mM calcium Burk medium); **3.1E-07** (0.81 mM magnesium and 2.32 mM calcium Burk medium); **6.6E-08** (0.81 mM magnesium and 2.90 mM calcium Burk medium); **≈2.5E-06** (1.62 mM magnesium and 0.58 mM calcium Burk medium). **3.64E-05** (Recipient UW - >50% nonencapsulated); **3E-06** (Recipient ATCC 12837 encapsulated); **7E-04** (recipient UW1 - noncapsulated); **1.65E-05** (recipient nif-5 encapsulated). auxotrophic to prototrophic transfer: **46E-06** (recipient ura-21); **18E-06** (recipient ade-15); **26E-06** (recipient hyp-18); **1620E-06** (recipient UW1). rifR transfer: **29E-06** (recipient ura-21); **6E-06** (recipient ade-15); **1.5E-06** (recipient hyp-18); **19.9E-06** (recipient UW1). auxotrophic to prototrophic and rifR transfer: **0.0044E-06** (recipient ura-21); **0.0001E-06** (recipient ade-15); **0.0189E-06** (recipient hyp-18).	-	late exponential phase	-	-	Solid and liquid; in vitro	-	-	-
Maier, R. J.	Transfer from Rhizobium japonicum to Azotobacter vinelandii of Genes Required for Nodulation	Journal of Bacteriology	1978	cell lysate	no	yes	9.7E-07	-	-	-	-	in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Lu, N.	Adsorption of Extracellular Chromosomal DNA and Its Effects on Natural Transformation of Azotobacter vinelandii	Applied and Environmental Microbiology	2010	chromosomal (dissolved or adsorbed)	yes	no	DNA dissolved in MOPS = **2E-05**; DNA dissolved in MOPS + 100 mM Na+ = **6E-05**; DNA dissolved in MOPS + 1 mM Ca2+ = **7E-05**. DNA adsorbed similar with TF obtained with dissolved DNA.	-	-	-	-	liquid; in vitro	-	-	-
Coukoulis, H.	Transformation in Bacillus amyloliquefaciens	Journal of Bacteriology	1971		yes	yes	-	-	180 min after start transformation protocol	120 min after start of transformation protocol	12 µg/ml	liquid; in vitro	-	-	-
Gwinn, D.D.	Transformation of Bacillus licheniformis	Journal of Bacteriology	1964	genomic DNA	yes	no	-	-	stationary phase	-	-	Solid and liquid; in vitro	-	20% (solid transformation) $10^{-3\%}$ (liquid transformation)	-
Jensen, K.K.	Protoplast Transformation of Bacillus licheniformis MC 14	Journal of General Microbiology	1989	plasmid	no	yes	0.1% (10 µg/ml); 1% (100 µg/ml)	-	-	-	-	liquid; in vitro	-	-	-
Zawadzki, P.	The Log-Linear Relationship Between Sexual Isolation and Sequence Divergence in Bacillus Transformation is Robust	Genetics	1995	genomic DNA; PCR-product	yes	yes	(all transformation frequencies were log_{10}-transformed)	-	-	-	3 µg/ml (B. subtilis 1A2)	-	yes: transformation by donor derived DNA in the presence of recipient-derived DNA was one-half than when only donor-derived DNA was present (for B. subtilis 1A2)	-	-
Mareckova, H.	Transformation in Rhizobium japonicum	Archiv für Mikrobiologie	1969	genomic DNA	yes	no	-	-	maximum competence 9 hours after bacterial growth	-	-	liquid; in vitro	-	0.01%	-
Raina, J.L.	Deoxyribonucleate Binding and Transformation in Rhizobium japonicum	Journal of Bacteriology	1972	purified ^{32}P-labeled DNA	yes	no	0.0001 to 0.01% depending on the media	-	late log phase	-	10 µg/ml	liquid; in vitro	yes: discrimination of the DNA binding between homologous DNA (9%) and heterologous H.	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
													influenzae DNA (2%)		
Singh, R.K.	Transfer of Nitrate Reductase Genes of the Cyanobacterium Nostoc muscorum into Rhizobium japonicum	Journal of General Microbiology	1983		no	yes	1.2E-05 (units ?)	-	exponentially growing cells	-	-	liquid; in vitro	-	-	-
Graham, J.B.	Genetic Exchange in Bacillus subtilis in Soil	Molecular and General Genetics	1978	spores; DNA	yes	no	(Units?) Mixed culture Recipient 168R + donor SB-/R: **1.8E-04** (day 1), **5.3E-05** (day 2), **2.6E-05** (day 3), **4.3E-06** (day 4), **2.7 E-08**(day 6), not detected (day 8). Mixed culture Recipient SB-/R + donor 168R: **1.8E-04** (day 1), **2.3E-04** (day 2), **5.1E-05** (day 3), **3.3E-06** (day 4), **1.6 E-05**(day 6), **2.0E-05** (day 8). gle strain culture 168R + DNA SB-/R: **3.5E-02** (day 1), **1.6E-02** (day 3), **1.7E-01** (day 4), **7.3E-04** (day 6). Single strain culture SB-/R + DNA 168R: not detected (day 1), **1.1E-08** (day 3), **1.6E-04** (day 4), **1.1E-01** (day 6).	-	-	-	(60 µg per pot were added)	soil; in situ	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Hauser, P.M.	A rapid and simple method for Bacillus subtilis transformation on on solid media	Microbiology	1994	chromosomal, (plasmid, recombinant phage DNA, lysate of E. coli infected by recombinant phage)	yes	yes	M22 leuA8 recipient + L1440 donor: **5E-06** (24 hr 10^8 cells per plate); **1E-05** (24 hr 10^7 cells per plate); **1E-05** (24 hr 10^6 cells per plate); **0** (24 hr 10^5 and 10^4 cells per plate); **5E-06** (48 hr 10^8 cells per plate); **3E-05** (48 hr 10^7 cells per plate); **4E-05** (48 hr 10^6 cells per plate); **2E-04** (48 hr 10^5 cells per plate); **3E-03** (48 hr 10^4 cells per plate). L5256 ilvA1 recipient + GSY1127 donor: **5E-06** (24 hr 10^8 cells per plate); **2E-05** (24 hr 10^7 cells per plate); **1E-05** (24 hr 10^6 cells per plate); **3E-05** (24 hr 3E5 and 10^5 cells per plate); **5E-06** (48 hr 10^8 cells per plate); **2E-05** (48 hr 10^7 cells per plate); **3E-05** (48 hr 10^6 cells per plate); **1E-04** (48 hr 3E05 cells per plate); **7E-04** (48 hr 10^4 cells per plate); **5E-03** (48 hr 3E04 cells per plate). 168 trpC2 recipient + L1440 donor: **3E-06** (24 hr 10^8 cells per plate); **5E-06** (24 hr 10^7 cells per plate); **3E-06** (24 hr 10^6 cells per plate); **3E-06** (48 hr 10^8 cells per plate); **5E-06** (48 hr 10^7 cells per plate); **3E-06** (48 hr 10^6 cells per plate). 2A1 thr recipient + 2A2 donor: **2E-05** (24 hr 10^7 cells per plate); **2E-05** (48 hr 10^7 cells per plate).	-	-	-	1 µg	Solid (and liquid); in vitro	-	-	-
Vojcic, L.	An efficient transformation method for Bacillus subtilis DB104	Applied Microbiology and Biotechnology	2012	plasmid	no	yes	-	donor pHY300Car: **8.0E4** to **1.5E05** (with histidine at 0, 10, 50, 200 and 500 µg/ml); 4.5E04 (histidine 1000 µg/ml). **1.3E04** (donor pC194); **1.2E05** (donor pHY300PLK); **1.5E05** (donor pHY300Car); **2.2E04** (donor pHCMC04).	end of exponential growth phase	-	(optimum: 5 to 10 ng of plasmid DNA)	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Wang, X.	Across Genus Plasmid Transformation Between Bacillus subtilis and *Escherichia coli* and the Effect of *Escherichia coli* on the transforming Ability of Free Plasmid DNA	Current Microbiology	2007	purified plasmid; LB cell suspension	no	yes	Recipient DB104: **4.800E-05** (donor HB101 pAPR8-1); **7.167E-06** (donor pMK4); **0.452E-06** (donor TG1^{m} pAPR8-1); **0.535E-06** (donor TG1 pAPR8-1).Recipient 168: **0.375E-05** (donor HB101 pAPR8-1). 20 mM MgCl2 = **8.18E-05**	-	-	exponentially growth phase of the recipients	-	Solid; in vitro	-	-	-
Kim., J.S.	Natural Transformation-Mediated Transfer of Erythromycin Resistance in Campylobacter coli Strains from Turkeys and Swine	Applied and Environmental Microbiology	2006	genomic DNA	yes	no	Liquid assays: Recipient 961, 3237, 3325, 7474 (turkey origin) and donor 1705, 1800r, 2901 = 10^{-05} to 10^{-06}. Recipient 614-3m, 426, P5, 2113, 5980 (swine origin) and donor 1705, 1800r, 2901 = 10^{-07} to 10^{-09}. Recipients 5980, 1684, WP145 were not transformable.	-	-	-	-	Solid and liquid; in vitro	-	-	-
Richardson, P.T.	Integration of Heterologous Plasmid DNA into Multiple Sites on the Genome ofCampylobacter coli following Natural Transformation	Journal of Bacteriology	1997	plasmid	yes (plasmids containing homologous region to the recipient genome derived from the katA gene: pC K1, pCK12, pCK2, pCK14, pCK17, pCK4 with homologous sequence with the length of 125, 175, 270, 286, 370 and 567 bp, respectively)	yes (pSP105 plasmid)	-	**4E-13** (donor pSP105), **3E-13** (donor pCK1), **3E-13** (donor pCK12), **2E-13** (donor pCK2), **8E-13** (pCK14), **5E-12** (donor pCk17), **2E-11** (donor pCK4)	-	-	-	-	-	-	-
Wang, Y.	Natural Transformation in Campylobacter Species	Journal of Bacteriology	1990	chromosomal plasmid	yes	yes (E. coli DNA and plamids)	**10^{-3}** in C. coli recipients and **10^{-4}** in C. jejuni recipients with homologous DNA. Transformation of C. coli UA585 with homologous DNA was **1.2E-03** (nalR marker), **4E-04** (strR marker) and **2E-07** (nalR and strR markers). no of transformants/spot (5E06 recipient cells): Donor C. coli UA417R:	Recipient C. coli UA585 and donor C. coli UA417: **4E05** (0.01 µg/ml DNA); **8E04** (1 µg/ml DNA)	early log phase (though the bacteria is competent	-	1 µg/ml	solid and biphasic; in vitro	yes: C. coli UA585 was transformed with C. jejuni UA466R DNA with 20% efficiency compared with	-	-

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
							3500 (Recipient C. coli UA417; marker strR); **0** (recipient C. coli UA420, marker nalR); **3000** (recipient C. coli UA420, marker strR); **6000** (recipient C. coli UA585, marker NalR); **4000** (recipient C. coli UA585, marker strR); **1000** (recipient C. coli UA724, marker NalR); **1000** (recipient C. coli UA724, marker strR). Donor C. jejuni UA466R: **0** (recipient C. jejuni UA67, marker strR); **300** (recipient C. jejuni UA466, marker NalR); **1000** (recipient C. jejuni UA466, marker StrR); **200** (recipient C. jejuni UA649, marker nalR); **800** (recipient C. jejuni UA649, marker strR); **150** (recipient C. jejuni UA580, marker nalR); **600** (recipient C. jejuni UA580, marker strR); **0** (recipient C. jejuni UA697, marker nalR); **0** (recipient C. jejuni UA697, marker strR); **0** (recipient C. jejuni C31, marker nalR); **0** (recipient C. jejuni C31, marker strR). Recipient C. coli UA417 = **0** (donor C. jejuni pILL550A, marker kanR); recipient C. coli UA585 = **8** (donor C. jejuni pILL550A, marker kanR); recipient C. coli UA585 = **3** (donor E. coli pUOA17, marker kanR); recipient C. coli UA585 = **6** (donor C. coli pUOA17, marker kanR); recipient C. coli UA585 pUOA15 = **500** (donor E. coli pUOA17, marker kanR); recipient C. coli UA585 pUOA15 = **640** (donor C. coli pUOA17, marker kanR); recipient C. coli UA585 pUOA13 = **1200** (donor C. coli pUOA15, marker TcR); recipient C. jejuni UA649 pUA649 = **0** (donor C. jejuni pILL550A, marker kanR); recipient C. jejuni UA649 pUA649 = **200** (donor C. jejuni pUA466 marker TcR); recipient UA650 = **0** (donor C. jejuni pUA466, marker TcR).		through out the entire growth cycle) (for C: coli)				homologous DNA; C. jejuni UA466 was transformed with C. coli UA417R DNA at 1% efficiency compared with homologous DNA.		
de Boer, P.	Generation of Campylobacter jejuni genetic diversity in vivo	Molecular Microbiology	2002	Cells	yes	yes (?) (different C. jejuni strain)	-	-	-	-	-	in vivo (chickens)	-	-	-
Vegge, C.S.	Natural Transformation of Campylobacter jejuni Occurs Beyond Limits of Growth	PLoS ONE	2012	chromosomal	yes	yes (competition experiments)	donor C. jejuni NCTC11168 Δtlp1::CamR: **3.6E-03 ± 1.8E-03** (using excess DNA). Recipient cells in exponential phase = **2.3E-03 ± 3.2E-04**. Recipient cells in late stationary phase = **5.9E-06 ± 2.6E-06**. Donor	donor C. jejuni NCTC11168 Δtlp1::CamR: **5.4E-03 ± 1E-03**	exponential growth (although is also compet	-	-	biphasic; in vitro	yes: two C. jejuni mutants readily transformed C. jejuni NCTC11168, while H. pylori	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
							rpsL^{Sm} + 20 fold diluted recipient cells = **4.1E-05 ± 5.7E-07**. Donor rpsLSm + undiluted recipient cells = **8.2E-07 ± 2.1E-07**.	**transformants/DNA molecule** (using limited amount of DNA)	ent in stationary phase)				and A. butzleri DNA could not outcompete transformation with isogenic DNA (natural transformation seems species- but not strain-specific)		
Wilson, D.L.	Variation of the natural transformation frequency of Campylobacter jejunin liquid shake culture	Microbiology	2003	chromosomal (and co-cultivation of cells)	yes	yes (competition experiments)	iquid transformation in 5% CO2 atmosphere: recipient 3130 and donor 3130 CR2161 = **8E-04 to 3E-06**; recipient 33292 and donor 33292 CR2162 = not detected; recipient 33560 and donor 33560 CR2162 = **7E-04 to 2E-08**; recipient 81-176 and donor CR2161 = **3E-04 to 8E-07**. Co-cultivation of C. jejuni 81-176-Tn5CmR19 and 81-176-23SK4: **3E-08 to 1.4E-05** (0.7% CO2 atmosphere); **4.1E-09 to 6.1E-08** (5% CO2 atmosphere); **9.0E-10 to 1.4E-08** (10% CO2 atmosphere, only occurring in 4 of the 16 data points).	-	early exponential growth	-	1 µg	liquid; in vitro	yes: preferential uptake of C. jejuni DNA rather than C. coli and E. coli DNA in competition experiments	-	-
Frigaard, N.U.	Chromosomal Gene Inactivation in the Green Sulfur Bacterium Chlorobium tepidum by Natural Transformation	Applied and Environmental Microbiology	2001	linearized and circular plasmid; Chromosomal	yes (plasmids containing homologous regions to the recipient)	yes (Synechococus DNA)	donor AhdI-digested pTN1G4: **2E-07 to 3E-07** (after 10 hr incubation); **4E-08** (0.1 µg DNA); **1E-07** (1 µg DNA); **1.4E.07** (10 µg DNA). 10 µg donor DNA: Donor AhdI-digested pTN1G4 (2.93 kb homologous region) = **1.3E-07**; donor EcoRI-digested pTN1G4 (2.93 kb homologous region) = **4E-07**; donor AhdI-digested pTN2G1 (1.08 kb homologous region) = **1.1E-07;** donor EcoRI-digested pTN2G1 (1.08 kb homologous region) = **7E-09**; donor AhdI-digested pTN3G11 (0.29 kb homologous region) = **0**; donor EcoRI-digested pTN3G11 (0.29 kb homologous region) = **0**; AhdI-digested pTN1G4 (GenR marker) = **3E-07**; AhdI-digested pTN1CE1 (EryR marker) = **6E-07**; AhdI-digested pTN1CE1 (StrpR and specR marker) = **3E-03**	donor AhdI-digested pTN1G4 = **10^2** (after 10 hr incubation)	late exponential phase (though stationary phase is also transformable)	-	10 µg (/3E-09 to 6E-09)	Solid; in vitro	no: transformation by Synechococcus decreased the transformation frequency of C. tepidum by AhdI-digested pTN1G4	-	-
Mattimore, V.	Genetic Characterization of Forty Ionizing Radiation-Sensitive Strains of Deinococcus radiodurans: Linkage Information from	Journal of Bacteriology	1995	chromosomal; linear plasmid	yes (D. radiodurans DNA and plasmids containing homologous regions to the recipient)	-	-	0.1% (transformants per µg DNA ??)	exponential phase	-	-	solid (dot transformation); in vitro	-	-	-

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
	transformation														
Tirgari, S.	Transformation in Micrococcus radiodurans: Measurement of Various Parameters and Evidence for Multiple, Independently segregating Genomes per Cell	Journal of General Microbiology	1980	genomic DNA	yes	no	**10^{-04}** (without added CaCl2; rifR marker); **10^{-03} - 10^{-02}** (with added CaCl2, rifR marker); **10^{-05} - 10^{-04}** (without added CaCl2; eryR marker); **10^{-04} -10^{-03}** (with added CaCl2, eryR marker). Marker rifR, CaCL2 conc ranging from 0 to 1.43 M: **10^{-03} to 10^{-02}.** 90 min of incubation with DNA = **7.6E-03**	-	exponential growth phase	-	no evidence of saturation (maximum tested 88 µg/ml)	liquid; in vitro	-		-
Kennan, R.M.	The Type IV Fimbrial Subunit Gene (fimA)of Dichelobacter nodosus Is Essential for Virulence, Protease Secretion, and Natural Competence	Journal of Bacteriology	2001	plasmid	yes (plasmids containing homologous regions to the host genome)	-	-	-	-	-	-	liquid; in vitro	-	-	type IV fimbriae
Villar, M.T.	Role of the Eikenella corrodens pilALocus in Pilus Function and Phase Variation	Journal of Bacteriology	2001	plasmid; linear	yes (plasmids containing homologous regions to the host genome)	-	1E-05 to 3E-05	30 to 60	-	-	-	in vitro	-	-	type IV pilus
Kristensen, B.M.	Natural Transformation of Gallibacterium anatis	Applied and Environmental Microbiology	2012	chromosomal DNA; circular plasmid; linear plasmid	yes	yes	Recipient 12656-12: Donor 12656-12 NalR: **6E-04** (in BHI broth medium); **3E-06** (in blood agar medium). Donor 12656-12 gtzA::KmR = **2E-04**. Donor 12656-12 gtxBD::KmR = **2E-03**. Donor circular pBA1100 = **2E-05**	Recipient 12656-12: Donor 12656-12 NalR = **3E05**. Donor 12656-12 gtzA::KmR = **1E06**. Donor 12656-12 gtxBD::KmR = **1E07**. Donor circular pBA1100 = 7E03	(induction of competence by transfer to M-IV starvation medium)	-	0.5 µg/10^{8} CFU recipient	liquid; in vitro	yes: E. coli DNA competes poorly with G. anatis DNA for uptake	-	low density of uptake sequences
Dargis, M.	Modification in Penicillin-Binding Proteins during In Vivo Development of Genetic Competence of Haemophilus influenzae Is Associated with a Rapid Change in the Physiological State of Cells	Infection and Immunity	1992	chromosomal	yes	no	in a graphic	-	stationary phase	-	-	in vivo (rats)	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Enne, V.I.	Sulfonamide Resistance in Haemophilus influenzae Mediated by Acquisition of sul2 or a Short Insertion in Chromosomal folP	Antimicrobial Agents and Chemotherapy	2002	total DNA; PCR-product	yes	no	-	Recipient Rd RM118: **2.3E-03** (donor total DNA); **1.6E-04** (donor PCR product) (Units?)	-	-	-	-	-	-	-
Takahata, S.	Horizontal Gene Transfer of ftsI, Encoding Penicillin-Binding Protein 3, in Haemophilus influenzae	Antimicrobial Agents and Chemotherapy	2007	cells	yes	no	**5.1E.-07** (donor MSC06647); **1.2E-06** (donor MSC06651); **1.5E-06** (donor MSC06663).	-	-	-	-	liquid; in vitro	-	-	USS
Gromkova, R.C.	Genetic Transformation in Haemophilus parainfluenzae Clinical Isolates	Current Microbiology	1998	genomic DNA	yes	no	-	Donor H. parainfluenzae mutant StrRNalR; recipients from biotype I: **8E03** (recipient 80, strR marker); **2E04** (recipient 80, nalR marker); **5E02** (recipient 81, strR marker); **8E02** (recipient 81, nalR marker); **3E04** (recipient 91, strR marker); **5E04** (recipient 91, nalR marker); **3E03** (recipient 96, strR marker); **6E03** (recipient 96, nalR marker); **4E02** (recipient NCTC 7857, strR marker); **8E02** (recipient NCTC 7857, nalR marker). Donor H.	late stationary phase of growth	-	1 µg/ml	liquid; in vitro	-	-	-

Bibliography				TRANSFORMATION											
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
								parainfluenzae mutant StrRNalR; recipients from biotype II: **5E05 to 6E06** (recipient group 1, marker strR); **2E06 to 8E06** (recipient group 1, marker nalR); **2E04 to 7E04** (recipient group 2, marker strR); **6E04 to 9E04** (recipient group 2, marker nalR); **4E03 to 8E03** (recipient group 3, marker strR); **7E03 to 9E03** (recipient group 3, marker nalR).								
Sisco, K.L.	Sequence-specific DNA uptake in Haemophilus transformation	PNAS	1979	plasmid; linear	yes (plasmid containing 8.1 Kb region from H. parainfluenzae)	-	-	-	-	-	1 µg	liquid; in vitro	yes: H. parainfluenzae cells distinguish between vector DNA and cloned H. parainfluenzae DNA.	-	uptake of DNA with a 8-12bp USS
Biggas, A.	Development of a genetic manipulation system for Haemophilus parasuis	Veterinary Micrpbiology	2005	chromosomal and plasmid	yes	yes	-	10^6 (1 to 5 µg DNA); 3E05 (0.1 ng DNA)	-	beginning of the incubation of the transformation mix	1 to 5 µg	Solid; in vitro	-	-	uptake of DNA with a specific USS

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
Baltrus, D.A.	Multiple phases of competence occur during the Helicobacter pylori growth cycle	FEMS Microbiology Letters	2005	plasmid, PCR-product, genomic DNA	yes	-	**6.27E-05** (recipient G27, donor pcagA::cat); **4.07E-05** (recipient G27, donor pSNV017). **2.32E-05** (recipient J99, donor pcagA::cat); **1.11E-07** (recipient J99, donor J99SR). **3.45E-05** (recipient 26695, donor pcagA::cat); **8.45E-07** (recipient 26695, donor pSNV017)	-	H. pylori G27: competence peaks during log and stationary phase (significant loss of transformability in the late log phase). H. pylori J99: two peaks during log phase and one during stationary phase. H. pylori 26695: variable	-	-	liquid; in vitro	-	-	-
Jiesong, H.	Helicobacter pylori acquistion of metronidazole resistance by natural transformation in vitro	World Journal of Gastroenterology	1998	genomic DNA	yes	no	Recipient NCTC 11637 and donor H38: **2.8E-06** to **5.9E-05**. Recipient H1, H9, H13, H41, H43, H46, H50, H53, H68: **3.4E-06** to **2.4E-04**.	-	-	24 hours after incubation of recipient cells	-	liquid; in vitro	-	-	-
Hofreuter, D.D.	Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system	Molecular Microbiology	2001	chromosomal; plasmid	yes	yes	-	-	-	-	-	liquid; in vitro	-	-	Type IV secretion system
Noto, J.M.	Genetic Manipulation of a Naturally Competent Bacterium, Helicobacter pylori	Methods in Molecular Biology	2012	-	-	-	-	-	-	-	-	Solid and liquid; in vitro	-	-	-
Israel, D.A.	Characteristics of Helicobacter pylori	FEMS Microbiology	2000	chromosomal linear	yes	yes (E. coli DNA)	1,30E-04	1,30E-04	mid- to late-lag phase	at the begin of the starting culture	6 ng /10E8 cells	solid and liquid; in vitro	yes: discrimination between E. coli	only a fraction	not clear

Bibliography				TRANSFORMATION												
1. Author	**Title**	**Journal**	**Year**	**Conformation of source DNA** (chromosomal, episomal; linear, circular)	**Transformation with homologous DNA**	**Transformation with heterologous DNA**	**Transformation frequency** (maximal number of transformants per CFU)	**Transformation efficiency** (transformants per µg DNA)	**Growth phase for optimal transformation**	**Optimal stage for DNA inoculation** (maximal numbers of transformants per CFU)	**DNA saturation** /transformation	**Transformation experiments** (solid, liquid, in situ, in vitro, in vivo)	**DNA uptake specificity**	**Percentage of population transformed**	**DNA uptake sequence**
	natural transformation	Letters							(0-10 h prior to the onset of exponential phase growth)				and H. pylori DNA in competition experiments		
Scocca, J. J.	Specificity in deoxyribonucleic acid uptake by transformable Haemophilus influenzae	Journal of Bacteriology	1974								5-10 ng / 10E8 cells (Scocca, J. J.; J. Bacteriol 1974)				yes
Mocci a, C.	The nucleotide excision repair (NER) system of Helicobacter pylori: role in mutation prevention and chromosomal import patterns after natural transformation	BMC Microbiology	2012												
Kulick, S.	Mosaic DNA Imports with Interspersions of Recipient Sequence after Natural Transformation of Helicobacter pylori	PLoS ONE	2008												DNA uptake: Type IV secretion system

Literature analysis F: Competence induction

Table 34: Literature analysis F: Competence

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Juni, E.	Transformation Assay for Identification of Psychrotrophic Achromobacters	Applied and Environmental Microbiology	1980	naturally competent	loopful of DNA + cell paste Hyx-7 -> suspension and spread over an area of 5 to 8 mm diameter -> 20°C ON (16hrs)	-	-	-		
Domingues, S.	Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species	PloS Pathogens	2012	naturally competent	solid LB + nitrocellulose filter + competent cells + DNA -> 30°C 24 hrs	-	-	-		
Hendrickx, L.	Natural Genetic Transformation in Monoculture Acinetobacter sp. Strain BD413 Biofilms	Applied and Environmental Microbiology	2003	naturally competent	biofilm grown in a flow cell in continuous mode + DNA during 1 h of continuous flow with DNA-containing medium	-	-	-	Williams, 1996; Lorenz 1992	A. baylyi BD413 is referred as Acinetobacter sp. Strain BD413. Young biofilms (1 day) are transformed at higher rates than mature biofilms (3 days); the presence of planktonic cells decreases/inhibits transformation; the transformants accumulate at the bottom of the biofilm
Lorenz, M.G.	Plasmid transformation of naturally competent Acinetobacter calcoaceticus in non-sterile soil extract and groundwater	Archives of Microbiology	1992	naturally competent	Competent cells (10^7 or 10^9 cell per ml, for transformation with chr or plasmid DNA) + medium + DNA (chr 0.45 µg/ml; plasmid 0.5 to 1 µg/ml) -> 30°C 45 to 60 min -> + DnaseI and cations -> 30°C 10min. Transformation in natural samples: competent cells pellet washed with sterile water + groundwater or soil extract + 30 ng pKT210 -> 30°C 1 hr -> + DNaseI + 2-fold concentrated LB broth -> 90 min	-	-	-		A. baylyi is refferred as Acinetobacter calcoaceticus. It was shown that DNA uptake by A. baylyi BD413 requires divalent cations
Nielsen, K.M.	Natural Transformation of Acinetobacter sp. Strain BD413 with Cell Lysates of Acinetobacter sp.,Pseudomonas fluorescens, and Burkholderia cepacia in Soil Microcosms	Applied and Environmental Microbiology	2000	naturally competent	Filter transformation: competent cells + lysed cels or naked DNA -> nitrocellulose filter on top of LB agar with antibiotic -> 30°C overnight. Transformation in soil microcosms: competent cells + sterile or nonsterile soil microcosms -> 24 hrs -> + nutrient solution -> 1 hr -> + cell lysates -> 24 h 20°C. Transformation in sterile soil microcosm: microcosm + cell lysates -> up to 4 days -> competent cells. Transformation in nonsterile soil microcosm: microcosm + cell lysates -> up to 24 hrs -> competent cells.	-	-	-		A. baylyi BD413 is referred as Acinetobacter sp. Strain BD413. Donor bacteria are common soil bacteria. Transformation efficiencies obtained in vitro and in situ with the various lysates were similar to or exceeded those obtained with conventionally purified DNA. The presence of cell debris did not inhibit transformation in soil, and the debris may protect DNA from rapid biological inactivation.

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Rizzi, A.	Strategy for In Situ Detection of Natural Transformation-Based Horizontal Gene Transfer Events	Applied and Environmental Microbiology	2008	naturally competent	in vitro: solid LB + nitrocellulose filter + competent cells + DNA -> 24 hrs in situ: solid LB (for leaf) or soil infusion (for root) + polycarbonate membrane filters + tobbaco leaf or root + competent cells + DNA -> 28°C up to 5 days	-	-	-	Hendrickx 2003	most of the relevant information is available in the supplemental material
Vaneechoutte, M.	Naturally Transformable Acinetobacter sp. Strain ADP1 Belongs to the Newly Described Species Acinetobacter baylyi	Applied and Environmental Microbiology	2006	naturally competent	as described in Young and Ornston, 2001: ON recipient culture + succinate media (1:10) -> 2 hrs 30°C (induction of competence) -> competent cells + donor DNA -> 1 hr 30°C	-	-	-		

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Williams, H.G.	Natural Transformation in River Epilithon	Applied and Environmental Microbiology	1996	naturally competent	In vitro transformation: ON recipient and donor DNA each on a nitrocellulose filter -> placed together on PCA -> 24h 20°C. Transformation in beaker microcosm: beaker with sterile/freshly river water + ON recipient and donor DNA each on a nitrocellulose filter -> placed together on sterile scrubbed slate disc and covered with larger filter -> 24h 20°C. Transformation in situ: ON recipient and donor DNA each on a nitrocellulose filter -> placed together on sterile scrubbed slate disc and covered with larger filter -> nylon mesh bag -> 24h. Transformation of cultures incorporated into river epilithon: LB culture of recipient or donor on filter -> sterile scrubbed stone -> mesh bag -> 24h in the river -> donor in contact with recipient -> 24h in river. Effect of growth of the recipient culture prior to mating: recipient grown for 15 min to 50h -> 1E7 cells per cm^2 and use in in vitro (PCA 20°C 24h; B22 salts solution solidified with agar 5 or 20°C) or in situ transformations. Effect of the amount of lysate added: in vitro transformation with various amounts of lysates - 0.01 to 10 ml.	-	-	-		A. baylyi is refferred as Acinetobacter calcoaceticus. The aquatic environment of river epilithon is able to support natural transformation. The Weeki Wachee River maintained a constant temperature of 24°C all year at the sample site. The river Taff and Hillsborough River varied between 2 and 22°C and between 18 and 30°C, respectively. Both rivers usually varied ±1°C during a 24-h period. The authors examine the effects of temperature, recipient age, and preincorporation of the recipient into the epilithic biofilm on gene transfer. A high degree of variablity between replicates in transformation experiments in situ reflects the dynamic and variable nature of the environment. The transformation frequency in situ generally increased with temperature. Temperature did not have as great an effect on transformation in laboratory experiments on agar. The age of the recipient culture affected the transformation frequencies in situ but did not significantly affect the transfer frequency on laboratory media.
Harding, C.M.	Acinetobacter baumannii Strain M2 Produces Type IV Pili Which Play a Role in Natural Transformation and Twitching Motility but Not Surface-Associated Motility	mBio	2013	naturally competent	fresh LB + ON culture (in stationary phase) (10:1) -> 2 hrs 37°C 180 rpm -> + DNA for 2 hrs 37°C	-	-	-		
Ramirez, M.S.	Naturally Competent Acinetobacter baumannii Clinical Isolate as a Convenient Model for Genetic Studies	Journal of Clinical Microbiology	2010	naturally competent	fresh LB + culture in stationary phase (1:1) + DNA -> 25 or 37°C for 1 hour	-	-	-		

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Ramirez, M.S.	Acinetobacter baumanniiis Able to Gain and Maintain a Plasmid Harbouring In35 Found inEnterobacteriaceaeIsolates From Argentina	Current Microbiology	2012	naturally competent	fresh LB + culture in stationary phase (1:1) + DNA -> 37°C for 1 hour	-	-	-		
Bosse, J.T.	Harnessing natural transformation in Actinobacillus pleuropneumoniae: a simple method for allelic replacements	FEMS Microbiology Letters	2004	naturally competent	liquid: 1 ml recipient cells grown in BHI_N + 1 µg chr DNA -> 37°C 25 min -> + 10 U Dnase I -> 37°C 5 min -> 2V BHI_N -> 37°C 100 min -> selective plates ON 37°C 5% CO2. Solid: ON recipient culture in agar -> 5 ml BHI_N -> until OD600 = 0.50 (10^9 cfu/ml) -> 20 µl BHI_L plates -> 37°C 100 min -> + 10 µl DNA (1 µg) on to the cells -> 37°C 4 h -> resuspension in 2 ml phosphate-buffered saline and plated	-	-	-		
Bosse, J.T.	Natural competence in strains of Actinobacillus pleuropneumoniae	FEMS Microbiology Letters	2009	naturally competent (high level competence can be induced for one strain)	Solid: 20 µl recipient cells on BHI_L plates (OD600 = 0.50) -> 37°C 100 min -> + 10 µl DNA (1 µg) on to the cells -> 37°C 4 h -> resuspension in 2 ml phosphate-buffered saline and plated. Liquid (MIV - starvation medium): cells growing exponentially in BHI-N or BHI-HN transferred to MIV -> 37°C 100 min -> + 1 µg DNA -> 37°C 15 min	-	-	-		competence may be correlated with serotype in A. pleuropneumoniae. Competence is regulated by the sfx gene (at least in one strain)
Redfield, R.J.	Evolution of competence and DNA uptake specificity in the Pasteurellaceae	BMC Evolutionary Biology	2006	Naturally competent	competent cells (made competent by transfer of exponential growing cells to MIV starvation medium) + 100 ng of conspecific DNA + 100, 300 or 900 ng of competing DNA -> 37°C 15 min -> Dnase I -> 5 min -> diluted and plated	-	-	-		
Fujise, O.	Clonal distribution of natural competence in Actinobacillus actinomycetemcomitans	Oral Microbiology and Immunology	2004	Naturally competent (strains from serotypes a, d and e)	Recipient cells grown ON on sTSB (Trypticase Soy Broth supplemented with 0.1% yeast extract, 5% heat inactivated horse serum and 1.5% agar) agar at 37°C in 5% CO2 -> adjusted to 10^9 CFU/ml -> 20 µl spotted onto prewarmed sTSB agar plate and spreaded (~10 mm diameter) -> 2hrs -> + 10 µl donor DNA (100 µg/ml) -> 5-6 hrs -> washed off the agar and plated	-	-	-		Actinobacillus actinomycetemcomitans is now designated Aggregatibacter actinomycetemcomitans
Tonjum, T.	Identification of Haemophilus aphrophilus and Actinobacillus actinomycetemcomitans by DNA-DNA Hybridization and Genetic Transformation	Journal of Clinical Microbiology	1990	naturally competent	as described in Tonjum et al, 1985 (I don't have access to it). In quantitative transformation assays: 0.5 ml of the recipient suspension was exposed to 0.05 ml of each DNA for 20 minutes	-	-	-		Haemophilus aphrophilus is now known as Aggregatibacter aphrophilus
Wang, Y.	Natural Transformation and DNA Uptake Signal Sequences in Actinobacillus actinomycetemcomitans	Journal of Bacteriology	2002	Naturally competent (higher competence induced by cAMP)	**1.** Recipient cells grown ON (15 to 20 hrs) on sTSB (Trypticase Soy Broth supplemented with 0.1% yeast extract, 5% heat inactivated horse serum and 1.5% agar) agar at 37°C in 5% CO2 -> adjusted to 10^9 (fimbriated wild-type strains) or 5 x 10^9 (nonfimbriated mutants) CFU/ml -> 20 µl spotted onto prewarmed sTSB agar plate and spreaded (~10 mm diameter) -> 2hrs -> + 10 µl donor DNA (100 µg/ml) -> 5 (nonfimbriated mutants) or 6 (fimbriated wild-type strains) hrs -> washed off the agar and plated 2 to 7 days. **2.** recipient cells grown ON -> sTSB broth at OD600 = 0.1 -> 30 min 37°C 5% CO2 without shaking -> + cAMP 2 mM -> 50 min -> + donor DNA 2 µg/ml -> 20 min -> DNase I -> plating.In transformation competition assays: the same as above, but 0.2 µg transforming DNA and 1.8 µg competing DNA -> 20 min -> + DNase I -> 5 (nonfimbriated mutants) or 6 (fimbriated wild-type strains) hrs.	-	-	-		Transformation frequency and efficiency of the nonfimbriated A. actinomycetemcomitans strain D7S-smooth were higher than those of the fimbriated, wild-type parental strain D7S. flp-encoded fimbriae are not required for the natural competence of this bacterium
Wang, Y.	Type IV pilus gene homologs pilABCD are required for natural transformation in Actinobacillus actinomycetemcomitans	Gene	2003	Naturally competent	As in Wang, 2002: Recipient cells grown ON (15 to 20 hrs) on sTSB (Trypticase Soy Broth supplemented with 0.1% yeast extract, 5% heat inactivated horse serum and 1.5% agar) agar at 37°C in 5% CO2 -> adjusted to 10^9 (fimbriated wild-type strains) or 5 x 10^9 (nonfimbriated mutants) CFU/ml -> 20 µl spotted onto prewarmed sTSB agar plate and spreaded (~10 mm diameter) -> 2hrs -> + 10 µl donor DNA (0.1 - 0.5 µg) -> 5 (nonfimbriated mutants) or 6 (fimbriated wild-type strains) hrs -> washed off the agar and plated 2 to 7 days.	-	-	-		
Buzby, J.S.	Plasmid Transformation in Agmenellum quadruplicatum PR-6: Construction of Biphasic Plasmids and	Journal of Bacteriology	1983	Naturally competent	As in Stevens,1980, with the follow modifications: competent cells + DNA -> 1 h -> diluted 20-fold in medium A -> 48 h in nonselective liquid culture under reduced lighting -> plated on medium A agar which ampicillin	-	-	-	Stevens, 1980	Agmenellum quadriplicatum is also designated Synechococcus sp.

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
	Characterization of Their Transformation Properties									
Essich, E.	Chromosomal Transformation in the Cyanobacterium Agmenellum quadruplicatum	Journal of Bacteriology	1990	Naturally competent	As in Stevens, 1980. Alternatively: recipient cells grown to 10^8 cells/ml in medium A with 5 g of NaNO3/l. 1 vol DNA + 9 vol recipient cells -> 27 to 30°C without bubbling -> varied incubation times -> + DNase I -> plated	-	-	-	Stevens, 1980	The absence of light before exposure of cells to exogenous DNA resulted in a marked decline in transformation frequency. The transformation frequency is higher when transformation is performed at 30 rather than at 39°C.
Stevens, S.E.	Transformation in Agmenellum quadruplicatum	PNAS	1980	Naturally competent	9 volumes recipient cells grown in medium A (4E-07 cells/ml) + 1 vol DNA (0.14 µg/ml) in 0.15 M NaCl/0.015 M Na3citrate -> 39°C (different periods of time) -> + 1 vol Dnase (10 µg/ml) -> cells plated in medium A agar -> 32 ± 2°C for enough time to allow strR expression -> sprayed str solution -> 32 ± 2°C 3-4 days. (Cells in liquid medium are grown in continous agitation with 1% CO2 and with illumination provided by F24T12 CW/HO fluorescent lamps providing 580 µeinsteins cm^{-2} sec^{-1} -> but the authirs showed that continuous exposure to light is not essential for transformation; cells in solid medium were grown under illumination provided by F96T12 CW fluorescent lamps providing 250 µeinsteins cm^{-2} sec^{-1})	-	-	-		
Stevens, S.E.	Heterospecific Transformation among Cyanobacteria	Journal of Bacteriology	1986	Naturally competent	As in Stevens, 1980: 9 volumes recipient cells grown in medium A (4E-07 cells/ml) + 1 vol DNA (final DNA concentration 1 µg/ml) -> 1 hr in light with Co2 (39°C for PR-6 and BG-1 recipients and 30°C for PCC6906 recipient)-> + DNase 10 µg/ml -> plated and incubated 48 hr -> srpayed with str solution -> 3 to 6 days				Stevens, 1980	
Demaneche, S.	Natural Transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in Soil	Applied and Environmental Microbiology	2001	Naturally competent	Soil microcosms: soil microcosm was seeded with concentrated recipient strain (0.25 ml in water) and donor DNA (15 µg in 0.25 ml water) (soil microcosm with 0.5 µg DNA and 5E09 bacteria g of dry $soil^{-1}$) or donor bacterial suspension (donor and recipient cells at the same concentration, with a final concentration of 5E09 bacteria g of dry $soil^{-1}$. -> 28°C 3 days -> + DNase I -> 3h 28°C in LB -> dilution and plating in LB -> 2 days. In vitro, solid: 20 µl of ON 100-fold-concentrated recipient cells + 0.5 µg DNA or 20 µl of 100-fold-concentrated donor cells -> GTTP filter -> 28°C 24 hrs ->cells resuspended and plated. In vitro, liquid: 40 µl ON recipient cells + 950 µl liquid media (different media were tested) + 10 µl of 100 µg/ml concentrated plasmid solution ->24 hrs 28°C with shaking -> plating.	-	-	-		transformation-mediated gene transfers can occur in soils. P. fluorescens LP59JG failed to produced transformants in vitro (using different media), although it was naturally transformed in situ
Daniell, H.	Transformation of the cyanobacterium Anacystis nidulans 6301 with the *Escherichia coli* plasmid pBR322	PNAS	1986	Naturally competent	1 ml of Permeaplasts (prepared by 2-hr treatment with lysozyme/EDTA) or cells (used as recipients) + 1 µg donor DNA -> 32°C with light with agitation. Selection of transformants: inoculation on agar plates containing 10 mM sodium acetate and 0.5 µg ampicillin per ml in Kratz and Myers medium after growing the transformants in the presence of ampicilin 0.25 µg/ml in fresh growth medium 2-3 days; or incubation in the absence of antibiotic at 32°C 24 h -> 400 µl ampicilin solution 50 µg/ml dispensed nderneath the agar slab. -> All plates 32°C 7 days	-	-	-		permeaplasts are highly permeable cells that have a high efficiency of cell-wall regeneration and subsequent division. Inclusion of sodium acetate in the plates favored faster division and growth of the transformants
Golden, S.S.	Optimal Conditions for Genetic Transformation of the Cyanobacterium Anacystis nidulans R2	Journal of Bacteriology	1984	Naturally competent	5E08/ml recipient cells in BG-11 medium (or iron-deficient BG-11) -> + DNA (30 ng to 3.5 µg/l) -> 28 to 30°C for different times (up to 18hrs) (with or without light) -> plated -> 30°C 3 to 12 hrs -> addition of selective agent -> 5 to 7 days	-	-	-		Preincubation of the recipient cells in the dark, decrease transformation frequency. Transformation is more efficient when occuring in dark conditions (the same is observed when cells were treated with chemicals that inhibit photosynthesis, such as DCMU and CCCP) -> disruption of the photosynthetic process increases transformation efficiency

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Shestakov, S.V.	Evidence for Genetic Transformation in Blue-Green Alga Anacystis nidulans	Molecular and General Genetics	1970	Naturally competent	Recipient cells in phosphate buffer ph 7.5 + donor DNA -> 38°C 2500 lux (or 2 hr 37°C 3000 lux ??) -> + Dnase -> 20 min 38°C -> plated in (modified C-medium) agar without antibiotic -> 2 days -> + antibiotic solution into the hole in agar near the wall of the dish -> 7-8 days; or 3 days -> plates covered with a second agar layer with antibiotic	-	-	-		active uptake of DNA and possibly the process of integration are directly dependent on the energy of photosynthesis.
David, M.	Transformation of Azotobacter vinelandii with Plasmids RP4 (IncP-i Group) and RSF1010 (IncQ Group)	Journal of Bacteriology	1981	CaCl2 induced transformation	Preparation of A. vinelandii competent cells: 30°C in C medium ->2E08 cells/ml -> 20 ml for 10 min 0°C -> cells harvested and washed -> 10 ml cold CaCl2 200 mM -> 0°C 20 min -> washed and suspended in 2ml cold CaCL2 200 mM. Transformation with A. vinelandii: competent cells + DNA (diluted in cold CaCl2) -> 0°C 60 min -> 42°C 2 min -> + C-Tris medium 30°C -> 30°C 20 hrs shaking -> plating on selective medium -> 4 days 30°C.Transformation of E. coli as in Cohen, 1972.	-	-	-	Page, 1976	
Glick, B.R.	Transformation of Azotobacter vinelandii with Plasmid DNA	Journal of Bacteriology	1985	Naturally competent	Single recipient colonies -> TF medium -> 170 rpm 30°C -> OD620nm <0.2 -> + fresh TF medium -> 170 rpm 30°C ->1.6E08 cells/ml -> 50 µl cells + 300 µl fresh TF medium + 50 µl DNA (22 µg/ml) -> 30 min 30°C -> centrifugation -> pellet + 400 µl fresh TF medium -> 60 min 30°C -> plated on AG medium with and without selection -> 30°C 72h.	-	-	-	Page, 1976; David, 1981	Growth of Azotobacter sp. In TF medium is sufficient to prepare the cells for the uptake of plasmid DNA
Page, W.J.	Physiological Factors Affecting Transformation of Azotobacter vinelandii	Journal of Bacteriology	1976	Naturally competent	Solid: recipient cells (2E07) + DNA (1 to 1.5 µg) in Burk medium -> 24 h 30°C -> suspended and diluted -> plated on selective medium -> 4 days 30°C. Liquid: recipients cells grown in liquid Burk medium with shaking -> 50 µl cells + DNA -> 30 min 30°C -> DNase -> 24 h 30°C	-	-	-		Transformation was most successful on solid medium. Optimal transformation was obtained at 30 C and pH 7.0. Phosphate 10 mM enhanced transformation. Magnesium ions enhanced transformation.
Maier, R. J.	Transfer from Rhizobium japonicum to Azotobacter vinelandii of Genes Required for Nodulation	Journal of Bacteriology	1978	Naturally competent	As in Page, 1976, except that 2% sucrose replaced glucose in the transformation medium.	-	-	-	Page, 1976	Intergeneric transformation
Lu, N.	Adsorption of Extracellular Chromosomal DNA and Its Effects on Natural Transformation of Azotobacter vinelandii	Applied and Environmental Microbiology	2010	Naturally competent	recipient cells grown on BN (modified Burk medium) plates 30°C 2 days -> liquid BN -> 30°C 170 rpm 18-20 hrs. Competent cells + MOPS buffer (plus specified ions) + DNA (2 µg) -> 20 to 35 min room temperature -> diluted and plated in B and BN plates -> 30°C 3 to 5 days.	-	-	-		
Coukoulis, H.	Transformation in Bacillus amyloliquefaciens	Journal of Bacteriology	1971	Naturally competent	recipient cells grown ON 37°C on Tryptose blood agar base -> penassay broth with CaCl2 10^{-3} M -> washed -> modified glucose minimal medium -> cell-growth transformation medium (2E7 CFU/ml). Cells + DNA (10 to 15 µg) -> 37°C 180 min -> suspended and plated.	-	-	-		
Gwinn, D.D.	Transformation of Bacillus licheniformis	Journal of Bacteriology	1964	Naturally competent	Solid: recipient cells (grown from spores) grown in NBY or NBSG +DNA (45 µg) -> minimal agar plate -> 37°C 4-5 days. Liquid: recipient cells in broth + DNA (10 to 50 µg) -> 37°C water bath 30 min -> +minimal broth -> plated on minimal agar plates + DNase	-	-	-		
Jensen, K.K.	Protoplast Transfor ma tion of Bacillus licheniformis MC 14	Journal of General Microbiology	1989	Naturally competent	Protoplast transformation: recipient cells grown ON in LB broth with glucose -> penassay broth, agitation 37°C, until OD540 = 0.5 ->SMMP -> lysozyme -> agitation 1h room temperature -> protoplasts suspended in SMMP -> + 10-100 µg DNA -> + PEG -> 2 min room temperature -> + SMMP -> diluted and plated in regeneration medium -> 46°C 3 to 5 days	-	-	-		
Zawadzki, P.	The Log-Linear Relationship Between Sexual Isolation and Sequence Divergence in Bacillus Transformation is Robust	Genetics	1995	Naturally competent	-	-	-	-		Shorter PCR-product gave lower transformation frequency than longer one.
Mareckova, H.	Transformation in Rhizobium japonicum	Archiv für Mikrobiologie	1969	Naturally competent	Recipient cells + DNA -> 90 min -> Dnase -> plating	-	-	-		
Raina, J.L.	Deoxyribonucleate Binding and Transformation in Rhizobium japonicum	Journal of Bacteriology	1972	Naturally competent	Competence medium + recipient cells + DNA -> 30°C 150rev/min -> + Dnase -> 15 min -> ice bath -> cell pellet + medium (SSC) -> diluted and plated -> 2 hrs 30°C -> + agar with selection -> 48 hrs 30°C	-	-	-		Development of competence in the late log phase is enhanced by the presence of casamino acids and Mg2+ ions

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Singh, R.K.	Transfer of Nitrate Reductase Genes of the Cyanobacterium Nostoc muscorum into Rhizobium japonicum	Journal of General Microbiology	1983	Naturally competent	As in Stevens 1980: recipient cells + DNA (10 µg/ml) -> 25°C, 3 hrs, gentle shaking -> Dnase I -> plating	-	-	-		Bradyrhizobium japonicum was previously known as Rhizobium japonicum
Graham, J.B.	Genetic Exchange in Bacillus subtilis in Soil	Molecular and General Genetics	1978	Naturally competent	Mixed strain cultures: Bacillus spores -> + Dnase I 30-60 min 37°C -> 80°C 20 min -> dripped into pots with soil and whased into the soil. Single strain culture: DNA dripped into pots with soil and whased into the soil -> bacillus spores prepared as above into soil ->37°C	-	-	-		
Hauser, P.M.	A rapid and simple method for Bacillus subtilis transformation on solid media	Microbiology	1994	Naturally competent	Recipient cells or spores -> spread onto LA plates -> ON 30°C -> cells resuspended -> +DNA -> TS supplemented plates -> 37°C	-	-	-		
Vojcic, L.	An efficient transformation method for Bacillus subtilis DB104	Applied Microbiology and Biotechnology	2012	Naturally competent	Preparation of competent cells: Single recipient colony -> SM1 liquid medium -> 37°C 14-16 h 250 rpm -> + SM1 to 2.9E7 cells/ml -> 37°C 3 hrs 200 rpm -> + SM2 medium + histidine -> 37°C 2 hrs 300 rpm (cells remained competent for 1 hr). Transformation: competent cells + DNA -> 37°C 30 min 200 rpm -> + fresh LB -> 37°C 30 min 200 rpm -> plated on selective plates	-	-	-		
Wang, X.	Across Genus Plasmid Transformation Between Bacillus subtilis and *Escherichia coli* and the Effect of *Escherichia coli* on the transforming Ability of Free Plasmid DNA	Current Microbiology	2007	Naturally competent	Exponential-phase recipient in LB + DNA -> MM (spizzen minimal medium) agar or Eppendorf tubes -> 28°C 40 min -> plating on selective plates. Single colony of recipient and donor in LB (separately) -> ON 37°C vigorous shaking -> 1:100 dilution in freah LB -> 37°C. Exponential-phase recipient + exponential-phse donor -> 40 min 28°C -> diluted and plated -> 37°C 12/22 hrs	-	-	-		
Kim., J.S.	Natural Transformation-Mediated Transfer of Erythromycin Resistance in Campylobacter coli Strains from Turkeys and Swine	Applied and Environmental Microbiology	2006	Naturally competent	Solid: recipient cells (grown ON on Muller-Hinton agar plates 42°C microareobically) spotted onto Muller-Hinton agar plates + 4 µl DNA -> mixed over a 0.5 cm region -> 15 to 17 hr 42°C microaerobic conditions -> plated in selective plates -> 42°C 48 hr microaerobic conditions. Liquid: recipient cells grown on sheep blood agar plates 42°C 48 hr microaerobic conditions. 1 colony -> MH broth -> 42°C 24 hr microaerobic conditions -> + MH broth -> 42°C 7 hrs microaerobic conditions -> + DNA (3 µg) -> 5 hr 42°C microaerobic conditions -> plated on selective plates ->36 to 48 hr 42°C.	-	-	-	Wang, 1990	
Richardson, P.T.	Integration of Heterologous Plasmid DNA into Multiple Sites on the Genome ofCampylobacter coli following Natural Transformation	Journal of Bacteriology	1997	Naturally competent	As in Wang, 1990.				Wang, 1990	Homologous recombination occurred with 286 homologous bp. Plasmids with less or no homology were still integrated in teh chromosome of C. coli, by illegitmate recombination
Wang, Y.	Natural Transformation in Campylobacter Species	Journal of Bacteriology	1990	Naturally competent	Solid: recipient cells grown on MH agar for 24 hr 37°C 7% CO2 -> spread on MH agar (5E07 cells per plate) -> 6 hrs -> + DNA (0.2 µg) spotted onto the inoculated agar -> 5 hrs -> plated in selective plates. Biphasic: cell suspension (1E07 to 5E07 cells/ml in MH broth) transfer to tube with MH agar -> 2 to 6hrs -> + DNA -> 3 to 5hrs -> + DNase I and MgCl2 -> plated in selective plates.	-	-	-		
de Boer, P.	Generation of Campylobacter jejuni genetic diversity in vivo	Molecular Microbiology	2002	Naturally competent	Animals inoculated with mixtures of two C. jejuni mutants (10^5 cfu each)	-	-	-		

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Vegge, C.S.	Natural Transformation of Campylobacter jejuni Occurs Beyond Limits of Growth	PLoS ONE	2012	Naturally competent	Exponentially growing C. jejuni in BHI broth transferred to BHI agar -> 2 hr microaerobic conditions 37°C -> + 10-400 ng DNA -> 2-4 hrs same conditions-> diluted and plated in selective media	-	-	-		Transformation was found to occur at all tested conditions (although at a highly variable degree): 20, 30, 37 and 42°C; aerobic, microaerobic and anaerobic conditions. The level of transformatio nwas highest under microaerobic conditions at 42°C. Natural transformation occurred very efficiently in the pH interval 7 to 11. Stress does not enhances or induces competence. Transformation was greatly reduced by respiratory inhibitors. The efficiency of natural transformation by C. jejuniis connected to growth rate.
Wilson, D.L.	Variation of the natural transformation frequency of Campylobacter jejuniin liquid shake culture	Microbiology	2003	Naturally competent	Recipient cells grown 48 hr on Bolton agar 37°C 5% CO2. Single colony + Bolton broth -> 37°C 250 rpm 0.7, 5 or 10% CO2 -> 12-24 hrs -> 1 to 5E08 cfu/ml + Bolton broth -> 37°C 250 rpm 0.7, 5 or 10% CO2 -> each 2 hrs transfer onto Bolton agar + DNA (1 or 10 µg) -> 37°C 250 rpm 0.7, 5 or 10% CO2 -> 30 min -> DNase I -> 2.5 hr -> plated on selective medium -> 37°C 5% CO2 3 days. Co-cultivation: cells prepared as described above. 1 to 5E08 cfu/ml + Bolton broth -> 37°C 250 rpm 0.7, 5 or 10% CO2 -> each 2 hrs transfer onto Bolton agar -> 37°C 250 rpm 0.7, 5 or 10% CO2 -> 30 min -> DNase I -> plated on selective medium -> 37°C 5% CO2 3 days.	-	-	-	de Boer, 2002	
Frigaard, N.U.	Chromosomal Gene Inactivation in the Green Sulfur Bacterium Chlorobium tepidum by Natural Transformation	Applied and Environmental Microbiology	2001	Naturally competent	ON recipient cells in late exponential growth phase (3 to 6E09 cells/ml) harvested -> resuspended in CL medium with 1 µg DNA -> spotted on nonselective CP plate -> jar, dark 1-2 hr -> light 40°C 18-20 hr -> suspended in CL -> plated on selective medium -> 5 to 6 days	-	-	-		
Mattimore, V.	Genetic Characterization of Forty Ionizing Radiation-Sensitive Strains of Deinococcus radiodurans: Linkage Information from transformation	Journal of Bacteriology	1995	Naturally competent	Recipient cells in the exponential growth phase -> TGY plate -> 4 to 6 hr 30°C ->dotting 3 to 7 µg DNA -> 24 hrs -> TGY agar -> selection -> 5 days	-	-	-		Deinococcus radiodurans was previously known as Micrococcus radiodurans
Tirgari, S.	Transformation in Micrococcus radiodurans : Measurement of Various Parameters and Evidence for Multiple, Independently segregating Genomes per Cell	Journal of General Microbiology	1980	Naturally competent	DNA in ice + recipient cells -> 30°C gentle shaking water bath -> 90 min -> + DNase -> 15 min -> diluted in TGY -> petri dish and recovered with TGY agar -> 30°C -> second layer of TGY with antibiotic (maximum freq if added after 2 to 3 hr)-> 30°C 4days	-	-	-		
Kennan, R.M.	The Type IV Fimbrial Subunit Gene (fimA)of Dichelobacter nodosus Is Essential for Virulence, Protease Secretion, and Natural Competence	Journal of Bacteriology	2001	Naturally competent	ON broth culture of recipient cells harvested -> fresh EYE broth -> + DNA (5 µg) -> room temperature, anaerobic environment 4 hrs -> + EYE broth -> ON 37°C anaerobically -> plated on selective media -> 37°C anaerobically 7 days.					
Villar, M.T.	Role of the Eikenella corrodens pilALocus in Pilus Function and Phase Variation	Journal of Bacteriology	2001	Naturally competent	Recipient cells grown on supplemented chocolate agar plates aerobically at 35°C -> 48 hrs -> harvested and resuspended in medium A (2.5E06 cfu/ml) -> + DNA (1 µg) -> 30°C 45 min -> supplemented chocolate agar -> 35°C 8 hr -> plated on selective media -> 72 hr	-	-	-		

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Kristensen, B.M.	Natural Transformation of Gallibacterium anatis	Applied and Environmental Microbiology	2012	Naturally competent	Recipient cells grown ON -> 1:100 in BHI -> 37°C shaking until OD600 = 0.2 -> M-IV medium -> 37°C 100 min -> + DNA (1 µg/ml chromosomal DNA or circular plasmid; 0.5 µg/ml linearized plasmid) -> + BHI ->1 h 37°C -> diluted and palted on slective media. When circular plasmid was used: competent cells in 30% glycerol	-	-	-		
Dargis, M.	Modification in Penicillin-Binding Proteins during In Vivo Development of Genetic Competence of Haemophilus influenzae Is Associated with a Rapid Change in the Physiological State of Cells	Infection and Immunity	1992	Naturally competent	In vivo: recipient cells ON culture -> supplemented BHI broth -> OD600 = 0.4 (exponential growth) -> washed in saline -> + DNA (1 µg per 0.5 ml cells) -> placed in diffusion chambers for implantation in rats -> plated on selective plates	-	-	-		
Enne, V.I.	Sulfonamide Resistance in Haemophilus influenzae Mediated by Acquisition of sul2 or a Short Insertion in Chromosomal folP	Antimicrobial Agents and Chemotherapy	2002	Naturally competent	M-IV method as described by Herriot et al, 1970 (but this reference only explains how to prepare the media that allow quick growth of the cells and reach 10^{10} cells in stationary phase)	-	-	-		
Takahata, S.	Horizontal Gene Transfer of ftsI, Encoding Penicillin-Binding Protein 3, in Haemophilus influenzae	Antimicrobial Agents and Chemotherapy	2007	Naturally competent	Cells were grown on chocolate II agar (independently) -> suspended in supplemented BHI broth (independently)-> 37°C 1 hr -> cells mixed at a ratio 1:1 + DNase I -> 37°C 2 hrs -> plated on selective plates.	ftsI	-	homologous recombination		
Gromkova, R.C.	Genetic Transformation in Haemophilus parainfluenzae Clinical Isolates	Current Microbiology	1998	Naturally competent	recipient cells in BHI broth supplemented with NAD -> placed in plastic petri dishes -> 18 hr 37°C without shaking -> competent cells added to supplemented BHI + DNA -> 5 hrs 37°C with shaking -> plated on selective media -> 18 hr 37°C.	-	-	-		Clumping of the recipient strains during growth possibly interferes with DNA transformation.
Sisco, K.L.	Sequence-specific DNA uptake in Haemophilus transformation	PNAS	1979	Naturally competent	Recipient cells incubated in MIV medium to induce competence -> BHI broth with 0.1 ml of 10^{10} recipient cells/ml + DNA (0.1 µg) ->37°C 5 min -> 0°C + Dnase -> 20 min -> washed and resuspended.	-	-	-		
Biggas, A.	Development of a genetic manipulation system for Haemophilus parasuis	Veterinary Micrpbiology	2005	Naturally competent	Recipient cells grown on chocolated blood plates ON 37°C -> resuspended in TYE broth (OD660 = 2 -> 10^{10} cfu/ml) -> 1/10 dilution spotted and spreaded (10 mm diameter) onto prewarmed chocolated blood plate -> + cAMP + DNA ->37°C -> cells resuspended and plated onto selective plates -> 37°C 2 days	-	-	-		Maximum frequency of transformation was achieved with 8 mM cAMP and no transformants were obtained in the absence of cAMP. The best transformation frequency is achieved when the DNA is incubated together with the cAMP and cells for 5 hrs.
Baltrus, D.A.	Multiple phases of competence occur during the Helicobacter pylori growth cycle	FEMS Microbiology Letters	2005	Naturally competent	Recipient cells grown on brucella broth supplemented with FBS -> 150 rpm -> + DNA (1 -plasmid and PCR product - or 10 - genomic - µg) each 4 hours (DNA is always added to a different culture) -> plated on selective blood agar.	-	-	-		
Jiesong, H.	Helicobacter pylori acquistion of metronidazole resistance by natural transformation in vitro	World Journal of Gastroenterology	1998	Naturally competent	Recipient cells grown on supplemented BHI broth ->37°C, 24 hrs, microaerobic conditions -> + DNA (50 µg) -> 37°C 6 hrs -> plated on selective agar	-	-	-		
Hofreuter, D.D.	Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system	Molecular Microbiology	2001	Naturally competent	recipient cells from serum plates -> BHI with 10% FCS (OD550 = 0.1) -> + DNA -> 4-6 hr microaerophilic conditions 37°C ->plated on selective medium	-	-	-		

Bibliography				COMPETENCE		Mosaic genes				
1. Author	**Title**	**Journal**	**Year**	**Competence** (inducible, constitutive)	**Environmental conditions** (to induce competence)	**Mosaic Gene(s)** (detected in the corresponding organism)	**incorporated fragment lengths**	**Recombination** (information on involved mechanisms and components)	**Associated publications**	**Comments**
Noto, J.M.	Genetic Manipulation of a Naturally Competent Bacterium, Helicobacter pylori	Methods in Molecular Biology	2012	Naturally competent	Solid: ON recip cells grown in TSA blood plate -> harvested and resuspended in Brucella broth -> cells spotted onto TSA blood plate -> 1 µg plasmid DNA ->37°C ON 5% CO2 -> transfer to selective plates -> 5 days. Liquid: Inoculation of H. pylori in Brucella broth (OD600 = 0.1-0.2) -> 1-2 µg plasmid DNA -> 16-18 hr 37°C 5% CO2 -> OD600 = 1.5-2.0 ->harvest and resuspend cells in Brucella broth -> plate right-side-up on selective plates ->37°C 5% CO2 -> upside down 37°C 5% CO2 5 days.	-	-	-		Paper with protocol for H. pylori transformation
Israel, D.A.	Characteristics of Helicobacter pylori natural transformation	FEMS Microbiology Letters	2000	Recipient strains naturally competent	solid: harvested colonies -> resuspended in PBS + DNA. Overnight incubation on non-selective plates - transformant counting liquid: PBS-washed cells incubated in Brucella broth + DNA addition	-	-	-		
Scocca, J. J.	Specificity in deoxyribonucleic acid uptake by transformable Haemophilus influenzae	Journal of Bacteriology	1974							
Moccia, C.	The nucleotide excision repair (NER) system of Helicobacter pylori: role in mutation prevention and chromosomal import patterns after natural transformation	BMC Microbiology	2012	naturally competent			1.3-3.8 kb + interspersed sequences		Kulik, 2008	
Kulick, S.	Mosaic DNA Imports with Interspersions of Recipient Sequence after Natural Transformation of Helicobacter pylori	PLoS ONE	2008	naturally competent			1.3-3.8 kb + interspersed sequences (82 bp) recombination import: 417 bp fragments of several 100 bp	far more common in H. pylori than in other bacteria. Numerous recombination events (single patient, samples recovered over a year)	Moccia, 2012	special feature of mosaic gene: short interspersed sequences; relevant for transformation/recombination

www.bmg.gv.at