Regulatory responses of Streptococcus pneumoniae to varying metal ion- and nitrogen availability
Kloosterman, Tomas

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 3

Development of Genomic Array Footprinting for Identification of Conditionally Essential Genes in *Streptococcus pneumoniae*

Jetta J.E. Bijlsma#, Peter Burghout#, Tomas G. Kloosterman#, Hester J. Bootsma#, Anne de Jong, Peter W. M. Hermans and Oscar P. Kuipers

Based on *Appl. Environ. Microbiol.* 73:1514-1524

# these authors contributed equally to the work described in this paper
Development of Genomic Array Footprinting

Abstract

*Streptococcus pneumoniae* is a major cause of serious infections such as pneumonia and meningitis in both children and adults worldwide. Here, we describe the development of a high-throughput genome-wide technique, genomic array footprinting (GAF), for the identification of genes essential for this bacterium at various stages during infection. GAF enables negative screens by means of a combination of transposon mutagenesis and microarray technology for the detection of transposon insertion sites. We tested several methods for the identification of transposon insertion sites and found that amplification of DNA adjacent to the insertion site by PCR resulted in nonreproducible results, even when combined with an adapter. However, restriction of genomic DNA followed directly by *in vitro* transcription circumvented these problems. Analysis of parallel reactions generated with this method on a large *mariner* transposon library, showed that it was highly reproducible and correctly identified essential genes. Comparison of a *mariner* library to one generated with the *in vivo* transposition plasmid pGh9:ISS1, showed that both have an equal degree of saturation, but that 9% of the genome is preferentially mutated by either one. The usefulness of GAF was demonstrated in a screen for genes essential for surviving Zn\(^{2+}\) stress. This identified a gene encoding a putative cation efflux transporter, and its deletion resulted in an inability to grow under high Zn\(^{2+}\) conditions. In conclusion, we developed a fast, versatile, specific, and high-throughput method for the identification of conditionally essential genes in *S. pneumoniae*.

Introduction

Infection with the Gram-positive pathogen *S. pneumoniae* is a worldwide cause of mortality and morbidity. Carriage of the bacterium can be asymptomatic, but often progresses to diseases such as sinusitis and otitis media and to more serious infections such as pneumonia, sepsis and meningitis. It is estimated that over a million people die of *S. pneumonia*‐related diseases every year (40,235). Current polysaccharide vaccines are effective but only protect against a fraction of the 90 serotypes known, and replacement and disease by nonvaccine serotypes is already observed (294). Treatment of *S. pneumoniae* infections is also confounded by the rise of strains resistant to the most commonly used antibiotics (193). Thus, there is an urgent need for the identification of new protein leads for the development of vaccines and antimicrobial drugs, preferably by using high throughput genome-wide screening methods.

Several methods have been used to determine which genes are needed by *S. pneumoniae* in the various niches it occupies in the host (conditionally essential genes), such as transcriptome analysis (243), differential fluorescence induction (207) and signature tagged mutagenesis (STM) (123,181,254). Of these methods, STM is the only one that enables negative screening and thus directly addresses the essentiality of a gene under a certain condition. However, STM is hampered by various drawbacks: every mutant has to be grown and stored separately, which makes it hard to generate large libraries with a more than 1x coverage of the genome; detection of the tags is cumbersome and identification of transposon insertion sites is labor intensive (17). In the three *in vivo* STM screens performed with *S. pneumoniae*, no more than 8,000 mutants in total have been tested. Furthermore, there is little overlap in the identified genes in each study, indicating that the screens were not saturated and more importantly, that not all existing conditionally essential genes have been identified (123,181,254).
Therefore, we set out to develop genomic array footprinting (GAF), a high throughput method to identify conditionally essential genes in *S. pneumoniae*, using a combination of random transposon mutagenesis and microarray technology (Fig. 1). GAF detects the transposon insertion sites in a library by amplifying and labeling the chromosomal DNA adjacent to the transposon and subsequent hybridization of these probes to a microarray. Identification of transposon insertion sites that have disappeared from the library due to selection, which represent conditionally essential genes, is achieved by differential hybridization of the probes generated from the library grown under two conditions to an array (Fig. 1). We anticipated that the most critical step in the whole procedure is the specific amplification and detection of the DNA adjacent to the transposon insertion sites, as this determines the rate of false positives. Recently, several methods to identify transposon insertion sites using microarrays have been described (55, 275, 279). In this paper, we describe (i) the pitfalls of several methods of amplification when applied to *S. pneumoniae* libraries; (ii) the successful development and validation of a specific and reproducible method for the detection of transposon insertion sites in *S. pneumonia*. Finally, we show with a biological screen for genes essential for survival of Zn\(^{2+}\) stress that GAF is a reliable and effective method to identify conditionally essential genes.

**Figure 1.** Schematic representation of the genomic array footprinting procedure (GAF). A large *S. pneumoniae* transposon library is grown under nonselective and selective conditions. Subsequently, chromosomal DNA containing transposons (gray rectangle) with outward-facing T7 RNA polymerase promoters (arrow with T7) is isolated from each population. The DNA is digested and the DNA adjacent to the transposon insertion site is amplified using *in vitro* transcription with T7 RNA-polymerase. The RNA is used in standard microarray procedures. Hybridization of these probes to a microarray will reveal which genes were disrupted in the mutants that disappeared during selection; those spots to which only material derived from the nonselective condition will hybridize (gray spots).

**Materials and Methods**

**Bacterial strains, growth conditions, DNA isolation and manipulations**

*S. pneumoniae* strains used in this study were the serotype type 2 isolate D39 (19) and its unencapsulated derivate R6 (133). *Lactococcus lactis* strain NZ9000 (161) was used to generate the pGH9T7:ISS1. *Escherichia coli* DH5αF was used as host for the pR412 plasmid and the generation of pR412T7. All strains were routinely cultured in TY broth or M17 broth containing glucose (GM17) as described previously (161). General DNA techniques were performed as described previously (161). Sequencing of the DNA adjacent to the insertion sites was performed using a single primer PCR essentially as described (152). An in-frame deletion of *czcD* (*SP1857*) was constructed using pORI280 as described before (161) with primer pairs *czcD*-KO-1/*czcD*-KO-2 and *czcD*-KO-3/*czcD*-KO-4 (Table 1), which removed 852 bp of the ORF; a detailed description of its construction is provided elsewhere (163).
Table 1. Primers used in this study. All primers were designed for this work, except CEKG2A and CEKG4, which were designed by N.R. Salama and co-workers (275).

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5’-3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBMrIRPi</td>
<td>AGACCGGGGACTTATCAGCC</td>
</tr>
<tr>
<td>PBMrStartT7</td>
<td>CCTATATGGAGTAGTTAGACACATAGATGGCGTCGC</td>
</tr>
<tr>
<td>PBMrEndT7</td>
<td>CCTATATGGAGTAGTTAGATACACATAGGTCCCTTCAAG</td>
</tr>
<tr>
<td>ISSsor</td>
<td>ATTTGAAAAAGGCAGCGCCAGTTGATATATCTCGCTGTC</td>
</tr>
<tr>
<td>MATT 1</td>
<td>AAGGCGATGGAACAAATTCGGG</td>
</tr>
<tr>
<td>MATT 2</td>
<td>GGAACGCTTCCGGATTTCGG</td>
</tr>
<tr>
<td>CEKG2A</td>
<td>CGCCACCGGTCGACTAGTACNNNNNNNNNAGAG</td>
</tr>
<tr>
<td>CEKG2D</td>
<td>CGCCACCGGTCGACTAGTACNNNNNNNNNATAG</td>
</tr>
<tr>
<td>CEKG2E</td>
<td>CGCCACCGGTCGACTAGTACNNNNNNNNNNAAT</td>
</tr>
<tr>
<td>CEKG4</td>
<td>CGCCACCGGTCGACTAGTAC</td>
</tr>
<tr>
<td>LinkAmp1</td>
<td>AGGCACAGTCGAGGACTTTATCCTA</td>
</tr>
<tr>
<td>LinkAmp2</td>
<td>GCCTCTGAATACCTCTACAACAAGTAC</td>
</tr>
<tr>
<td>PBMrTn5</td>
<td>CGGGAATCAAGGTGTTGGTTAGT</td>
</tr>
<tr>
<td>PBMrTn3</td>
<td>TACTAGCGACGCCATCTATGTC</td>
</tr>
<tr>
<td>TMr_1</td>
<td>TGCTTTAATACTAGCGCCATCTATGTGTC</td>
</tr>
<tr>
<td>ISS1_T7_up</td>
<td>CGATAAGCTTTAGAGAGAAATGGTTCTGTGCAAGAGTTTCTGAT</td>
</tr>
<tr>
<td></td>
<td>AAGTCTATTTTAGTGTAAATGAATACCTCCTATAGTGAGTCGA</td>
</tr>
<tr>
<td></td>
<td>TTAATAATACGGAGGATATACATCAAGT</td>
</tr>
<tr>
<td>ISS1_T7_down</td>
<td>CAGGAATTCATGCTCTAGACCCATCTCTGTGCAAGAGTTT</td>
</tr>
<tr>
<td></td>
<td>AAAATCAATAACAGGTGCTCTCTATAGTGACTGCTATTATTTAATC</td>
</tr>
<tr>
<td></td>
<td>CTTCTGTATCTAAAGCAATATTCC</td>
</tr>
<tr>
<td>Adapter1</td>
<td>AATTACCACGCCA</td>
</tr>
<tr>
<td>Adapter2</td>
<td>CGACCACGCCA</td>
</tr>
<tr>
<td>Adapter3</td>
<td>AGTCTCGCGATGTAAGGGTTCTGTGCAAGAGTTTCTGAT</td>
</tr>
<tr>
<td>Adapter4</td>
<td>GTCCAGTCTCAGCAAGTATAAGGGTTCTGTGCAAG</td>
</tr>
<tr>
<td>czcD-KO-1</td>
<td>TGCTCTAGAAGGTCAATGCTCTGAAAG</td>
</tr>
<tr>
<td>czcD-KO-2</td>
<td>AGCATATTTTGCCCTCATATTTC</td>
</tr>
<tr>
<td>czcD-KO-3</td>
<td>ATATGAAAGCAGACTTGTATGATGACGCAACATATTAG</td>
</tr>
<tr>
<td>czcDKO-4</td>
<td>GAAGATCTCCTGTAGCTGAGACAAGGC</td>
</tr>
</tbody>
</table>

Construction of pGh9:ISS1 and pR412 derivates with outward facing T7 RNA polymerase promoters in their transposons

A pGh9:ISS1 derivative with two outward facing T7 promoters in the ISS1 element was constructed as follows. The ISS1 element from pGh9:ISS1 (203) was removed by digestion with EcoRI and HindIII and replaced by a HindIII, EcoRI digested PCR product of the ISS1 element generated on pGh8:ISS1 (203) with primers ISS1_T7_up (Table 1) and ISS1_T7_down (Table 1). This resulted in pGh9T7:ISS1 which contains the ISS1 element (with the T7 promoters on both sides) in an orientation opposite compared to that in the pGh9:ISS1 plasmid.

The backbone of transposon donor plasmid pR412 (209) including the ILR and IRR of the transposon, was amplified with Pwo DNA polymerase (Roche) with a single phosphorylated primer PBMrIRPi (Table 1); PCR cycling conditions were as follows: 94°C for 4 min; 30 cycles of 94°C for 1 min, 55°C for 1 min, 68°C for 5 min; 68°C for 5 min. The spectinomycin-resistance cassette was PCR amplified with primers containing a T7-tag, i.e., PBMrStartT7 (Table 1) and primer PBMrEndT7 (Table 1); PCR cycling conditions were as follows: 94°C for 4 min; 30 cycles of 94°C for 1 min, 55°C for 1 min, 68°C for 1.5 min; 68°C for 4 min. The two PCR products were ligated to obtain the transposon donor plasmid pR412T7.

Generation of transposon libraries

Random transposon libraries were routinely generated in S. pneumoniae as described before, using the in vitro mariner transposon (5) and the in vivo pGh:ISS1 system (203).
Recombinant HimarC9 transposase was purified from the *E. coli* strain BL21 (DE3) (Novagen) containing plasmid pET29C9 (174) essentially as described before (175). However, to prevent DNAse contamination of the transposase stock, the DNAseI treatment of the purification step was omitted from the protocol. For *in vitro* transposition, 1 µg of purified chromosomal DNA was incubated with 0.5 µg pR412T7, which contains the *marinerT7* transposon, and 0.5 µg purified recombinant HimarC9 transposase in a 20 µl reaction. After repair of the transposition reaction with T4 DNA polymerase and *E. coli* DNA ligase (123), the DNA was used for transformation of strain R6 (5,174). After overnight growth on selective plates containing 120 mg L⁻¹ spectinomycin, colonies were scraped, pooled and stored at -80°C. For construction of pGh9T7:ISS1 mutant libraries (203), the R6 and D39 strains were transformed with the pGh9T7:ISS1 plasmid and grown at 30°C on selective plates containing 0.25 mg L⁻¹ erythromycin. Several colonies of *S. pneumoniae* containing the pGh9T7:ISS1 plasmid were pooled, grown overnight in GM17 broth with erythromycin at 30°C and stored at –80°C. From this stock, bacteria were inoculated for overnight growth in GM17 broth at 30°C with the appropriate antibiotic, the next day the cultures were diluted 1 to 50 in GM17 broth without antibiotics and after three hours of growth shifted to 37°C. After another three hours of growth at the non-permissive temperature, 10⁻³ and 10⁻⁴ dilutions were plated on selective medium. Approximately 20,000-30,000 CFU were scraped from plate, pooled and stored at –80°C for further use. The same *marinerT7* and pGh9T7:ISS1 transposon libraries were used for all experiments to determine the reproducibility and specificity of each method. New libraries were generated for the comparison of the *marinerT7* and pGh9T7:ISS1 transposon libraries and the GAF Zn²⁺ screen.

**Microarray construction, labeling, hybridization and analysis**

*S. pneumoniae* microarrays were produced essentially as described (170,332) and contained amplicons of on average 600 bp representing 2,087 ORFs of *S. pneumoniae* TIGR4, each of which is present twice on the array (317) as an internal control to monitor variation in hybridization efficiency per slide. Purified RNA obtained after *in vitro* transcription was used to generate fluorescent DNA probes by direct or indirect labeling using standard methods. Hybridization, washing and acquisition of array images were performed as described previously (162). Spots were screened visually to identify those of low quality. The amount of low quality spots varied per slide but was never more than 10 %; they were removed from the dataset prior to analysis. Slide data were processed and normalized using MicroPreP (333). A net mean intensity filter based on hybridization signals obtained with amplicons representing open reading frames unique for *S. pneumoniae* strain TIGR4 (317), was applied in all experiments. For one-array experiments (Fig. 2 and 3) only those genes were analyzed that had reliable measurements (*i.e.*, passed filters) for both duplicate spots present on the array. For the other experiments, microarray data were analyzed when at least 3 measurements were available, using a CyberT implementation of the Student’s *t*-test (198). False discovery rates were calculated as described previously (332). Correlation between normalized signal intensities per gene in both channels was determined by linear regression analysis.

**Identification of transposon insertion sites using Micro Array Tracking of Transposon mutants (MATT)**

For the identification of transposon insertion sites, MATT was performed as described previously (275) with some modifications. A linear PCR with AccuPrime (Invitrogen) polymerase was performed on chromosomal DNA containing pGh9T7:ISS1 insertions as
Development of Genomic Array Footprinting

template and the MATT1 (Table 1) primer located on the transposon, for 30 cycles at 56 °C. The single-stranded fragments were purified with a GFX column (GE Healthcare, life sciences) and 100 ng was used as template for a semi-random PCR in a 20 µl reaction containing 2 µM each of an anchored random primer (CEKG2A, CEKG2D, or CEKG2E, Table 1), MATT1 and 1 U Taq polymerase (Invitrogen) with the following cycling reactions: 1 cycle at 94 °C for 4 min; 12 cycles at 94 °C for 30s, 42 °C for 30 s (- 0.5 °C each cycle) and 72 °C for 2 min; and 30 cycles at 94 °C for 30s; 65 °C for 30 and 72 °C for 2 min. The reaction products were diluted five- fold and 1 µl was used as template for a subsequent reaction using a nested transposon specific primer Matt2 (Table 1) and the anchor primer CEKG 4 (Table 1) in a 100 µl reaction using the following cycling conditions: 1 cycle at 94 °C for 4 min; 30 cycles at 94 °C for 30s; 65 °C for 30 and 72 °C for 2 min. The PCR products were purified using the Roche PCR purification kit and used as input for a T7 RNA polymerase reaction using the T7 MEGAscript kit (Ambion).

Identification of transposon insertion sites using a TOPO adapter

Genomic DNA was diluted in 20 µl H2O to 0.5 µg/µl and fragmented by sonication with the microtip of a Branson digital sonifier (8 times for 0.5 s, output 20%). To prepare double-stranded, blunt end DNA, fragmented DNA was successively treated with S1 nuclease, T4 DNA polymerase, and E. coli DNA ligase. To facilitate the ligation of a TOPO-isomerase activated linker (TOPO-Walker Kit, Invitrogen), the blunt-end double-stranded DNA was incubated with calf intestinal phosphatase for dephosphorylation and with Taq DNA polymerase (Thermoperfect, Integro) to create 3’ A-tailed DNA-ends. Ligated DNA was used as template in two consecutive PCR reactions with Taq DNA polymerase: the first reaction contained primer s LinkAmp1 (Table 1) and primer PBMrTn5 (Table 1), the second nested PCR primers LinkAmp2 (Table 1) and PBMrTn3 (Table 1). PCR cycling conditions for both reactions were as follows: 94°C for 4 min; 20 cycles of 94°C for 30s, 72°C for 30 s and 72°C for 1 min; 72°C for 1.5 min; 72°C for 3 min. PCR products were loaded on a 2% agarose gel and DNA fragments with the preferred size were excised and extracted with a Qiaquick gel purification kit (Qiagen). In vitro transcription reaction and subsequent DNAseI digestion were performed with the T7 MEGAscript kit (Ambion).

Identification of transposon insertion sites using TraSH

TraSH was performed essentially as described previously (279) with a few modifications. Genomic DNA was digested with ApoI or TaqI, precipitated and ligated overnight in a 15 µl volume with an approximately 1,000-fold molar excess of the appropriate adapter (Adapter1 for Apol- and Adapter2 for Taql- digested DNA, Table 1). The ligated products were purified and 0.3 to 1 µl was used as template in the following PCR reaction containing primer TmR1 (Table1): 1 cycle at 95 °C for 1 min; 5 cycles at 95 °C for 30s, 72 °C for 30 s and 72 °C for 1 min; subsequently the appropriate adapter primer (Adapter3 for Apol- and Adapter4 for Taql-digested DNA, Table1) was added, and the reaction was continued with the following cycling conditions: 5 cycles at 95°C for 30s, 72°C for 30 s and 72°C for 1 min and 30 cycles at 94°C for 30s; 65°C for 30 (- 0.3°C each cycle) and 72°C for 70 s. PCR products were purified and used as template for a T7 RNA polymerase reaction using the MEGAscript kit (Ambion).

Identification of transposon insertion sites using the T7 RNA polymerase promoter

The procedure was essentially performed as described previously (55). Genomic DNA (>10 µg) was digested overnight with TaqI, DdeI or Alul and the purified DNA was
Chapter 3

used for an in vitro transcription reaction with the MEGAscript T7 RNA polymerase kit (Ambion) in a total volume of 20 μl.

Selection for genes essential for survival of high Zn\(^{2+}\) conditions.

Aliquots of a marinerT7 transposon library generated in S. pneumoniae R6 containing approximately 20,000 independent mutants were diluted 1 to 20 in GM17 broth or in GM17 broth containing 0.5 mM ZnSO\(_4\). Both cultures were grown for 10 generations as determined by OD\(_{600}\) after which the bacteria were spun down (9,000g, 4 min) and used for the extraction of genomic DNA.

Microarray data accession numbers

The microarray data have been deposited in the NCBI Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/ under GEO Series accession numbers GSM158348, GSM158350, GSM158558, GSM158562, GSM158563, GSM158564, GSM158565, GSM158566, GSM158567, GSM158568, GSM158569, and GSM158570.

Results

Prerequisites for the development of GAF

Important considerations for the development of GAF were that the method should be sensitive enough to detect all mutants in a library and should be reproducible. Reproducibility, specificity and sensitivity are in large part determined by the amplification and detection step (Fig. 1). Therefore, we tested the performance of various methods for transposon insertion site detection (55,275,279) on both small (<100 CFU) and large (>20,000 CFU) mutant libraries.

Although we used strains R6 and D39 throughout this study, we wanted our method to be applicable to (in principle) all S. pneumoniae strains, some of which are not as easily transformable. Therefore we established two methods for the generation of random mutant libraries; in vitro mariner mutagenesis (5), which depends on a high transformation efficiency of the accepting strain for the generation of large libraries and is not expected to have polar effects on downstream genes (274), and the in vivo pGh:IS\(_{S1}\) system (203), which consists of a plasmid with a temperature-sensitive origin of replication, the IS\(_{S1}\) insertion element and an erythromycin-resistance gene with a weak terminator (-7.5 kcal mol\(^{-1}\)); thus insertion of pGh:IS\(_{S1}\) could cause some polar effects. Growth of bacteria that contain the plasmid for several generations at 30 °C, the permissive temperature, followed by a temperature shift to 37 °C induces the insertion of the plasmid into the chromosome. Thus, a one-time introduction of this plasmid in a strain with low transformation efficiency is sufficient to generate large (>20,000 CFU) libraries, avoiding the need for high transformation efficiencies in accepting strains. In each transposon an outward facing T7 RNA polymerase promoter close to both ends of the insertion element was introduced for amplification purposes. Both transposons readily generated large (>20,000 CFU) libraries in S. pneumoniae strains R6 and D39 as expected, and the pGh:IS\(_{S1}\) system in strain Tigr4 (34) as well (results not shown). Mutant libraries generated with these transposons were used exclusively throughout this study, and the RNA generated with the aid of the T7 RNA polymerase promoter was used in our standard microarray procedures.

Identification of the transposon insertion sites using MATT

Microarray tracking of transposon mutants combines two rounds of PCR: first, a reaction with specific and non-specific primers, followed by a nested PCR to further amplify
the material (275). The specific primers hybridize to sequences on the transposon and the nonspecific primers contain 10 degenerative nucleotides followed by a unique anchor sequence. Several anchor sequences were chosen that are present on average every 400 bp in the genome. MATT was found to be specific and reproducible when used for artificial GAF experiments on a few mutants (comparison of 12 mutants with the same set in which three mutants were removed; results not shown). However, when two parallel reactions on one large (>20,000 CFU) genome-wide library were compared, it was obvious that the method was not reproducible (Fig. 2A) as few genes displayed equal signal intensities in both channels ($R^2: 0.036$). When the products of the second round of PCR were visualized on an agarose gel, several distinct bands of varying size in each reaction were visible, indicating that some DNA fragments were selected randomly during the PCR steps (results not shown).

**Figure 2.** Detection of transposon insertion sites using PCR based methods. (A) Scatterplot of the signal intensities of two parallel reactions on the chromosomal DNA of a genome-wide library using MATT. There is no correlation between the signal intensities for each gene ($R^2: 0.036$), indicating that the procedure is non-reproducible. (B) Scatterplot of the signal intensities of two parallel reactions on the chromosomal DNA of a genome-wide library using a TOPO-isomerase adapter combined with a nested PCR. There is no correlation between the signal intensities in each gene ($R^2: 0.010$), indicating that the procedure is non-reproducible. (C) Scatterplot of the signal intensities of two parallel reactions on the chromosomal DNA of a genome-wide library using TraSH. Correlation between the signal intensities for each gene indicates that the procedure is fairly reproducible ($R^2: 0.813$).

**Identification of transposon insertion sites using a TOPO-isomerase adapter or the adapter-based TraSH method.**

Next, we tested a method that consists of the ligation of a TOPO-isomerase activated adapter to sheared chromosomal DNA, followed by two PCR steps using nested primers that anneal specifically to the adapter and the transposon. Similar to MATT, this method was
specific and reproducible when applied to artificial GAF experiments on a small number of mutants (i.e., comparison of 21 mutants with the same set from which 6 mutants were removed, results not shown), but clearly nonreproducible when used on a genome-wide library (Fig. 2B). When the PCR products of separate reactions were examined, on agarose gels, different distinct banding patterns were visible, again indicating that the PCR steps introduced a bias in the whole procedure. Repeated attempts to improve the PCR steps in both the MATT and the TOPO-adapter method by using Taq polymerase variants with increased specificity, touchdown PCR, and even more stringent annealing conditions were unsuccessful (results not shown).

**Figure 3.** *In vitro* transcription on purified chromosomal DNA restriction fragments leads to specific and sensitive detection of transposon insertion sites. (A) Scatterplot of the comparison of reactions on the chromosomal DNA of a mutant library consisting of 90 unique *marinerT7* mutants and the DNA of the same mutants to which the chromosomal DNA of 9 defined mutants has been added. Straight lines denote a two-fold difference in signal intensities between the Cy5 and Cy3 channels. Genes identified by sequencing as added to the library are indicated as black squares. Scatterplots of the signal intensities of two parallel reactions on the chromosomal DNA of a genome-wide library digested with TaqI (B), Ddel (C) and AluI (D). Correlation between the signal intensities in each plot is >0.96, indicating that the procedure is highly reproducible.

Subsequently, we tested the TraSH method, which consists of the ligation of an adapter to chromosomal DNA digested with TaqI or ApoI, followed by a short linear PCR with a transposon specific primer, which is then used as template for a PCR reaction with transposon-specific and adapter-specific primers (279). This method was fairly specific and reproducible both when used on a small number of mutants (i.e., comparison of 21 mutants with the same set of mutants from which 6 mutants were removed; results not shown) and on a large (>20,000 CFU) genome-wide library (Fig. 2C), as the majority of the genes had comparable signal intensities in both channels ($r^2$: 0.83). However, the optimal annealing
temperature and template concentration had to be determined for each new sample used in TraSH, as standardized PCR conditions resulted in distinct banding patterns in the PCR products and lower correlations between signal intensities in each channel. As the continuous optimization of the PCR conditions made this method cumbersome, we did not pursue it further.

**Identification of transposon insertion sites using *in vitro* transcription on digested DNA**

To avoid the bias introduced by methods employing PCR, we directly used purified chromosomal fragments digested with TaqI (which on average cuts every 442 bp) as input for an *in vitro* transcription reaction using the T7 RNA polymerase promoter that we had introduced in both transposons (Fig. 1) (55,182). Comparison of duplicate reactions on a small set of 90 mutants with this method gave rise to specific and reproducible results (data not shown). Next, we tested the feasibility of this method by performing an artificial GAF experiment. To a small library of 90 mutants, 9 defined mutants (transposon insertion site determined by sequencing) were added (90 + 9 library). Subsequently, transposon insertion sites in both libraries were determined by microarray analysis as described above. As expected, the signal intensities of most genes were found to be similar in both channels (i.e., ratios between 0.5 and 2) indicating that mutants in these genes were present in both populations. Furthermore, the signal intensities of 13 genes in the 90 + 9 library were increased significantly in one channel (Fig. 3A), leading to ratios >9 (Table 2), indicating that transposon insertions in or near these genes were not present in the 90 mutant library. All 13 of these genes corresponded to ORFs present on the TaqI fragments of 8 defined mutants added to the library (Fig. 3A black squares, Table 2). The corresponding amplicons of the genes on the TaqI restriction fragment in one defined mutant (*SP2129* to *SP2131*) had high signal intensities in both channels (ratio of 1), indicating that a mutant with a transposon

### Table 2. ORFs present on the TaqI fragments on that contain the insertion sites of the 9 mutants added to the library and the corresponding normalized signals and ratios. ND, not detected signal below background filter, NA, not applicable. NP, ORF not represented by an amplicon on the microarray.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>ORFs present on TaqI fragment</th>
<th>Gene ID on array</th>
<th>Cy5 signal</th>
<th>Cy3 signal</th>
<th>Ratio 90 + 9 mutants/90 mutants</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td><em>SP0319</em></td>
<td><em>SP0319</em></td>
<td>15761</td>
<td>600</td>
<td>25.9</td>
</tr>
<tr>
<td></td>
<td><em>SP0320</em></td>
<td><em>SP0320</em></td>
<td>14689</td>
<td>354</td>
<td>41.5</td>
</tr>
<tr>
<td>A2</td>
<td><em>SP1262</em></td>
<td>NP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>SP0029</em></td>
<td><em>SP0029</em></td>
<td>6076</td>
<td>665</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td><em>SP0031</em></td>
<td>NP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>SP0032</em></td>
<td><em>SP0032</em></td>
<td>14318</td>
<td>1484</td>
<td>9.6</td>
</tr>
<tr>
<td>A4</td>
<td><em>SP0943</em></td>
<td><em>SP0943</em></td>
<td>19237</td>
<td>441</td>
<td>43.6</td>
</tr>
<tr>
<td>A6</td>
<td><em>SP2001</em></td>
<td><em>SP2001</em></td>
<td>6152</td>
<td>99</td>
<td>62.1</td>
</tr>
<tr>
<td></td>
<td><em>SP2002</em></td>
<td><em>SP2002</em></td>
<td>27091</td>
<td>692</td>
<td>39.1</td>
</tr>
<tr>
<td>A8</td>
<td><em>SP1533</em></td>
<td><em>SP1533</em></td>
<td>32682</td>
<td>1204</td>
<td>27.2</td>
</tr>
<tr>
<td>A10</td>
<td><em>SP2239</em></td>
<td><em>SP2239</em></td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td><em>SP2240</em></td>
<td><em>SP2240</em></td>
<td>21523</td>
<td>501</td>
<td>42.9</td>
</tr>
<tr>
<td>B7</td>
<td><em>SP2151</em></td>
<td><em>SP2151</em></td>
<td>3063</td>
<td>187</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td><em>SP2152</em></td>
<td><em>SP2152</em></td>
<td>42633</td>
<td>718</td>
<td>59.4</td>
</tr>
<tr>
<td></td>
<td><em>SP2153</em></td>
<td><em>SP2153</em></td>
<td>23473</td>
<td>454</td>
<td>51.7</td>
</tr>
<tr>
<td>B8</td>
<td><em>SP2129</em></td>
<td><em>SP2129</em></td>
<td>53211</td>
<td>48992</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td><em>SP2130</em></td>
<td><em>SP2130</em></td>
<td>2486</td>
<td>3021</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td><em>SP2131</em></td>
<td><em>SP2131</em></td>
<td>30685</td>
<td>25927</td>
<td>1.2</td>
</tr>
<tr>
<td>C2</td>
<td><em>SP1999</em></td>
<td><em>SP1999</em></td>
<td>26273</td>
<td>430</td>
<td>61.0</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td><em>SP1792</em></td>
<td>9058</td>
<td>2955</td>
<td>3.1</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td><em>SP2169</em></td>
<td>1715</td>
<td>577</td>
<td>3.0</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td><em>SP0001</em></td>
<td>1814</td>
<td>430</td>
<td>4.2</td>
</tr>
</tbody>
</table>
insertion in this location was already present in the 90 mutant library. Three additional genes were detected as added to the library (Fig. 3A, gray spots outside the lines that denote the 0.5 and 2.0 cut-off, Table 2). The detection of \textit{SP1792} is probably due to the fact that the amplicon has significant homology with amongst others, ORF \textit{SP1262} (data not shown) and thus the Cy5 signal is probably due to cross hybridization with the fragments generated by added mutant A2. \textit{SP0001} is adjacent to ORF \textit{SP2240} on the genome and is present on the TaqI fragment of added mutant A10, and the intermediate-level signals in the Cy5 channel could be due to the occurrence of some partially digested fragments. It is not clear why gene \textit{SP2169} is detected; however, the signals in both channels and the ratio are close to the arbitrarily chosen cutoff value of 2 and differ substantially from the signal intensities of the amplicons that represent the 9 added mutants.

Subsequently, this procedure was tested on one genome-wide \textit{marinerT7} library (>100,000 CFU) grown for approximately 20 generations in GM17. Three different restriction enzymes were used, TaqI, DdeI, and AluI, which have an increasing number of sites in the \textit{S. pneumoniae} genome; on average they cut every 442, 337 and 204 bp, respectively. The scatterplots derived from the analysis of each separate hybridization (Fig. 3B, C, D) clearly show a strong correlation between the signal intensities in each channel for each gene ($r^2 > 0.96$ in all cases) (the normalized signal data and cyberT output data for these arrays are available at http://molgen.biol.rug.nl/publication/GAF_data/). Restriction with DdeI resulted in the detection of 1301 genes and the numbers for the TaqI and AluI digests were comparable (1193 and 1265, respectively). Therefore, the three duplicate hybridizations were combined for further analysis. The library was generated in R6, which contains 2,116 ORFs, 1,995 of which are homologous to ORFS in strain TIGR4 and thus present on our array (133). As expected, all amplicons that represent ORFs present in strain TIGR4 but not R6 (317) (TIGR4 specific) had low signal intensities. These signals were used to manually generate a background filter that was applied to the analysis of the data from all three restrictions. Of the 1995 R6 ORFs on the array, 1322 (66\%) were detected in at least two restrictions and had a less than twofold ratio difference, demonstrating the reproducibility of this method. In principle, no mutants in genes that are essential for the growth of \textit{S. pneumoniae} can be present in the library; thus, amplicons representing these genes should not give signals above background (i.e., should be classified as nondetected). Comparison of the 673 nondetected R6 genes with the reported essential genes of \textit{S. pneumoniae} (5,54,177,184,226,298,318,352), showed that 80 had been designated as essential and 97 were located adjacent to putative essential genes (Supplementary Table S1, http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1828782&blobname=aem_73_5_1514_index.html). These analyses showed that this method was robust and reproducible for incorporation in GAF and that all three restriction enzymes can be used. The latter is important as not all restriction enzymes cut in all genes and the use of several restriction enzymes on the same DNA improves the chance of detecting all mutants that have disappeared from a population.

\textbf{Comparison of the pGhT7:ISS1 and marinerT7 transposon libraries}

Previous experiments were performed on \textit{marinerT7}-generated libraries. To test whether the method could also be applied to pGh9T7:ISS1 libraries, and to assess the level of saturation and randomness achieved by each transposon, we compared a large (>20,000 CFU) \textit{marinerT7} library with a pGh9T7:ISS1 one, each grown for approximately 20 generations in GM17. After isolation, the DNA was digested with DdeI and AluI and each detection reaction was performed in duplicate. Microarray data were analyzed as before using the background filter generated with the TIGR4-specific amplicons (the normalized signal data and cyberT output data for these arrays are available at
Development of Genomic Array Footprinting

http://molgen.biol.rug.nl/publication/GAF_data). This analysis resulted in the detection of 1,473 genes and the nondetected genes were similar to those not detected in the analysis of the three marinerT7 digests. The ratio of the majority of the genes was between 0.5 and 2, indicating that they are mutated by both transposons. However, 139 (9%) genes had a ratio lower than 0.5 or higher than 2, indicating that these genes are only, or preferentially, mutagenized by one of the two transposons. These “hotspots” are distributed throughout the genome (Table 3). There were 60 genes preferentially mutated by pGh9T7:ISS1, whereas 79 were preferentially mutated by marinerT7, indicating that it has a slightly higher saturation rate than pGh9T7:ISS1. Anchored PCR using primers located on some of the ORFs that had a ratio close to the cut-off point of 0.5 indicated that these genes were indeed preferentially hit by either transposon and that these ratios were not due to slight differences in, for instance, T7 RNA polymerase efficiency (data not shown).

Table 3. Summary of genes preferentially mutated by marinerT7 (ratio of <0.5, first two columns) or by pGh9T7:ISS1 (ratio of >2, last two columns), sorted by their location on the genome to show clustering. Gene ID refers to TIGR4 locus tags. Bayesian \( p \)-value of <0.0001 by CyberT, false discovery rate of <0.002 for all genes.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Ratio pGh9T7:ISS1/ marinerT7</th>
<th>Gene ID</th>
<th>Ratio pGh9T7:ISS1/ marinerT7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP0024-SP0027</td>
<td>0.40; 0.50; 0.26; 0.32</td>
<td>SP0054</td>
<td>2.45</td>
</tr>
<tr>
<td>SP0115</td>
<td>0.48</td>
<td>SP0069</td>
<td>2.37</td>
</tr>
<tr>
<td>SP0151</td>
<td>0.47</td>
<td>SP0071</td>
<td>2.52</td>
</tr>
<tr>
<td>SP0240</td>
<td>0.45</td>
<td>SP0097</td>
<td>2.15</td>
</tr>
<tr>
<td>SP0242</td>
<td>0.13</td>
<td>SP0105</td>
<td>2.03</td>
</tr>
<tr>
<td>SP0278</td>
<td>0.42</td>
<td>SP0112-SP0113</td>
<td>3.05; 3.52</td>
</tr>
<tr>
<td>SP0319</td>
<td>0.44</td>
<td>SP0116</td>
<td>2.62</td>
</tr>
<tr>
<td>SP0321</td>
<td>0.45</td>
<td>SP0123</td>
<td>2.15</td>
</tr>
<tr>
<td>SP0322</td>
<td>0.46</td>
<td>SP0340</td>
<td>2.03</td>
</tr>
<tr>
<td>SP0385</td>
<td>0.41</td>
<td>SP0412</td>
<td>2.06</td>
</tr>
<tr>
<td>SP0558</td>
<td>0.49</td>
<td>SP0439</td>
<td>2.41</td>
</tr>
<tr>
<td>SP0574</td>
<td>0.47</td>
<td>SP0530</td>
<td>2.19</td>
</tr>
<tr>
<td>SP0593</td>
<td>0.47</td>
<td>SP0585</td>
<td>2.50</td>
</tr>
<tr>
<td>SP0690-SP0693</td>
<td>0.34; 0.33; 0.33; 0.38</td>
<td>SP0609</td>
<td>2.43</td>
</tr>
<tr>
<td>SP0741-SP0743</td>
<td>0.13; 0.18; 0.35</td>
<td>SP0617-SP0618</td>
<td>2.65; 2.03</td>
</tr>
<tr>
<td>SP0846-SP0847</td>
<td>0.31; 0.29</td>
<td>SP0681</td>
<td>2.20</td>
</tr>
<tr>
<td>SP0943</td>
<td>0.28</td>
<td>SP0764-SP765</td>
<td>2.92; 2.11</td>
</tr>
<tr>
<td>SP0987-SP0988</td>
<td>0.41; 0.49</td>
<td>SP0800</td>
<td>2.50</td>
</tr>
<tr>
<td>SP1025</td>
<td>0.44</td>
<td>SP0881-SP0882</td>
<td>2.99; 2.61</td>
</tr>
<tr>
<td>SP1027</td>
<td>0.37</td>
<td>SP1018</td>
<td>2.08</td>
</tr>
<tr>
<td>SP1035</td>
<td>0.44</td>
<td>SP1104-SP1105</td>
<td>2.22; 2.04</td>
</tr>
<tr>
<td>SP1040-SP1043</td>
<td>0.35; 0.27; 0.47; 0.35</td>
<td>SP1126</td>
<td>2.16</td>
</tr>
<tr>
<td>SP1045</td>
<td>0.48</td>
<td>SP1149</td>
<td>2.23</td>
</tr>
<tr>
<td>SP1051</td>
<td>0.47</td>
<td>SP1376</td>
<td>2.26</td>
</tr>
<tr>
<td>SP1053</td>
<td>0.47</td>
<td>SP1505</td>
<td>2.83</td>
</tr>
<tr>
<td>SP1069</td>
<td>0.45</td>
<td>SP1578</td>
<td>2.04</td>
</tr>
<tr>
<td>SP1089</td>
<td>0.35</td>
<td>SP1586</td>
<td>2.16</td>
</tr>
<tr>
<td>SP1116</td>
<td>0.48</td>
<td>SP1609</td>
<td>2.70</td>
</tr>
<tr>
<td>SP1121</td>
<td>0.47</td>
<td>SP1627</td>
<td>2.07</td>
</tr>
<tr>
<td>SP1122</td>
<td>0.46</td>
<td>SP1659</td>
<td>3.00</td>
</tr>
<tr>
<td>SP1159-SP1164</td>
<td>0.47; 0.11; 0.38; 0.33; 0.34; 0.35</td>
<td>SP1691</td>
<td>2.11</td>
</tr>
<tr>
<td>SP1166</td>
<td>0.45</td>
<td>SP1743</td>
<td>2.13</td>
</tr>
<tr>
<td>SP1170-SP1171</td>
<td>0.49; 0.35</td>
<td>SP1790</td>
<td>2.56</td>
</tr>
<tr>
<td>SP1173</td>
<td>0.39</td>
<td>SP1795</td>
<td>4.36</td>
</tr>
<tr>
<td>SP1184</td>
<td>0.49</td>
<td>SP1852</td>
<td>2.74</td>
</tr>
<tr>
<td>SP1186-SP1187</td>
<td>0.49; 0.26</td>
<td>SP1981-SP1983</td>
<td>2.95; 5.43; 3.45</td>
</tr>
<tr>
<td>SP1190-SP1192</td>
<td>0.33; 0.40; 0.50</td>
<td>SP1996</td>
<td>2.04</td>
</tr>
</tbody>
</table>
Table 4. Summary of genes essential for surviving high Zn\(^{2+}\) concentrations identified with GAF. Genes were designated conditionally essential when the ratio GM17/GM17 + Zn\(^{2+}\) was <0.5, the Bayesian p-value <0.001 and the false discovery rate <0.05. Gene ID refers to TIGR4 locus tags, function is based on the TIGR annotation (317).

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Function</th>
<th>Ratio GM17 + Zn(^{2+})/GM17</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP1857</td>
<td>Cation efflux system protein CzcD</td>
<td>0.43</td>
</tr>
<tr>
<td>SP0856</td>
<td>Branched-chain amino acid aminotransferase IlvE</td>
<td>0.47</td>
</tr>
<tr>
<td>SP0857</td>
<td>ABC transporter substrate binding protein</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Figure 4. *czcD* is a conditionally essential gene when *S. pneumoniae* is grown under high Zn\(^{2+}\) conditions. (A) Growth of the wild-type (filled triangles) and a *czcD* deletion mutant (open triangles) in GM17. (B) Growth of the wild-type (filled triangles) and a *czcD* deletion mutant (open triangles) in GM17 to which 0.5 mM ZnSO\(_4\) was added. Results are representative of at least three replicate experiments; the error bars indicate the standard deviation. (C) Schematic representation of *in silico* digestion of the R6 chromosomal region containing *czcD* and its neighbouring genes with TaqI. Arrows indicate the location of genes, grey rectangles indicate the genomic regions that are present as amplicon on the microarray.
Identification of genes essential for survival of Zn$^{2+}$ stress

As all components for GAF were successfully developed, we tested its ability to identify conditionally essential genes from a library grown under in vitro stress conditions. Although Zn$^{2+}$ is an essential ion, it is toxic for bacteria in high concentrations. Bacteria often contain specific proteins for Zn$^{2+}$ secretion (229), but none have been identified in \textit{S. pneumoniae} so far. The concentration of Zn$^{2+}$ in serum is reported to be around 15.3 µM (1 mg L$^{-1}$) and in lung tissue 229.4 µM (15 µg/g wet tissue) (335), and during inflammation concentrations increase in blood and other body sites (216,321), indicating that \textit{S. pneumoniae} is likely to encounter Zn$^{2+}$ stress during infection. Therefore, we decided to screen for genes essential for the survival of Zn$^{2+}$ stress. In addition, a Zn$^{2+}$ exporter had already been identified by other means in our group (163). A large (>20,000 CFU) \textit{marinerT7} library in strain R6 was grown in quadruplicate in GM17 or GM17 containing 500 µM ZnSO$_4$, which is approximately half the maximal concentration that \textit{S. pneumoniae} can withstand (163), for approximately 10 generations. Subsequent detection of transposon insertion sites in both populations using a TaqI restriction showed that gene \textit{SP1857} and a cluster of two genes, \textit{SP0856-SP0857}, had a significantly lower ratio (Table 4), suggesting that these are essential genes for growth in high Zn$^{2+}$ concentrations. \textit{SP0856-SP0857} are annotated as \textit{ilvE} and as a oligopeptide-binding protein, respectively. \textit{SP1857}, annotated as \textit{czcD}, encodes a protein homologous to cation exporters (317) and was the exporter already identified as being involved in Zn$^{2+}$ resistance (163). A mutant generated in this gene was indeed unable to grow in GM17 containing 500 µM ZnSO$_4$ (Fig. 4A and B). Although the mutant was at least ten times more sensitive to ZnSO$_4$ than the parent strain, the ratio in the GAF experiment was 0.43. \textit{In silico} TaqI restriction analysis of the R6 nucleotide sequence of the homologue of SP1857 and the chromosomal region surrounding it, results in a DNA fragment that contains SP1857, its promoter and part of the adjacent gene, \textit{SP1858} (Fig 4C). As polar effects are unlikely to occur, mutants containing a transposon inserted before the \textit{czcD} ORF probably have no growth defect in GM17 plus ZnSO$_4$ and will not disappear from the population. The presence of such mutants will result in a DNA fragment that also hybridizes to the \textit{SP1857} amplicon, which possibly explains why the ratio was close to 0.5. However, this ratio was found to be significant by statistical analysis and the phenotype of the deletion mutant confirmed our GAF findings. Thus, we have successfully developed GAF as it correctly identifies conditionally essential genes in \textit{S. pneumoniae}.

Discussion

Here, we describe the development of GAF for \textit{S. pneumoniae}, a technique enabling genome-wide negative screens in a high-throughput fashion. Selection for genes unable to survive high Zn$^{2+}$ concentrations with GAF correctly identified \textit{czcD} as the main determinant for this resistance, which was also shown in a study exploring the response of \textit{S. pneumoniae} to Zn$^{2+}$ stress (163). One of the other genes identified in this screen, \textit{SP0857}, encodes a putative ABC transporter, suggesting that it could be involved in the secretion of Zn$^{2+}$. The function of \textit{ilvE} (\textit{SP0856}) in resistance to Zn$^{2+}$ stress is not immediately obvious; it could either be that mutation of this gene confers a growth defect under these particular conditions, or that there is a polar effect of the insertion on \textit{SP0857} or the other two downstream genes that encode membrane proteins. GAF has also been applied successfully to identify genes essential for natural competence (our unpublished results) and we are currently using it to identify genes essential in several different disease aspects.
GAF has several advantages: detection of transposon insertion sites in a library is performed in a genome-wide manner using microarrays, bypassing the need to generate mutants with unique tags and/or store all mutants separately; this allows for the use of large mutant libraries that have more than ten times coverage of the genome. It also eliminates the need to sequence each individual mutant to determine the transposon insertion site, which accelerates the whole screening procedure and makes it possible to go from recovering the mutants after selection to identifying all transposon insertion sites in about one week.

Recently, several methods have been developed to identify transposon insertion sites using microarrays (275, 279, 324). For all these approaches it is important to reduce the number of false positives, as the mutants are not stored separately and site-directed mutants have to be generated to confirm that the identified genes are indeed conditionally essential. Therefore, we have looked carefully into the specificity and reproducibility of these methods, an issue that is rarely addressed. We showed that methods that use a PCR step or a combination of PCR and adapters introduce a random selection of certain fragments in the procedure when used on genome-wide libraries, which leads to non-reproducible results. A similar effect was observed by Tong et al., and they showed that this is not dependent on the use of amplicon arrays (324). It seems more likely that this random selection of certain fragments is caused by the presence of more than 500 mutants at once, as sensitive and specific results with PCR-based methods can be obtained when small libraries (<500 CFU) or sub-pools of a large library are used (324, results not shown).

When the PCR step is omitted and restricted DNA is used directly as template for an in vitro transcription reaction, the detection of transposon insertion sites becomes reproducible and specific (Fig. 3). We used the signal present on TIGR4-specific amplicons for the generation of a background filter. Application of this filter resulted in the detection of 66% of the ORFS of strain R6 on our array, which is in agreement with the proposed S. pneumoniae core genome consisting of 73% of the TIGR4 ORFs (236). As expected, the set of nondetected genes contained many essential genes or genes adjacent to essential genes. The latter indicates that there are quite some polar effects in the studies investigating the essential genes of S. pneumoniae. Other nondetected genes are either hypothetical, or expected to be essential as they for instance encode parts of the ribosome (e.g. SP0213, SP0214, SP0232-234, Supplementary Table S1). A few are represented by short amplicons and it could be that insertions in this region, combined with the restriction enzymes used, do not generate fragments that hybridize to these amplicons (this could be the case for e.g. SP0540-548 which encode a putative bacteriocin operon and we do not expect these genes to be essential (Supplementary Table S1). We also showed that large genome-wide libraries (>20,000 CFU) generated with marinerT7 and pGh9T7:ISS1 have a similar degree of saturation but that each transposon is missing approximately 4% of the total number of genes. These “blind spots” are probably due to differences in the insertion site preferences of each transposon, although the genes missed by marinerT7 do contain its recognition site TA. This also shows that the duplication of pGh9T7:ISS1 in the genome does not interfere with the detection of the transposon insertion sites and that this system is a good alternative for the marinerT7 system for use in strains that do not have a high transformation efficiency.

In contrast to transcriptome analysis, the fold-change obtained by GAF analysis bears no direct relation to the functional impairment caused by mutation of a gene. For instance, the ratio of czcD in the Zn²⁺ stress GAF screen was 0.43, but a deletion mutant was at least ten times more sensitive to high Zn²⁺ concentrations than the wild-type. This apparent discrepancy has two probable causes; i) the ratio is dependent on the generated restriction fragments, which prevents control over the length of the probes generated for hybridization. It is conceivable that an essential gene is also represented on a fragment derived from a transposon insertion in an adjacent, non-essential gene, thereby masking the effect of the
disappearance of the mutant in the conditionally essential gene. This is probably the case for czcD. ii) Not all transposon insertions in a gene abolish gene function completely; however, these mutants will also generate a signal during the GAF procedure, again leading to a masking of the effect of the complete disruption of gene function. These effects are amplified when large, complex libraries are used, which explains why the ratios in the artificial 90 + 9 GAF experiment are much higher than those obtained when genome-wide libraries are used.

In conclusion, GAF is a versatile, high-throughput method for the identification of conditionally essential genes in S. pneumoniae. GAF should be easily adaptable to other microorganisms, as the only prerequisites for this technique are the availability of random (transposon) mutant libraries and DNA microarrays.

Acknowledgements

We thank Marc Prudhomme and Jean-Pierre Claverys (CNRS, Toulouse, France), and Andrew Camilli (Tufts University School of Medicine, Boston, USA) for kindly providing plasmid pR412 and pET29C9. We thank Don Morrison (University of Chicago) for his generous gift of competence-stimulating peptide. We thank Magda van der Kooi-Pol for skillful technical assistance with the Zn2+ growth experiments. This work was supported by IOP genomic grant IGE3002 of the Dutch Ministry of Economic Affairs.
Chapter 4