A Novel Gene Amplification Causes Up-regulation of the PatAB ABC Transporter and Fluoroquinolone Resistance in *Streptococcus pneumoniae*.

Alison J. Baylay¹, Alasdair Ivens² and Laura J. V. Piddock¹*

¹Antimicrobials Research Group, School of Immunity and Infection, Institute of Microbiology and Infection and College of Medical and Dental Sciences, University of Birmingham, United Kingdom, B15 2TT

²Centre for Immunity, Infection and Evolution, Ashworth Laboratories, King's Buildings, University of Edinburgh, Edinburgh, United Kingdom, EH9 3JT

Keywords: Multi-drug resistance, Efflux, gene regulation, genomics

Running Title: Gene amplification of *patAB*

* Corresponding author

Email: l.j.v.piddock@bham.ac.uk

Tel: 0121-414-6966

Fax: 0121-414-6819
Abstract

Over-expression of the ABC transporter genes *patA* and *patB* confers efflux-mediated fluoroquinolone resistance in *Streptococcus pneumoniae*, and is also linked to pneumococcal stress responses. Although up-regulation of *patAB* has been observed in many laboratory mutants and clinical isolates, regulatory mechanisms controlling expression of these genes are unknown. In this study, we aimed to identify the cause of high-level constitutive over-expression of *patAB* in M184, a multidrug resistant mutant of *S. pneumoniae*. Using a whole genome transformation and sequencing approach, we identified a novel duplication of a 9.2 kb region of the M184 genome, which included the *patAB* genes. This duplication did not affect growth and was semi-stable with a low segregation rate. Expression levels of *patAB* in M184 were much higher than could be fully explained by doubling of gene dosage alone, and inactivation of the first copy of *patA* had no effect on multidrug resistance. Using a GFP reporter system, increased *patAB* expression was ascribed to transcriptional read-through from a tRNA gene upstream of the second copy of *patAB*. This is the first report of a large genomic duplication causing antibiotic resistance in *S. pneumoniae*, and also of a genomic duplication causing antibiotic resistance by a promoter switching mechanism.
**Introduction**

Gene amplification, the transient generation of tandem repeats of large chromosomal regions, is an important source of variability in bacterial populations. Data from *Salmonella* suggest that, in non-selected cultures, transient duplications occur at frequencies of $10^{-4}$ to $10^{-2}$, meaning that 10%-30% cells in a non-selected culture have a duplication somewhere in their genome at any one time (1). Due to this high prevalence it is thought that gene amplification, rather than point mutations, provides a pool of preexisting variation allowing bacterial populations to adapt to various stresses, such as antibiotic resistance or growth on unusual carbon sources (2). However, gene amplification events are difficult to detect and study, as the tandem repeat units are identical in sequence, and the gene amplifications are often unstable and are lost by recombination between repeat units once the selective pressure has been removed.

*S. pneumoniae* is the main bacterial cause of community acquired pneumonia and represents a major disease burden worldwide (3). Despite the recent introduction of the heptavalent pneumococcal conjugate vaccine, antibiotic resistance is an increasing problem in this organism due to the spread of multidrug resistant clones and increases in antimicrobial resistance among non-vaccine serotypes (4).

The *patA* (spr1887) and *patB* (spr1885) genes, which encode ABC half-transporters, have been associated with intrinsic resistance of *S. pneumoniae* to some fluoroquinolone antibiotics. Each half-transporter consists of a nucleotide binding domain and a membrane spanning domain, and heterodimerisation of PatA and PatB is required to form a functional transporter (5). Constitutive over-expression of *patA* and *patB* has been observed in both laboratory mutants and clinical isolates and causes decreased susceptibility to ciprofloxacin and norfloxacin, as well as other agents such as dyes and the biocide cetrimide (6–9). In all cases, *patA* and *patB* are co-regulated, despite the presence of a degenerate transposase gene (spr1886) encoded in the opposite direction separating the two genes. Inactivation of *patA* or *patB* either genetically or using efflux inhibitors results in a hyper-antibiotic-
susceptible phenotype in laboratory mutants (7), and in clinical isolates increases ciprofloxacin susceptibility even if the isolate also contains topoisomerase mutations (9).

In addition to transporting fluoroquinolones, recent evidence suggests that PatAB plays a role in pneumococcal stress responses. Fluoroquinolone transport is unlikely to be the primary physiological role of PatAB as fluoroquinolones are synthetic antibiotics that have only been in use for 30 years. Expression of patA and patB can be induced by several fluoroquinolones, including those that are thought not to be substrates for transport, as well as the non-quinolone DNA damaging agent mitomycin C (8, 10). This indicates that PatAB is upregulated as part of a response to DNA damage. Additionally, over-expression of patA and patB has been shown to alleviate a fitness cost of linezolid resistance (11), and a role has been suggested for PatAB in response to pH stress (12).

Despite the potential role of PatA and PatB in pneumococcal stress responses, no regulatory pathways controlling expression of patA and patB have been identified. Mutations affecting a Rho-independent transcriptional terminator upstream of patA were shown to cause patAB over-expression in three previous studies (11, 13, 14). However, mutations affecting this terminator are not found in all patAB over-expressing isolates (7, 9), and in these cases the genetic changes leading to over-expression are unknown.

M184 is a mutant of the unencapsulated laboratory strain R6 (Genbank accession: AE007317.1) characterized previously by Garvey and Piddock (7). It was generated by exposure of R6 to the efflux inhibitor reserpine, followed by transformation of the reserpine resistant phenotype of one of the selected mutants back into R6. As well as having reduced susceptibility to reserpine, M184 was found to be multi-drug resistant including resistance to fluoroquinolone antibiotics such as ciprofloxacin and norfloxacin, the dyes ethidium bromide and acriflavine, and the biocide cetrimide. This multiple drug resistance phenotype was shown to be due to the increased expression of patA and patB. In this study, the whole genome sequences of M184 and three second-generation transformants were analyzed to identify mutations that might be responsible for patAB over-expression. This revealed a
novel gene amplification, which caused high levels of \textit{patAB} expression due to a change in genomic context of the promoter of the second copy of \textit{patA}. 
**Materials and Methods**

**Bacterial strains and growth**

All *S. pneumoniae* used in this study are listed in Table 2. All *S. pneumoniae* strains and isolates were grown statically at 37°C in an atmosphere of 5% CO2 in Brain Heart Infusion broth (BHI; Oxoid, Basingstoke, UK) or on Columbia agar (Oxoid, Basingstoke, UK) supplemented with 5% defibrinated horse blood (TCS Biosciences, Buckingham, UK). Growth of strains in broth was assessed by measurement of optical density at 660 nm (OD660) of liquid cultures and by viable cell counts.

*E. coli* strains used for plasmid construction were grown overnight on Luria-Bertani (LB) agar at 37°C or in LB broth at 37°C with shaking at 180 rpm.

**DNA extraction, whole genome sequencing and data analysis**

Genomic DNA was extracted from all strains using the Wizard Genomic DNA extraction kit (Promega) following the manufacturer’s instructions.

Genome sequencing was carried out by Genepool (Edinburgh, UK) on the Illumina HiSeq (Illumina, Saffron Walden, UK) according to standard protocols to give 100 bp paired-end reads. Reads were mapped against the published R6 genome sequence (Genbank accession: AE007317.1) using Bowtie2 (15) using the --very-sensitive-local setting, and a consensus pileup was generated using mpileup from Samtools version 0.1.18 (16) using the faidx-indexed R6 genome as a reference. Raw sets of putative SNPs and small indels were generated from the pileup using BCFtools version 0.1.17-dev and filtered to remove variants with a quality score of less than 50 and covering read depth of less than 50. Identified mutations were confirmed by PCR and Sanger sequencing. To generate local assemblies, reads meeting the desired criteria were extracted from the read alignment using Samtools and assembled with Velvet version 1.1.07 (17), using the VelvetOptimiser script by Simon.
Gladman to determine the optimal parameter settings. Read alignments were visualized and per gene RPKM values calculated using Artemis version 14.0.0 (18).

**PCR and DNA sequencing**

The primer sequences and PCR conditions used in this study are listed in Table 3. PCR reactions were carried out essentially as described previously (7), although Extensor PCR mastermix was used to amplify the 9.2 kb duplicated region from strain M184 and its transformants.

PCR amplimers were sequenced by the Functional Genomics Laboratory (School of Biosciences, University of Birmingham, Birmingham, UK) as described previously (9).

**Transformation**

Transformations were carried out as previously described (7). Briefly, mid-logarithmic phase cultures of the recipient strain were diluted 1:20 in competence medium (Todd-Hewitt broth [Oxoid, Basingstoke, UK] containing 1 mM calcium chloride [Sigma Aldritch Ltd, Dorset, UK], 0.2% bovine serum albumin (BSA; Sigma Aldritch Ltd, Dorset, UK) and 100 ng/mL competence stimulating peptide 1 (CSP1; Mimotopes, Clayton, Victoria, Australia). Donor DNA was added at varying concentrations to 500 aliquots of competent cell suspension. Transformation reactions were incubated for three hours at 37°C, then 20 µL and 200 µL volumes were spread onto selective agar plates. Viable counts were determined in parallel to allow estimation of the transformation frequency.

**Antibiotics and susceptibility determination**

The MICs of various agents for the strains used in this study were measured using the standardized agar doubling dilution method according to the British Society for Antimicrobial Chemotherapy (19).

**Measurement of intracellular accumulation of ethidium bromide**
Efflux activities of strains were measured by monitoring uptake of ethidium bromide, which fluoresces when intercalated into DNA. Logarithmic phase cultures were pelleted by centrifugation at 2200 x g and resuspended in PBS. Cell suspensions were adjusted to an OD660 of 0.1 with fresh PBS, and 100 µL cell suspension was added in triplicate to a black microtitre tray. Fluorescence was measured at excitation and emission wavelengths of 530 and 600 nm using a FLUOstar Optima plate reader (BMG Labtech, Aylesbury, UK) every two min for a total of 30 min. A final concentration of 100 µM ethidium bromide was added to each well by injection on the second cycle of measurement.

**Measurement of expression of patA and patB**

RNA was extracted from three biological replicates of each strain to be tested using a Promega SV Total RNA Isolation kit according to the manufacturers’ instructions (Promega, Southampton, UK). RNA and contaminating DNA concentrations were determined using a Qubit fluorimeter (Life Technologies, Paisley, UK). Residual DNA contamination was removed by treatment with Turbo DNase (Life Technologies), and cDNA was generated using Superscript III reverse transcriptase (Life Technologies) following the first-strand synthesis protocol supplied by the manufacturer. Expression of patA and patB was measured relative to expression of rpoB by quantitative real-time PCR by monitoring fluorescence of SYBR-Green dye. Reactions consisted of 12.5 µL IQ SYBR Green supermix (Bio-Rad, Hemel Hempstead, UK), 375 nM each of forward and reverse primers and 1 µL cDNA in a 25 reaction. Real-time PCR was carried out using a Bio-Rad CFX96 thermal cycler with the following protocol: 3 min at 95°C, followed by 40 cycles of 10 seconds at 95°C, 30 seconds at 54.5°C. Expression values were calculated from the fluorescence data using the Pfaffl method (20).

**Insertional inactivation of individual copies of patA in M184**

An internal region of patA containing the magellan2 minitransposon from strain M246 (R6 patA::magellan2) was amplified by PCR using primer pair 5 (Table 2). The PCR amplifier was diluted 1:10 six times, and 10 µL of each concentration was transformed into M184 as
described above. Transformants were selected on 100 mg/L spectinomycin to select for successful incorporation of the minitransposon which contains a spectinomycin resistance gene, \textit{aad9}. Primer pairs specific for the N-terminal half of each copy of \textit{patA} were used to determine which copy of \textit{patA} contained the insert.

\textit{Growth kinetics and determination of stability of gene amplification}

To measure growth kinetics, a 1\% inoculum of the strain of interest was added to 50 mL BHI broth. Growth was monitored at hourly intervals by measurement of OD$\text{660}$ for eight hours.

To measure the stability of the gene amplification, M184 cultures were grown overnight in triplicate on 8 mg/L ethidium bromide to select for cells containing the duplication. Cells were scraped directly off the plates, resuspended in BHI broth containing 10\% glycerol, and stored at -80\°C. These time zero cultures were then diluted 1:1000 in fresh BHI broth and grown exponentially without selection for 12 hours (corresponding to approximately 15 generations). Samples of culture from zero and twelve hours were serially diluted in PBS and grown on non-selective agar to obtain single colonies. For each biological replicate and time point, 20 colonies were picked at random and restreaked onto both non-selective agar and plates containing 8 mg/L norfloxacin. The proportion of colonies retaining norfloxacin resistance at zero and twelve hours was compared to calculate the probability of loss of the duplication using the following equation. Assuming that at each generation a proportion $p$ of the population will lose the duplication, the proportion of cells retaining the duplication after $n$ generations is $(1-p)^n$ (21). To confirm that loss of norfloxacin resistance correlated with loss of the duplication, DNA was extracted from a representative sample of candidate colonies and checked for the presence of the duplication junction by PCR.

\textit{Construction of pBAV1K2 reporter plasmid, cloning of patA promoter and tRNA gene, and reporter assay}

To generate plasmid pBAV1K2, plasmids pBAV1K-T5-\textit{gfp} (22) and pMW82 (23) were digested with EcoRI and PstI (Thermo Scientific) and digestion products separated on a 1\% agarose gel. The 2800 bp plasmid backbone of pBAVK1K-T5-\textit{gfp} and the 824 bp band
corresponding to the gfp gene of pMW82 were extracted from the gel using a Qiagen gel
extraction kit according to the manufacturer’s instructions. The DNA fragments were ligated
using QuickStick ligase (Bioline) and used to transform E. coli Top10 cells (Life Technologies).
Transformants were selected on 50 mg/L kanamycin. To remove DNA methylation for
subsequent digestions, extracted plasmid was transformed into E. coli JM110 chemically
competent cells (kindly provided by Yanina Sevastsyanovich) and re-extracted.

To clone the patA promoter into pBAV1K2, giving pBAV1K2p, a 144 bp DNA fragment was
synthesized by GeneArt (Life Technologies) that covered the region from 12 to 146 bp
upstream of the patA start codon and incorporated a 5’ EcoRI site and a 3’ XbaI site. The
promoter fragment was digested with EcoRI and XbaI, purified using a QIAquick PCR
purification kit (Qiagen) and combined in a 100:1 ratio with pBAV1K2 DNA linearized with
XbaI and EcoRI. The DNA fragments were ligated with QuickStick ligase and used to
transform E. coli Top10 cells. Transformants were selected on 50 mg/L kanamycin and
screened for successful incorporation of the insert by PCR with primer pair six. To clone the
tRNA-Glu gene upstream of the patA promoter, giving plasmid pBAV1K2tp, a naturally
occurring SpeI site 134 bp upstream of the patA start codon was exploited. A 681 bp region
spanning the duplication junction in M184 was amplified using primer pair seven, which
introduces a 5’ EcoRI site. This was digested with EcoRI and SpeI, and combined in a 100:1
molar ratio with EcoRl- and SpeI-linearized pBAV1K2p. Ligation, transformation,
transformant selection and PCR screening was performed as before.

Plasmid DNA was extracted from E. coli cells using a QIAquick miniprep kit according to the
manufacturer’s instructions, and 20 µL of each plasmid preparation was used to transform
S. pneumoniae R6 as described above. Successful transformants were selected on plates
containing 100 mg/L kanamycin.

R6 cells containing pBAV1K2, pBAV1K2p, or pBAV1K2tp were grown in triplicate in BHI broth
supplemented with 100 mg/L kanamycin to an OD660 of 0.5. One mL of each culture was
transferred to a microcentrifuge tube and cells harvested by centrifugation (14,000 x g, 2
Cells were resuspended in 200 µL sterile PBS, and transferred to a black microtitre tray. Fluorescence was measured using a Fluostar Optima (BMG LabTech) with excitation and emission wavelengths of 492 and 520 nm, respectively.
Results

Over-expression of patAB can be transferred from M184 to R6 in a single transformation step

To identify the mutation responsible for increased expression of patA and patB in M184, the genome sequences of M184 and our laboratory stock of R6 were determined by Illumina sequencing. Reads were aligned against the published R6 genome sequence to identify single nucleotide polymorphisms (SNPs) and small insertion or deletion events (indels). This analysis identified 27 point mutations and small indels present in M184 that were not found in R6 (see the supplemental material at http://epapers.bham.ac.uk/1958/). Excluding reads that mapped to more than one position in the genome did not change the set of variants identified (data not shown).

To separate mutations causing patAB overexpression from compensatory mutations and bystander mutations, M184 DNA was transformed into R6 and transformants selected on 8 mg/L ethidium bromide. Reserpine was not used to select transformants as the mechanism of resistance to this agent appears to be more complex; inactivation of patA increased susceptibility to reserpine, but not inactivation of patB (7). To avoid selecting fluoroquinolone resistant topoisomerase mutants, ethidium bromide, another substrate of the PatAB transporter, was used to select transformants that had incorporated mutations causing patAB over-expression. Ethidium bromide-resistant transformants were selected in a single step, but at a transformation frequency of $10^{-6}$-$10^{-7}$, 10-20-fold lower than a spectinomycin resistance cassette transformed under the same conditions. No spontaneous ethidium bromide-resistant mutants were selected on DNA-free control plates.

Three transformants were chosen at random for further study, named R6M184T1, T2 and T3. Expression of patA and patB in all three transformants was measured by quantitative real-time PCR to confirm successful transfer of the patAB over-expression phenotype. In
M184 and all three transformants both genes were expressed at a level 100 - 1000 times higher than in R6 (Figure 1A).

The phenotypes of M184 and the three transformants were compared to confirm that the observed upregulation of patAB in the transformants conferred a phenotype similar to that of M184. Firstly, MICs of ciprofloxacin, norfloxacin and ethidium bromide for each of the strains were determined in the presence or absence of the ABC efflux pump inhibitor sodium orthovanadate for R6, M184, the three transformants and M240, an R6 derivative where patB has been inactivated by insertion of a magellan2 minitransposon (7). As expected, M184 and the three transformants showed decreased susceptibility to all three agents compared to R6 and M240, and this was reversible by addition of sodium orthovanadate (Table 1). However, the transformants differed slightly from M184 in that addition of sodium orthovanadate appeared to increase susceptibility to norfloxacin to a greater degree (at least two dilutions) in the transformants than in M184.

Accumulation of ethidium bromide by R6, R6 patB::magellan2, M184 and the three transformants was also measured. After ten minutes all three transformants accumulated significantly lower levels of ethidium bromide than R6 (1.4 to 1.6-fold reduction compared to R6, p<0.05, Figure 1B). However, the decrease in ethidium bromide accumulation in the transformants was less pronounced than in M184, which accumulated 2.6-fold lower levels of ethidium bromide than R6.

**Comparison of R6, M184 and transformants by whole genome resequencing reveals a novel gene amplification**

To identify which mutation or combination of mutations was transferred from M184 to the three transformants to give patAB over-expression, the whole genomes of the transformants were sequenced by Illumina sequencing. Reads were aligned against the published R6 genome and SNPs and small indels identified as before. Six mutations were found in transformant 1, three in transformant 2 and nine in transformant 3 (http://epapers.bham.ac.uk/1958/). Unexpectedly however, there were no mutations
shared between M184 and the three transformants, indicating that none of the 27
mutations found in M184 were the cause of patAB over-expression. This was confirmed by
PCR and Sanger sequencing of all candidate mutations. It was then hypothesized that patAB
over-expression was caused by a larger genomic rearrangement that is undetectable using
the SNP calling approach. This was investigated in two ways.

Firstly, the alignment of reads against the R6 genome was examined for evidence of
genomic rearrangements by identifying clusters of reads that did not align as “proper” pairs.
Proper pairs are defined as both reads aligning within 250 bp of each other and in the
correct relative orientation. Illumina sequence reads aligned against the published R6
genome were filtered to exclude all properly paired reads. Visual inspection of the filtered
alignment identified a region with a high density of “improperly paired” reads upstream of
patA in M184 and the transformants that was not present in R6 (Figure 2A). The pairs of
these reads were extracted from the raw data files by matching read names, and realigned
to the published R6 genome. They aligned to four regions of the genome, corresponding to
an area just upstream of each of the four copies of the rRNA genes. A local assembly
generated from these read pairs resulted in a single contig of 608 bp. This was used as a
query sequence to search the published R6 genome sequence. Hits were found between the
3’ 326 bp of the contig and regions upstream of the rRNA loci, with the closest matches
being with copies two and four (Figure 2B). The remaining 282 bp matched to a region
encompassing a region 183 to 464 bp upstream of patA. This suggested that a genomic
rearrangement had occurred involving the patA region.

Secondly, to look for evidence of changes in copy number of parts of the genome,
normalized read depths, expressed as reads per kilobase per million reads mapped (RPKM),
were calculated for each annotated CDS of more than 200 bp from the read alignments
from each strain, and the profiles from M184 and the three transformants were compared
against R6 (Figure 3). This analysis identified six genes that had significantly higher read
depths in M184 and the three transformants than in R6 (p < 0.05, one-tailed Student’s t-
test). These genes were contiguous and corresponded to spr1880 to spr1887 (spr1883 was
not included in the analysis as it is less than 200 bp in length). The average RPKM ratio for these genes relative to R6 was 1.2 (standard deviation = 0.4) across all four patAB-over-expressing strains, compared to 0.6 (standard deviation = 0.1) for the rest of the tested CDSs. This region encompassed patA (spr1887) and patB (spr1885), and the final gene in the region, spr1880, is located directly upstream of the fourth rRNA locus. Taken together, these results suggest that a 9.2 kb region of the genome, containing patA and patB has been tandemly duplicated in M184 and the three transformants.

To confirm the duplication suggested by the genome sequence analysis, PCR primers were designed to amplify the predicted novel junction between the two duplicated copies. PCR amplimers were obtained when DNA from M184 or the three transformants was used as a template, but not when R6 DNA was used (http://epapers.bham.ac.uk/1958/). The sequences of the resulting PCR amplimers were determined and confirmed to be the same as the contig assembled from the “improperly paired” reads.

Secondly, to confirm that the two copies of the locus were arranged in tandem, a second PCR reaction was carried out using primers specific for the 5’ end of patA and the 3’ end of the upstream gene hexA, which is not part of the duplication. In R6, these primers amplify the expected 357 bp region between patA and hexA. However, when DNA from M184 or the three transformants was used as a template and the PCR extension time was increased to ten minutes, a second PCR amplimer was observed of approximately 9 kb (http://epapers.bham.ac.uk/1958/). This corresponds to amplification from the patA primer binding to the second copy of patA; it was absent when R6 DNA was used as a template.

The gene amplification does not affect growth and has a low segregation rate

Generation times of R6, M184 and the three transformants were measured during growth in liquid medium. M184 had a significantly longer generation time than R6 (46±4 mins compared to 34±2 mins, p < 0.05), but no significant growth defect was observed in the three transformants. This implies that carriage of the duplication itself does not affect
growth of R6 in liquid medium. We hypothesize that growth of M184 is impaired because of one or more of the other mutations it carries. For example, mutations were found in genes encoding proteins involved in key cellular processes, such as the translation factor Ef-Tu (http://epapers.bham.ac.uk/1958/).

Large genomic duplications are often unstable as recombination between the two copies of the duplicated locus can cause loss of the intervening DNA (2). The rate of loss of the duplication from M184 and the transformants was measured during growth in liquid medium in the absence of antibiotic. Following twelve hours of exponential phase growth in BHI broth without antibiotic 72%±11% of colonies tested had lost resistance to norfloxacin, suggesting a probability of loss per cell per generation of 0.018±0.008 (n=8). To confirm that loss of norfloxacin resistance correlated with the loss of the duplication, seven norfloxacin sensitive and seven norfloxacin resistant colonies were tested by PCR for the presence of the duplication junction. PCR amplimers corresponding to the duplication junction were obtained from all seven of the resistant colonies but none of the sensitive colonies.

Based on these observations, and assuming continuous exponential growth at a rate equivalent to R6 (generation time of 34 minutes), the proportion of cells carrying the duplication would drop below the detection limit of this experiment (5%) after 2.5 to 7 days.

The duplication mechanism is unknown

To determine the mechanism of gene amplification in M184, 50 base pairs either side of the start and end point of the duplication and the duplication junction were aligned and examined for the presence of repeat sequences. When comparing the start and end points of the duplication no obvious repeat sequences were found. Only two identical bases (TA) were precisely conserved between the start and end of the duplication, and there were only seven identical bases in total in the 20 bp region surrounding the join point (http://epapers.bham.ac.uk/1958/). Duplication of the region results in duplication of the conserved TA motif at the junction point, but introducing gaps into the alignment to take
this into account still did not reveal any larger conserved repeat sequences (data not shown). Interestingly, the nine bp of sequence either side of the end point of the duplication formed a near-perfect inverted repeat with only one mismatch (CTACAACAT-AAGTTGTAG). However, this sequence was not present in the sequence around the start of the duplication, or at the duplication junction. The RegTransBase database of prokaryotic cis-regulatory elements (24) was searched for motifs similar to this inverted repeat using the STAMP motif alignment tool (25), but no matches were found. Illegitimate end-joining by DNA gyrase has been implicated in the formation of some gene amplifications, so the duplication start and end points were examined for similarity to known pneumococcal DNA gyrase and topoisomerase IV cleavage site consensus sequences previously reported (26). No convincing matches were found.

**M184 also over-expresses guaA but not the other duplicated genes**

As well as patA and patB, the 9 kb duplicated region contains six other protein coding genes and a tRNA gene, represented in Figure 4A. Expression levels of five of these, spr1886, guaA (spr1884), gpi (spr1882), gltX (spr1881) and spr1880, were measured by qRT-PCR from R6, M184 and the three transformants. In M184 and the three transformants, guaA was expressed at a level 20-50 fold higher than that of R6. This gene encodes a type 1 glutamine amidotransferase enzyme (NCBI conserved domain: cd01745). Approximately seven-fold higher levels of transcript were observed in M184 for the degenerate transposase gene spr1886, which is encoded between the patA and patB genes in the opposite orientation. There was no difference in expression for any of the other tested genes (Figure 4B). Expression levels of the tRNA gene could not be measured as five copies of this gene exist in the R6 genome, meaning that unique primers could not be designed.

**Over-expression of patA and patB in M184 comes from the second copy of the locus.**
Duplication of \textit{patA} and \textit{patB} did not fully explain the very high levels of \textit{patAB} over-expression measured by qRT-PCR in M184 and the three transformants. Wild-type levels of expression from two identical gene copies could theoretically result in a two-fold increase in expression, but instead large (100- to 1000-fold) increases in expression were observed.

There are two possible explanations for this observation. Firstly, the duplication changes the genetic context of the second copy of \textit{patA} and \textit{patB}, which may result in increased expression of these genes due to loss of regulatory control. Alternatively, both copies of \textit{patA} and \textit{patB} could be over-expressed due to a positive feedback loop if the initial increase of expression due to the duplication itself is sufficient to increase \textit{PatAB} levels above a threshold value. If the first hypothesis is correct, inactivation of the first copy of \textit{patA} should have no effect on the phenotype of M184, while inactivation of the second copy should reduce efflux (and increase accumulation) to R6 levels. On the other hand, if the second hypothesis is correct, inactivation of either copy of \textit{patA} would reverse antibiotic resistance in M184.

To test this, M184 cells were transformed with donor DNA amplified from M246, a strain of R6 where \textit{patA} is inactivated by insertion of the \textit{magellan2} minitransposon (7). The donor PCR amplimer contained the Mariner minitransposon flanked by internal \textit{patA} sequence, meaning that recombination with either copy of \textit{patA} in M184 should be equally likely. It is known that multiple recombination events can occur simultaneously in one pneumococcal cell (27) so, to select transformants where only one copy of \textit{patA} was inactivated, cells were transformed with a series of dilutions of the PCR amplimer and candidates selected from the lowest donor DNA concentration that produced viable colonies. It was expected that roughly equal numbers of transformants with the transposon inserted in each copy of \textit{patA} would be obtained. However, a total of 52 spectinomycin resistant transformants were obtained from three separate experiments and all contained the \textit{magellan2} transposon inserted in the first copy of \textit{patA}. Three transformants containing the insertion in copy one of \textit{patA} were selected at random for further testing. The MICs of ciprofloxacin, norfloxacin and ethidium bromide were measured for these strains in the presence and absence of
sodium orthovanadate. The MICs obtained for all three strains were identical to those for M184, suggesting that inactivation of the first copy of patA in M184 does not abolish PatAB over-expression.

**Over-expression from copy two of the patAB locus could be caused by read-through from a tRNA gene**

To determine why the second copy of patA appeared to contribute more to the M184 phenotype than the first copy, the genomic context of the second copy of patA was examined. The duplication results in the start codon of the second copy of patA being located 365 bp downstream of one of the five copies of a glutamate-specific tRNA gene found in the R6 genome. A list of *S. pneumoniae* transcriptional terminator predictions made by TransTermHP (28) was examined, and no Rho-independent terminator predictions were found in the downstream region of the tRNA gene. This suggested that increased expression of the second copy of the patAB locus could be conferred by transcriptional read-through from the tRNA gene. To test this prediction, a promoter probe plasmid was constructed by replacing a T5 promoter-gfp cassette from the broad host range plasmid pBAV1K-T5-gfp (22) with a promoterless gene coding for an unstable GFP variant from plasmid pMW82 (23), generating vector pBAV1K2. Unstable GFP was used to prevent changes in expression of GFP from the tested promoter from being obscured by fluorescence from accumulated stable GFP. Two different DNA fragments were cloned upstream of the promoterless gfp (Figure 5A). The first contained the predicted patA promoter alone (pBAV1K2p), while the second, larger region spanned the M184 duplication junction from the end of copy 1 of spr1880 to the start of copy 2 of patA, and contained the tRNA gene upstream of the patA promoter (pBAV1K2tp).

The promoter constructs were transformed into R6 and GFP fluorescence was measured during mid-logarithmic phase growth (Figure 5B). Fluorescence from pBAV1K2p was 2.4-fold higher than R6 harboring empty pBAV1K2 vector (574±132 fluorescence units compared to 236±104 units). Fluorescence levels observed for pBAV1K2tp were 6.3-fold higher than for...
the patA promoter alone (3682±327 fluorescence units). However, this may be an underestimate of true promoter activity as this strain had a severe growth defect compared to R6 containing pBAV1K2 or pBAV1K2p (generation time of 56 mins compared to 35 mins, respectively). Although not confirmed, this may be a side-effect of GFP toxicity.
Discussion

Over-expression of *patA* and *patB* confers clinically relevant resistance to fluoroquinolone antibiotics (9). In M184, over-expression of these genes was conferred by a duplication of a 9.2 kb section of the genome, not a point mutation. To our knowledge, this is the first example of a large (>1kb) gene amplification in *S. pneumoniae*. The only previously reported instance was an 18bp duplication of a ribosome binding site in the *ermA* gene which caused macrolide resistance (29). Gene amplifications resulting in antibiotic resistance have been observed in several bacterial species, and have been reviewed elsewhere (30).

Gene duplications can form by RecA-dependent and independent mechanisms. RecA-dependent duplications are formed by homologous recombination between long repeats, where 25-40 nt of sequence identity is required (31, 32). Common manifestations of this are recombinations between rRNA operons or IS elements (1, 33). Two mechanisms have been proposed to explain RecA-independent duplications. The first is strand slippage during rolling circle DNA replication, which requires short areas of sequence identity (34). The second is illegitimate end joining during repair of double stranded DNA breaks catalyzed by DNA gyrase (35). This does not require sequence homology at the duplication ends (36). No direct repeats were found at the ends of the duplicated region in M184, suggesting that the formation of this duplication is likely to be due to illegitimate end joining. Expression of *patAB* has been previously shown to be induced by fluoroquinolones and mitomycin C, which are DNA damaging agents (8, 10). It is tempting to speculate that this could have been due to transient duplication promoted by DNA gyrase inhibition.

Similar homology-independent duplication junctions were observed previously in a study of 104 amplification mutants of *Acinetobacter* with increased ability to grow on benzoate, where 36 of 104 distinct duplication junctions identified showed no sequence identity at the join point (37). Several of these homology-independent duplications were identified in multiple independent mutants, suggesting that homology-independent duplications can be formed in a site-specific manner. This could be explained by site-specificity in DNA gyrase.
cleavage sites, as it has been previously shown that DNA gyrase cleaves preferentially at
certain sites (38). In M184 the sites upstream and downstream of the duplication did not
match the reported S. pneumoniae DNA gyrase consensus cleavage site (26). However,
gyrase cleavage sites are highly degenerate so site-specific gyrase cleavage cannot be
completely ruled out. It has also been suggested that site-specificity in formation and
deletion of tandem repeats could be mediated by DNA secondary or tertiary structure
promoting recombination between particular sites through protein-protein or protein-DNA
interactions (37, 39, 40). It is interesting that there is an inverted repeat motif at the 3’ end
of the duplicated region that could be a binding site for an unknown protein factor.

Following an initial tandem duplication event, further gene amplification can occur by RecA-
dependent homologous recombination between copies of the duplicated locus, or by
rolling-circle replication mechanisms (30). In several previous studies of adaptive gene
amplification observed copy numbers were greater than two, indicating that further
amplification has taken place. In Salmonella enterica five- to 40-fold amplification of genes
encoding initiator tRNAs was shown to alleviate a fitness cost of resistance to peptide
deformylase inhibitors (41). A five-fold amplification of the blaA gene caused ampicillin
resistance in Yersinia enterolitica (42). In the A. baumannii study described above, mutants
with increased ability to grow on benzoate were found to possess between three and 45
copies of the cat genes (37). Additionally, in a recent study, 10- to 110-fold amplifications of
a Tn6020 transposon containing an aminoglycoside resistance gene was shown to be the
cause of tobramycin treatment failures in A. baumannii clinical isolates (43). A four-fold
amplification of the folCEPBK genes was found in Streptococcus agalactiae, which resulted
in decreased susceptibility to sulfonamides and trimethoprim (21). In these cases the
adaptive benefit of the amplification was either shown or assumed to be linked to the copy
number of the amplified genes.

Further amplification of the duplicated region has not occurred in M184, and the copy
number is therefore only two. However, this duplication results in a much higher expression
of patA and patB than can be explained by increased gene dosage alone. Insertional
inactivation of the first copy of \textit{patA} had no effect on the antibiotic resistance phenotype, indicating that the increased production of PatAB originates from the second copy of the locus. The duplication event alters the genomic context of the second copy of \textit{patAB}, and brings it under the control of the promoter of a tRNA gene.

Therefore, in contrast to previous studies, although the adaptive mutation in this strain is a gene amplification, the primary cause of the increased antibiotic resistance is altered expression of the \textit{patAB} locus, rather than increased gene dosage. Antibiotic resistance mediated by promoter switching caused by a gene deletion event has been previously identified (44). However, to our knowledge, this is the first observation of a transient gene duplication causing antibiotic resistance due to placing resistance genes under the control of a highly active promoter. In the sulfonamide and trimethoprim resistant \textit{S. agalactiae} isolate (21), expression of the second, third and fourth copies of the \textit{fol} genes was ascribed to an alternative promoter, but this only led to a five-fold increase in gene expression, compared to the four-fold increase that would be expected from gene dosage alone.

Most gene duplication events are unstable and are lost in the absence of selection due to recombination between identical copies of the locus (2). Typical segregation rates range between 0.01 and 0.15 loss events per cell per generation (2), although the 13.5kb quadruplication event causing sulfonamide and trimethoprim resistance in \textit{S. agalactiae} was highly stable and only lost at a rate of 0.003 per cell per generation (21). The duplication observed in M184 was unstable as expected, and lost at an estimated rate of between 0.01 and 0.03 per cell per generation. Although not approaching the level of stability observed in the \textit{S. agalactiae} strain, this is still toward the more stable end of the range of previously observed segregation rates. The observed rate of duplication loss suggests that the duplication would be undetectable by the method used here within 2.5 to seven days; however, this is assuming continuous exponential growth under laboratory conditions, whereas growth in a physiological situation may be limited by a number of factors. Therefore, it is possible that the stability of this duplication, and therefore increased expression of \textit{patAB}, could be higher in an environment in which antibiotics are used.
Over-expression of guaA, which is adjacent to patB in the S. pneumoniae genome, was also observed in M184 and the three transformants. This gene encodes a type 1 glutamine amidotransferase enzyme, which is a diverse family of enzymes catalyzing transfer of ammonia from glutamine to a variety of substrates (45). However, over-expression of guaA was not observed in previous studies where expression of this gene was measured, either as part of a global transcriptome analysis (10) or directly by qRT-PCR (46). The gene was over-expressed to a lesser extent than patA and patB (20-50 fold compared to 100-1000 fold) so it is unclear whether this shows that guaA is co-expressed with patAB on a single transcript, or whether the observed over-expression is due to leaky termination of transcription from patB. However, over-expression of patAB has been linked to DNA damage (8), and glutamine amidotransferases are known to be involved in nucleotide biosynthesis pathways (47); a functional link is therefore possible. The gpi, gltX and spr1880 genes downstream of guaA were not upregulated in this study. Promoter prediction software (48) suggests the presence of a promoter sequence upstream of these genes (data not shown) so it is likely that they are expressed on at least one separate transcript.

Data presented here represent the first report of a large gene amplification in S. pneumoniae causing antibiotic resistance. This is also the first time that capture of a resistance gene by a strong promoter has been observed due to transient gene amplification. Taken together, these results add to the growing body of evidence suggesting that gene amplification is an important adaptive mechanism in bacteria allowing them to survive antibiotic exposure.
Acknowledgments

This work was supported by a Medical Research Council Doctoral Training Grant (grant number DKAA GAS0025) to LJVP.
References

1. Anderson P, Roth J. 1981. Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons. Proc. Natl. Acad. Sci. U. S. A. 78:3113–7.

2. Andersson DI, Hughes D. 2009. Gene amplification and adaptive evolution in bacteria. Annu. Rev. Genet. 43:167–95.

3. Isaacman DJ, McIntosh ED, Reiners RR. 2010. Burden of invasive pneumococcal disease and serotype distribution among Streptococcus pneumoniae isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. Int. J. Infect. Dis. 14:e197–209.

4. Cornick JE, Bentley SD. 2012. Streptococcus pneumoniae: the evolution of antimicrobial resistance to beta-lactams, fluoroquinolones and macrolides. Microbes Infect. 14:573–83.

5. Boncoeur E, Durmort C, Bernay B, Ebel C, Di Guilmi AM, Croizé J, Vernet T, Jault J-M. 2012. PatA and PatB form a functional heterodimeric ABC multidrug efflux transporter responsible for the resistance of Streptococcus pneumoniae to fluoroquinolones. Biochemistry 51:7755–65.

6. Marrer E, Schad K, Satoh AT, Page MGP, Johnson MM, Piddock LJV. 2006. Involvement of the putative ATP-dependent efflux proteins PatA and PatB in fluoroquinolone resistance of a multidrug-resistant mutant of Streptococcus pneumoniae. Antimicrob. Agents Chemother. 50:685.

7. Garvey MI, Piddock LJ V. 2008. The efflux pump inhibitor reserpine selects multidrug-resistant Streptococcus pneumoniae strains that overexpress the ABC transporters PatA and PatB. Antimicrob. Agents Chemother. 52:1677–85.

8. El Garch F, Lismond A, Piddock LJ V, Courvalin P, Tulkens PM, Van Bambeke F. 2010. Fluoroquinolones induce the expression of patA and patB, which encode ABC efflux pumps in Streptococcus pneumoniae. J. Antimicrob. Chemother. 65:2076–82.

9. Garvey MI, Baylay AJ, Wong RL, Piddock LJ V. 2011. Overexpression of patA and patB, which encode ABC transporters, is associated with fluoroquinolone resistance in clinical isolates of Streptococcus pneumoniae. Antimicrob. Agents Chemother. 55:190–6.

10. Marrer E, Satoh AT, Johnson MM, Piddock LJV, Page MGP. 2006. Global transcriptome analysis of the responses of a fluoroquinolone-resistant Streptococcus pneumoniae mutant and its parent to ciprofloxacin. Antimicrob. Agents Chemother. 50:269.

11. Billal DS, Feng J, Leprohon P, Légaré D, Ouellette M. 2011. Whole genome analysis of linezolid resistance in Streptococcus pneumoniae reveals resistance and compensatory mutations. BMC Genomics 12:512.

12. Van Opinjten T, Camilli A. 2012. A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res. 22:2541–51.
13. Croucher NJ, Mitchell AM, Gould K a, Inverarity D, Barquist L, Feltwell T, Fookes MC, Harris SR, Dordel J, Salter SJ, Browall S, Zemlickova H, Parkhill J, Normark S, Henriques-Normark B, Hinds J, Mitchell TJ, Bentley SD. 2013. Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. PLoS Genet. 9:e1003868.

14. Baylay AJ, Piddock LJ V. Clinically Relevant Fluoroquinolone Resistance due to Constitutive over-expression of the PatAB ABC transporter in Streptococcus pneumoniae is conferred by disruption of a transcriptional attenuator. J. Antimicrob. Chemother.

15. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–9.

16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–9.

17. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–9.

18. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28:464–9.

19. Andrews JM. 2001. The development of the BSAC standardized method of disc diffusion testing. J. Antimicrob. Chemother. 48 Suppl 1:29–42.

20. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29:e45.

21. Brochet M, Couvé E, Zouine M, Poyart C, Glaser P. 2008. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in Streptococcus agalactiae. J. Bacteriol. 190:672–80.

22. Bryksin A V, Matsumura I. 2010. Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. PLoS One 5:e13244.

23. Bumann D, Valdivia RH. 2007. Identification of host-induced pathogen genes by differential fluorescence induction reporter systems. Nat. Protoc. 2:770–7.

24. Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov D V, Arkin A, Mironov AA, Gelfand MS, Dubchak I. 2007. RegTransBase—a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. Nucleic Acids Res. 35:D407–12.

25. Mahony S, Benos P V. 2007. STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res. 35:W253–8.

26. Leo E, Gould K a, Pan X-S, Capranico G, Sanderson MR, Palumbo M, Fisher LM. 2005. Novel symmetric and asymmetric DNA scission determinants for Streptococcus pneumoniae topoisomerase IV and gyrase are clustered at the DNA breakage site. J. Biol. Chem. 280:14252–63.

27. Croucher NJ, Harris SR, Barquist L, Parkhill J, Bentley SD. 2012. A high-resolution view of genome-wide pneumococcal transformation. PLoS Pathog. 8:e1002745.
28. Kingsford CL, Ayanbule K, Salzberg SL. 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol. 8:R22.

29. Musher DM, Dowell ME, Shortridge VD, Flamm RK, Jorgensen JH, Le Magueres P, Krause KL. 2002. Emergence of macrolide resistance during treatment of pneumococcal pneumonia. N. Engl. J. Med. 346:630–1.

30. Sandegren L, Andersson DI. 2009. Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat. Rev. Microbiol. 7:578–88.

31. Shen P, Huang H V. 1986. Homologous recombination in Escherichia coli: dependence on substrate length and homology. Genetics 112:441–57.

32. Watt VM, Ingles CJ, Urdea MS, Rutter WJ. 1985. Homology requirements for recombination in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 82:4768–72.

33. Haack KR, Roth JR. 1995. Recombination between chromosomal IS200 elements supports frequent duplication formation in Salmonella typhimurium. Genetics 141:1245–52.

34. Trinh TQ, Sinden RR. 1993. The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in Escherichia coli. Genetics 134:409–22.

35. Ikeda H, Shiraishi K, Ogata Y. 2004. Illegitimate recombination mediated by double-strand break and end-joining in Escherichia coli. Adv. Biophys. 38:3–20.

36. Shimizu H, Yamaguchi H, Ashizawa Y, Kohno Y, Asami M, Kato J, Ikeda H. 1997. Short-homology-independent illegitimate recombination in Escherichia coli: distinct mechanism from short-homology-dependent illegitimate recombination. J. Mol. Biol. 266:297–305.

37. Reams AB, Neidle EL. 2004. Gene amplification involves site-specific short homology-independent illegitimate recombination in Acinetobacter sp. strain ADP1. J. Mol. Biol. 338:643–56.

38. Ikeda H, Aoki K, Naito A. 1982. Illegitimate recombination mediated in vitro by DNA gyrase of Escherichia coli: structure of recombinant DNA molecules. Proc. Natl. Acad. Sci. U. S. A. 79:3724–8.

39. De Vries J, Wackernagel W. 2002. Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. Proc. Natl. Acad. Sci. U. S. A. 99:2094–9.

40. Bzymek M, Lovett ST. 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U. S. A. 98:8319–8325.

41. Nilsson Al, Zorzet A, Kanth A, Dahlström S, Berg OG, Andersson DI. 2006. Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. Proc. Natl. Acad. Sci. U. S. A. 103:6976–81.

42. Seoane A, Sánchez E, García-Lobo JM. 2003. Tandem amplification of a 28-kilobase region from the Yersinia enterocolitica chromosome containing the blaA gene. Antimicrob. Agents Chemother. 47:682–8.
43. McGann P, Courvalin P, Snesrud E, Clifford RJ, Yoon EJ, Onmus-Leone F, Ong AC, Kwak YI, Grillot-Courvalin C, Lesho E, Waterman PE. 2014. Amplification of aminoglycoside resistance gene aphA1 in Acinetobacter baumannii results in tobramycin therapy failure. MBio 5.

44. Magnet S, Courvalin P, Lambert T. 1999. Activation of the cryptic aac(6')-Iy aminoglycoside resistance gene of Salmonella by a chromosomal deletion generating a transcriptional fusion. J. Bacteriol. 181:6650–6655.

45. Zalkin H, Smith JL. 1998. Enzymes utilizing glutamine as an amide donor. Adv. Enzymol. Relat. Areas Mol. Biol. 72:87–144.

46. Wasserscheid J, Dewar K, Feng J, Ouellette M, Le D. 2009. Genome sequencing of linezolid-resistant Streptococcus pneumoniae mutants reveals novel mechanisms of resistance. Genome Res. 1214–1223.

47. Huang M, Graves LM. 2003. De novo synthesis of pyrimidine nucleotides; emerging interfaces with signal transduction pathways. Cell. Mol. Life Sci. 60:321–36.

48. De Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J. 2012. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. BMC Genomics 13:299.

49. Hoskins J, Alborn WE, Arnold J, Blaszczak LC, Burgett S, Hoff BSDE, Estrem ST, Fritz L, Fu D, Fuller W, Geringer C, Gilmour R, Glass JS, Khoja H, Kraft AR, Lagace RE, Blanc DJLE, Lee LN, Lefkowitz EJ, Lu JIN, Matsushima P, Ahren SMMC, Henney MMC, Leaster KMC, Mundy CW, Nicas TI, Norris FH, Gara MO, Peery RB, Robertson GT, Rockey P, Sun P, Winkler ME, Yang Y, Young-bellido M, Zhao G, Zook CA, Baltz RH, Jaskunas SR, Rosteck PR, Skatrud PL, Glass JI. 2001. Genome of the Bacterium Streptococcus pneumoniae Strain R6. Society 183:5709–5717.
Tables

Table 1. MICs (mg/L) of ciprofloxacin, norfloxacin and ethidium bromide for four R6<sup>M184</sup> transformants in the presence and absence of sodium orthovanadate.

|       | Cip | Cip + NaO | Nor | Nor + NaO | EtBr | EtBr + NaO |
|-------|-----|-----------|-----|-----------|------|------------|
| R6    | 0.5 | 0.25      | 1   | <1        | 2    | <1         |
| M184  | 2   | 0.5       | 8   | 2         | 16   | <1         |
| R6<sup>M184</sup> T1 | 2   | 0.25      | 8   | <1        | 16   | <1         |
| R6<sup>M184</sup> T2 | 2   | 0.25      | 8   | <1        | 16   | <1         |
| R6<sup>M184</sup> T3 | 2   | 0.25      | 8   | <1        | 16   | <1         |

Cip, ciprofloxacin; Nor, norfloxacin; EtBr, ethidium bromide; NaO, 50 μM sodium orthovanadate. MIC values are the mode of at least three independent experiments. Bold text indicates MIC values that are two or more dilutions greater than R6.
Table 2. Bacterial strains used in this study.

| Strain     | Description                                      | Reference |
|------------|--------------------------------------------------|-----------|
| R6         | Unencapsulated wild-type strain derived from D39 | (49)      |
| M169       | Spontaneous reserpine resistant mutant           | (7)       |
| M184       | Reserpine resistant R6 transformant of M169      | (7)       |
| R6<sup>M184</sup> T1 | EtBr resistant R6 transformant of M184 | This study |
| R6<sup>M184</sup> T1 | EtBr resistant R6 transformant of M184 | This study |
| R6<sup>M184</sup> T1 | EtBr resistant R6 transformant of M184 | This study |
| M240       | R6<sup>patB::magellan2</sup>                     | (7)       |
| M246       | R6<sup>patA::magellan2</sup>                     | (7)       |

EtBr, ethidium bromide
Table 3. Details of PCR primers used in this study

| Pair | Forward primer sequence | Reverse primer sequence | Amplicon |
|------|--------------------------|-------------------------|----------|
| 1    | CTTATTGGTGGGGAGAAGAA     | GATAACGCGGTTGCAAGAGT    | Duplication junction |
| 2    | GATAGGGCAGAAGAGCATCC     | GATAACGCGGTTGCAAGAGT    | Intergenic region between *hexA* and *patA* |
| 3    | TCTTGCTCAGTCCATCATCGAATAT | CAGCATGGTTCCTTGC       | *patA* copy 1 |
| 4    | CTTATTGGTGGGGAGAAGAA     | CAGCATGGTTCCTTGC       | *patA* copy 2 |
| 5    | ATGGGTCCTGCGCATCGATAT    | CAGCATGGTTCCTTGC       | *patA* internal region containing *megalion2* insertion |
| 6    | TAGATCGACGAGCCAGATT      | GCTAGTTGAACGCTCCATC     | Cloning site of plasmid pBAV1K2 |
| 7    | TAAGAATTCTTTATGAGCGATAGAAGCGATCGAAGT | GATAACGCGGTTGCAAGAGT   | Duplication junction, adding 5' EcoRI site (underlined text) |
Legends to Figures

Figure 1. Expression of patA and patB and accumulation of ethidium bromide in M184 and R6^{str} transformants. A. Expression of patA (black bars) and patB (grey bars) relative to rpoB. B. Accumulation of ethidium bromide after ten minutes of incubation with 100 μM ethidium bromide. *, accumulation significantly lower than that of R6, p<0.05.

Figure 2. Evidence of genomic rearrangements from improperly paired reads. A. Depth of reads aligned to the hexA and patA region following removal of properly paired reads by filtering for M184 (solid line) and R6 (dashed line). B. Matches found by Blast between a contig assembled from improperly paired reads against the R6 genome.

Figure 3. Evidence of duplication of a genomic region including patA and patB from comparison of per gene normalized read depths. A. Per gene RPKM values for M184 plotted against R6. Six genes (filled circles) appeared to be present at higher copy number in M184 than in R6 than the remaining genes (grey crosses). On further inspection these genes were contiguous, and included patA and patB (indicated). The hexA gene immediately upstream of patA (filled triangle) was not included in the higher copy number region. B. The same six genes were present at increased copy number in M184 and the three R6^{str} transformants. The slope of the regression line calculated for the high copy number genes (upper) was divided by the slope of the regression line for the remainder of the genome (lower) to obtain an estimate of the copy number of the amplified genes.

Figure 4. Measurement of expression of genes contained within the duplicated region of M184. A. Representation of the genomic region around patA and patB. Dotted lines represent the extent of the duplicated region in M184 and the three transformants. B. Expression of
seven genes contained within the duplication measured by qRT-PCR from R6 (black bars) and M184 (white bars). Error bars represent the standard deviation of three biological replicates.

**Figure 5.** Measurement of activity of the *patA* promoter with and without the tRNA gene found upstream of *patA* copy two in strains containing the duplication. **A.** Representation of the cloning of the tRNA gene and *patA* promoter upstream of a promoterless *gfp* gene in vector pBAV1K2. **B.** Fluorescence levels measured from R6 cells containing pBAV1K2 (Empty vector), pBAV1K2p (p), and pBAV1K2tp (tp). Error bars represent the standard deviation of three biological replicates. *, fluorescence significantly greater than from pBAV1K2p, one-tailed Student’s t-test, p<0.05.
a

b

|       | Upper |      | Lower |      | Ratio |
|-------|-------|------|-------|------|-------|
|       | Slope | $R^2$| Slope | $R^2$|       |
| M184  | 0.91  | 0.971| 0.47  | 0.95 | 1.94  |
| T1    | 0.97  | 0.995| 0.62  | 0.965| 1.56  |
| T2    | 1.86  | 0.993| 0.78  | 0.973| 2.41  |
| T3    | 1.17  | 0.955| 0.51  | 0.901| 2.29  |
