

The impact of phages on bacterial chromosomes: a genomic-based analysis

Doctoral Thesis

Author(s):

Canchaya Sanchez, Carlos A.

Publication date:

2004

Permanent link:

<https://doi.org/10.3929/ethz-a-004829261>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Doctoral Thesis ETH No. 15651

**THE IMPACT OF PHAGES ON BACTERIAL CHROMOSOMES:
A GENOMICS-BASED ANALYSIS**

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Natural Sciences

Presented by

CARLOS A. CANCHAYA SANCHEZ
Bachiller en Ciencias – Biología, UNALM
DES en Bioinformatique - FUNDP
Born April 27th, 1971
Citizen of Peru

Accepted on the recommendation of

Prof. Dr. Martin Loessner, examiner
Prof. Dr. Wolf Dietrich Hardt, co-examiner
Dr. Harald Brüssow, co-examiner

Zürich 2004

Summary

Bacteriophages have played, and continue to play a central role in bacterial genetics and molecular biology. Historically, they were used to define the nature of the gene and early studies of gene regulation. Phages were also the first genomes to be sequenced. However, phage research was not leading the genomics revolution, which was clearly dominated by bacterial genomics. It was realized that an important underused source for phage sequences are bacterial genome projects. Microbial genomics has shown that a substantial amount of bacterial DNA is represented by DNA of prophages and phage-like elements. A detailed database mining of these sequences gave us further insights into the role of prophages in the bacterial genome and their possible interactions.

To assure their ecological and evolutionary success, prophages have to encode genetic functions that are of selective advantage to the lysogenic host. One of these genes is the phage immunity functions. It comprises two groups of genes: the *ci*-like repressor that protects the cell against superinfection with temperate phages sharing identical or related repressor DNA recognition sites, and super infection exclusion genes. Candidates for both genes were identified in our model *S. pyogenes* phage SF370.1. The genome comparison of the three complete prophages of the *S. pyogenes* strain SF370 shows other gene candidates of possible selective advantage to the host: hyaluronidase, DNase, toxins, mitogenic factors, superantigens. They are mostly localized downstream of the lysin gene next to the *attR* in a segment called lysogenic conversion region.

The SF370 strain shows a broad range of prophage-host genome interaction: from a fully inducible prophage over biologically inactive, but apparently complete prophage genomes to a prophage remnant that has lost substantial amounts of the prophage genome and finally isolated phage genes. This series of events fits with theoretical predictions of the arms race between phage and bacterial genomes. One might speculate that this pattern also represents a time series. Under this view the inducible prophage represents a recent prophage acquisition.

Experimental work was performed to elucidate if the genes are silent or are actively transcribed in *L. johnsonii* and *L. plantarum*. Genes with sequence similarity to candidate lysogenic conversion genes from *S. pyogenes* prophages were identified in *L. plantarum* prophages Lp1 and Lp2 as well as in *L. johnsonii* NCC533 prophages Lj965 and Lj928. These genes were localized between the phage integrase and repressor gene on one side and between the phage lysin and *attR* on the other. Further extra genes were an array of four tRNAs and an orf between the lysin gene and *attR*.

Additional experiments were done to study the contribution of phages to the individuality of the bacteria. Microarray studies were performed using the genome of *L. johnsonii* NCC533 strain. The results showed that the DNA of phages is not present in any other strain of *L. johnsonii* species. They constitute a high proportion of the DNA unique to the sequenced NCC533 strain. Prophages make then an important contribution to the individuality of bacteria.

The role of prophages for the lysogen might go beyond the contribution of genes. The three prophages from *S. pyogenes* strain SF370 shared high DNA sequence identity over the putative tail fiber genes and DNA sequence identity was even more pronounced between inducible prophages from two different *S. pyogenes* lysogens (SF370 and NIH1.1). Regions of DNA identity would be sites for homologous recombination between prophages.

Some bacterial species are represented by more than one genome in the public database. To identify the main contributors to the individuality of genomes belonging to the same species we compared their genomes by aligning them. In some cases, the alignment produced the straight line indicating perfectly aligned genomes, in other cases the straight line of the dotplot is interrupted by gaps.

Tracing back the history of rearrangements in *Xylella* identified prophages as the most probable regions for chromosome recombination. Postulating three prophage-mediated rearrangement events, we reconstructed a linear alignment of the two *Xylella* genomes. Prophages could then be considered as "hotspots" of chromosomal rearrangements, increasing in this way the diversification of both host and prophage.

Zusammenfassung

In der vorliegenden Arbeit wurde die Rolle von Prophagen in bakteriellen Genomen untersucht. Prophagen sind die DNA von Bakterienviren, die in das bakterielle Chromosom integriert wurden. Dieses Phänomen wurde zuerst an Hand von drei Bakterienstämmen untersucht. Diese Stämme repräsentierten ein bakterielles Pathogen und zwei Kommensalstämmen des Alimentationstraktes. Der pathogene Vertreter war *Streptococcus pyogenes* mit dem von einer menschlichen Wunde isolierten Stamm SF370. Dieser hochinvasive Stamm enthält einen vollständigen, induzierbaren Prophagen, zwei nicht-induzierbare Prophagen und einen unvollständigen Prophagen-Rest. Der induzierbare Prophage kodiert-zusätzlich zu den Phagen-Genen, die in allen Sfi11-artigen temperenten Siphophagen gefunden werden - auch für drei Virulenzfaktoren. Diese waren eine Hyaluronidase, Exotoxin C und ein mitogener Faktor. Alle drei Faktoren spielen eine Rolle in der Pathogenese von invasiven Streptokokken Infektionen. Der Prophage-Rest besteht aus dem Lysogenie und DNA Replikations Modul. Alle übrigen Phagen Gene wurden durch Deletion verloren. Dieser unvollständige Prophage enthält keine erkennbaren Virulenzgene, aber seine Integration in das bakterielle Genom trennt die funktionell zusammenarbeitenden DNA Reparatur-Gene *mutS* und *mutL*, was auf die genetische Stabilität von dem Bakterienstamm Auswirkungen haben kann. Die zwei Kommensalstämmen kamen aus der Gattung Laktobazillus. *Lactobacillus plantarum* Stamm WCFS1 wurde aus der Mundhöhle, *Lactobacillus johnsonii* Stamm NCC533 wurde aus dem Stuhl isoliert. Beide Laktobazillen Stämme enthalten multiple Prophagen, die sich jedoch als nicht induzierbar erwiesen. Der Grund ist ersichtlich für die *L. plantarum* Prophagen: sie sind entweder unvollständig, das Resultat von Inter-Prophage Rekombinationen oder sie enthalten mutierte essentielle Prophagen Gene, wie z. B. Prophage Lp2, der einen Stopp Kodon mitten in seinem Terminase-Gen aufwies. Wenn die *L. plantarum* Prophagen auch keine Virulenzgene aufwiesen, so enthielten sie jedoch Gene, die noch Sequenz-Verwandtschaft zu potentiellen Virulenzfaktoren in *S. pyogenes* Prophagen aufwiesen. Diese Gene gehörten auch zu den wenigen Prophagen Genen, die von dem Prophagen exprimiert wurden. Ein recht ähnliches Bild ergab sich für die Prophagen von *L. johnsonii*. Diese Beobachtungen führten zu einer Arbeitshypothese, wonach Prophagen eine Rolle in der Anpassung ihres Wirtsstammes in seiner ökologischen Nische spielen. Mikroarray Daten zeigten, dass Prophagen-DNA etwa die Hälfte der Stamm-spezifischen DNA in *L. johnsonii* ausmachen. Nach diesen Vorarbeiten wurde eine Gesamtanalyse der zur Zeit in Datenbanken gespeicherten komplett sequenzierten bakterieller Genome vorgenommen. Ein automatisches Prophagen-Suchprogramm wurde entwickelt, die identifizierten Prophagen wurden annotiert und mittels vergleichender Genomanalyse untersucht. Prophagen stellten sich als ein wesentlicher Teil vieler bakteriellen Genome heraus. Ihre Bedeutung geht jedoch darüber hinaus. Prophagen treiben offensichtlich die Evolution einer Reihe von bakteriellen Pathogenen. Zwei Situation konnten unterschieden werden. Prophagen kodieren für entscheidende krankmachende Toxine und Prophagen bringen eine Vielzahl von Faktoren bei, die in ihrer Kombination zu pathogenen Bakterien führen. Darüberhinaus vermitteln Prophagen auch chromosomale Inversionen und Rearrangierungen des genetischen Materials der Bakterien. Die bioinformatische Analyse zeigte ebenfalls, dass Phagen ein Motor für horizontalen Gentransfer zwischen Bakterien sind.