Targeted control of toxin production by a mobile genetic element in *Streptococcus pneumoniae*.

Emily Stevens[1], Daniel J. Morse[1], Dora Bonini[1], Seána Duggan[1], Tarcisio Brignoli[1], Mario Recker[2], John A. Lees[3], Nicholas J. Croucher[3], Stephen Bentley[4], Daniel J. Wilson[5], Sarah G. Earle[5], Robert Dixon[5], Angela Nobbs[6], Howard Jenkinson[6], Tim van Opijnen[7], Derek Thibault[7], Oliver J. Wilkinson[8], Mark S. Dillingham[8], and Ruth C. Massey[1]*.

1: School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD, UK.

2: College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK.

3: MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, St. Mary's Campus, Imperial College London, London, W2 1PG, UK.

4: The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

5: Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.

6: Bristol Dental School, University of Bristol, Bristol, BS1 2LY, UK.

7: Biology Department, Boston College, Chestnut Hill, MA, USA.

8: DNA-Protein Interactions Unit, School of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK.

* Corresponding author (email: ruth.massey@bristol.ac.uk)

Introductory Paragraph

Streptococcus pneumoniae is a major human pathogen that can cause severe invasive diseases such as pneumonia, septicaemia and meningitis[1-3]. Young children are at a particularly high risk, with an estimated 800,000 deaths worldwide in those under five attributable to invasive pneumococcal disease each year[1-3]. The cytolytic toxin pneumolysin (Ply) is a primary virulence factor for this bacterium, however, despite its importance to both the colonisation and pathogenic capabilities of this pathogen, the regulation of its expression is not well defined[4-7]. Using a genome-wide association approach we identified over a hundred potential affectors of Ply activity, including the Integrative and Conjugative Element (ICE) ICE*Sp*23FST81[8]. This regulatory effect is mediated through the activity of a novel modular protein, ZomB, which has an N-terminal UvrD-like helicase domain followed by two Cas4-like nuclease domains. The ZomB protein has potent ATP-dependent nuclease activity and binds specifically to the DNA containing a BOX repeat region that forms part of the *ply* operon. We hypothesise that with over 100 BOX regions across the pneumococcal genome, the acquisition of the *zomB* gene on ICE*Sp*23FST81 has the potential to re-wire the transcriptional landscape of this major human pathogen.

Main Text

Ply is a cytolytic pore-forming toxin that binds to cholesterol in the membranes of host cells, where it inserts into the lipid bilayer forming a transmembrane pore that ultimately causes lysis of the host cell[4-7]. Ply has also been shown to interfere with several aspects of the host immune system[9,10], been implicated in the initial colonization of the nasopharynx[11], and shown to promote transmission of the bacteria to new hosts[12]. Given its importance, we sought to define the genetic basis of Ply activity by analysing a collection of 165 isolates belonging to the *S. pneumoniae* PMEN1 lineage. We constructed a *ply* mutant in the pneumococcal isolate ATCC 700669[8] by replacing the *ply* gene with an erythromycin resistance cassette. Using this wild type and mutant strain as positive and negative controls we quantified the Ply activity (lysis of sheep red blood cells (RBCs)) of 165 clinical isolates in

triplicate, demonstrating that it varied significantly across this collection of closely related isolates (fig. 1a).

As the genomes of each of these isolates has been sequenced[13], we applied three complementary genome-wide association (GWAS) approaches to identify loci associated with Ply activity. In addition to a linear regression approach using the SNP (single nucleotide polymorphism) data, we also applied two methods that make use of kmer data: *BugWAS*[14] and *pyseer*[15]. The results from each of these methods are presented in supplementary tables 1-3. A Manhattan plot in Fig. 1b shows the significantly associated genetic loci determined through the linear regression method, highlighting those loci where the GWAS methods agree. There were two notable observations from these analyses. The first was that five loci were associated with Ply activity across all three methods, with the *pbpX* gene which encodes the Penicillin Binding Protein 2x being the most significantly associated locus. This was followed in order of significance by the intergenic region between a gene with the locus tag SPN23F05820 and *bgaA,* a gene with the locus tag SPN23F00840, the intergenic region between a gene with the locus tag SPN23F19120 and *msmG,* and the intergenic region between a gene with the locus tag SPN23F14800 and *greA*. What is striking about these four latter loci is that they each contain a BOX repeat region, which are a family of short repetitive sequences of DNA unique to the pneumococcal species[16]. Interestingly, a further 67 BOX containing regions were associated with Ply activity across the three GWAS analyses. The second notable observation was that 48 individual genes or intergenic regions on the Integrative and Conjugative Element (ICE) ICE*Sp*23FST81 were associated with Ply activity.

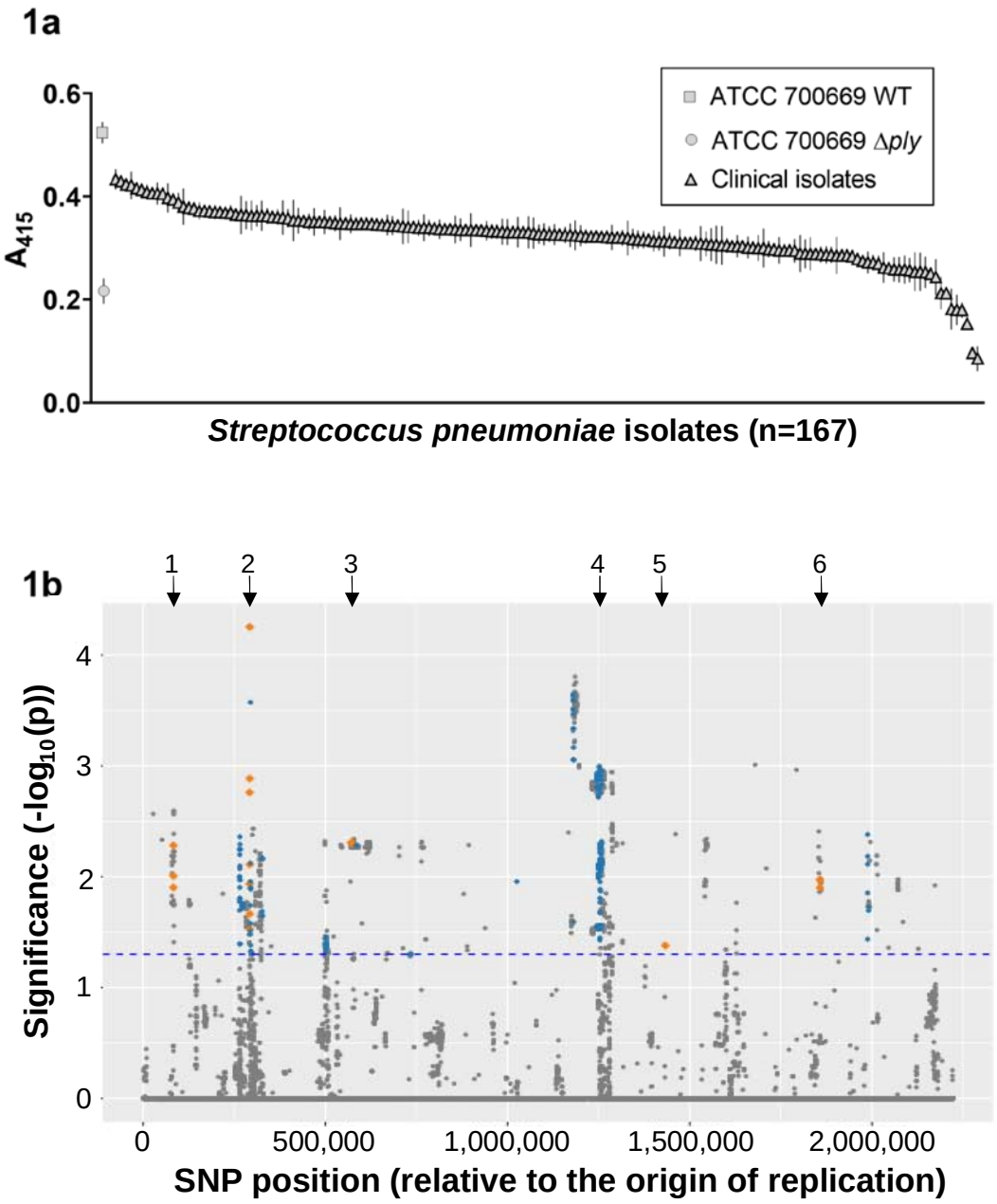


Fig. 1: Genome wide association study (GWAS) to identify novel effectors of pneumolysin (Ply) production by *S. pneumoniae*. (**a**) Pneumolysin activity of 165 *S. pneumoniae* of the PMEN1 lineage, as measured by cell lysis. A wild type (WT) and isogenic pneumolysin mutant (Δ*ply*) have been included as controls. (**b**) Manhattan plot of the GWAS results using the linear regression method. The blue dotted line indicates the threshold for significance. SNPs in loci identified by two or all three GWAS methods are indicated in blue and orange respectively. The black arrows indicate the regions of interest; 1: SPN23F00840, 2: *pbpX,* 3: intergenic between SPN23F05820 and *bgaA*, 4: ICE*Sp*23FST81, 5: intergenic between SPN23F14800 and *greA,* and 6: intergenic between SPN23F19120 and *msmG.*

Fitness trade-offs between antibiotic resistance and virulence in bacteria are well established[17-20], and for both clinical isolates and isogenic mutants the acquisition of penicillin resistance has been shown to decrease the virulence of *S. pneumoniae* in murine models of infection[21,22]. Given the association of the *pbpX* gene and Ply activity by all three GWAS methods, we hypothesised that a similar trade-off may be occurring here, such that the polymorphisms in the *pbpX* gene may be increasing the levels of resistance to penicillin, which may as a consequence reduce the levels of Ply being produced, or vice versa. Although bacterial GWAS results are typically validated through mutation of the associated locus, the contribution of the protein encoded by *pbpX* to the biosynthesis of the peptidoglycan layers in the bacterial cell wall is such that it is essential and cannot be inactivated. Instead, we tested our hypothesis by examining the levels of resistance to penicillin (minimum inhibitory concentrations, MICs) for the isolates, but found no significant correlation between the MICs and Ply activity (Pearson product-moment correlation r^2 = 0.076, P = 0.36). An alternative hypothesis is that the polymorphisms in the *pbpX* gene affects the stem peptide composition of peptidoglycan, as this has been shown to inhibit the release of Ply from the bacterial cells[23]. Further work to test this hypothesis is underway.

The high number of associated loci on ICE*Sp*23FST81 is particularly intriguing. This mobile genetic element (MGE), which can be found both integrated and in plasmid form (illustrated in fig. 2a), is believed to be critical to the success of this lineage of *S. pneumoniae* due to the antibiotic resistance capabilities it brings to the bacteria[8]. Given its ability to move horizontally as a single contiguous unit between bacteria, it is likely that of the associated loci only one is an affector of Ply activity, whilst the others are associated through their physical linkage to this. A survey of the putative activity of all 48 associated loci revealed that the majority of these are genes typical to such elements, involved in antibiotic resistance and the mechanics of its movement. An interesting exception is a gene with the locus tag SPN23F12470. This locus has been annotated as encoding a UvrD-like helicase, a family of proteins typically associated with core house-keeping activities for bacteria. Further *in silico*

analysis of the encoded protein, that we have named ZomB, suggests that it is a multi-domain protein with a putative DNA binding helicase domain, followed by two Cas4-like nuclease domains which are predicted to harbour 4Fe-4S clusters[24] (fig. 2b). With several known examples of Cas-like proteins regulating bacterial virulence[25-27], we hypothesised that it is this gene on ICE*Sp*23FST81 that is the affector of Ply activity.

To establish the role of ZomB in Ply activity we replaced the *zomB* gene with an erythromycin resistance cassette in *S. pneumoniae* strain ATCC 700669. The RBC lytic activity of the *zomB* mutant was significantly impaired relative to the wild type strain (fig. 2c). Using anti-Ply antibodies in a Western blot, we further demonstrated that this reduction in lytic activity was due to a decrease in the amount of Ply protein being produced by the bacteria (fig. 2d). We also quantified the relative transcription of the *ply* gene by qRT-PCR and found that the reduced abundance of Ply was likely due to a significant decrease in *ply* transcription (fig. 2e). Multiple unsuccessful attempts to construct a strain containing the *zomB* gene on a plasmid to complement its activity in the mutant led us to conclude that high levels of expression of this protein are toxic to the bacterial cells.

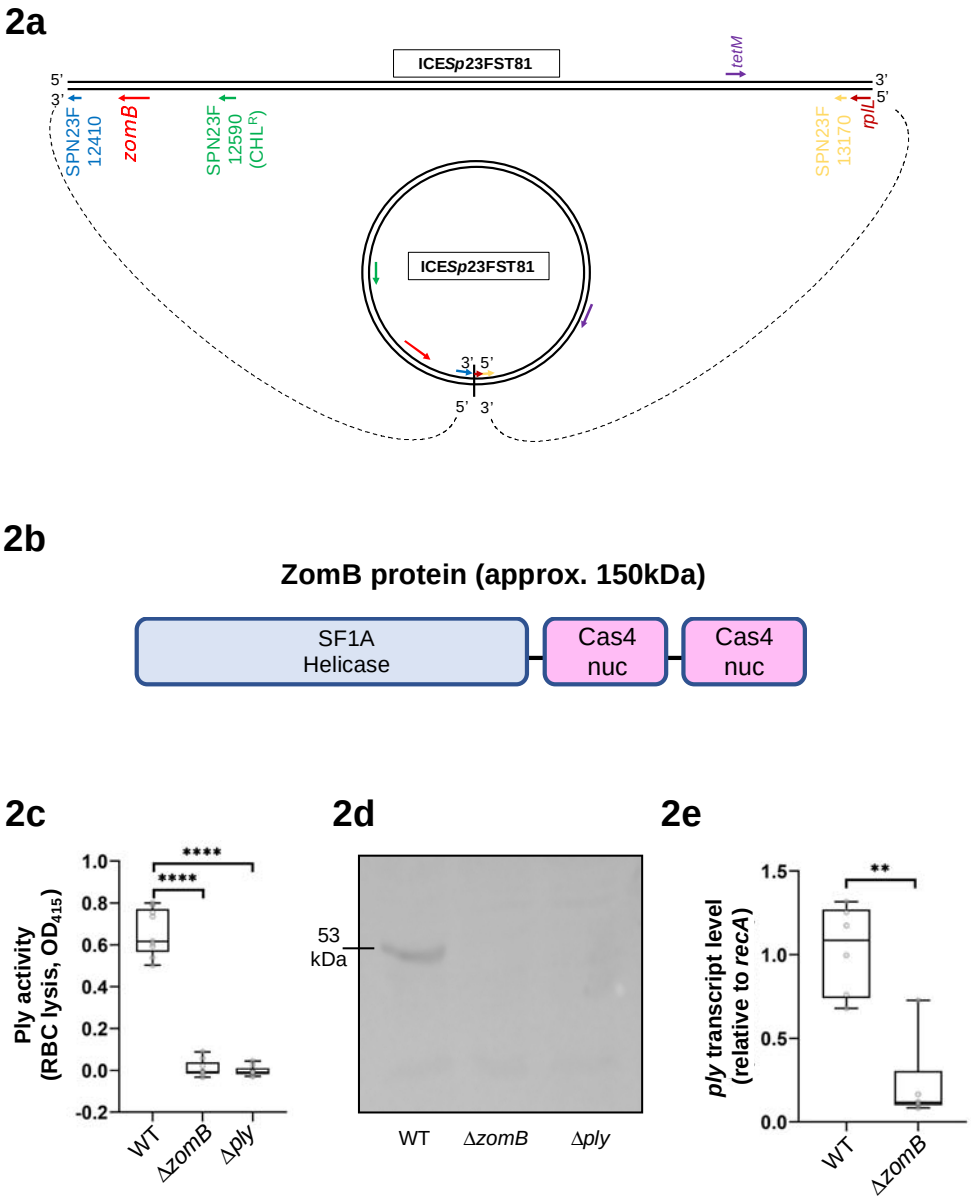


Fig. 2: Inactivation of the *zomB* gene on the ICE*Sp*23FST81 affects pneumolysin production. **(a)** Cartoon illustration of ICE*Sp*23FST81 in both its linear chromosomally integrated form and circularised plasmid form. Some genes of interest have been included: SPN23F12410, SPN23F13170 and *rplL* because these flank the element in its linear form and come into close association when in the plasmid form. Genes encoding the antibiotic resistance gene are also indicated. **(b)** Schematic of the ZomB protein with its helicase and two Cas4-like nuclease domains indicated. (**c**) the lytic activity of the *zomB* mutant is comparable to that of the *ply* mutant as determined by a sheep RBC lysis assay. (**d**) The presence of Ply in the extracellular medium is reduced in the *zomB* mutant, determined using anti-Ply antibodies in a Western blot on concentrated bacterial supernatant. (**e**) The effect the inactivation of *zomB* has on Ply production is mediated at a transcriptional level, as determined by qRT-PCR. This data was made relative to the expression of the

housekeeping gene *recA* in each sample, and normalised to the level of expression on the *ply* gene in the wild type strain. In the box plots each dot represents an individual replicate data point.

The *ply* gene is transcribed as part of an operon with four other genes that encode a transcriptional regulator (SPN23F19460) and proteins implicated in the movement of Ply from the cytosol to the bacterial cell wall (SPN23F19460-19500) (fig. 3a) [28]. It also contains a BOX repeat region immediately downstream of the *ply* gene. Due to their internal repeating sequences, BOX regions can form stable secondary structures, and their presence has been associated with effects on the transcription of neighbouring genes[29]. The molecular details of how their presence affects gene transcription has not yet been determined but is likely due to these secondary structures where they can either enhance or interfere with transcription processes depending on their relative positioning[30]. As ZomB appears to be a positive effector of *ply* transcription, and given its likely DNA binding capability, we hypothesised that it may be interacting with this BOX regions to elicit its effect on the expression of *ply*.

To characterise the biochemical activities of the ZomB protein, we expressed and purified recombinant ZomB with an N-terminal 6x histidine tag (fig. 3b). Using a coupled-assay, we found that ZomB hydrolyses ATP with Michaelis-Menten kinetics and displays a turnover number of approximately 20 s^{-1} and a K_m value of 90 µM ATP (fig. 3c). These experiments were performed in the presence of saturating quantities of ssDNA, which was shown to strongly stimulate ATP hydrolysis with an apparent dissociation constant of approximately 1 µM (ntds) (fig. 3d). This behaviour is typical of the UvrD-like DNA helicases of which ZomB is a member[31]. To examine the putative nuclease activity, ZomB protein (50nM) was incubated with duplex DNA with and without ATP and Mg^{2+}. Degradation of the DNA by ZomB was monitored by gel electrophoresis, which showed that the protein possesses a potent ATP-dependent nuclease activity (fig. 3e).

We next amplified the region of DNA containing the BOX element from within *ply* operon, as well as an equivalently-sized region of control DNA within the *ply* coding region, and performed electrophoretic mobility shift assays (EMSAs) with increasing concentrations of ZomB protein. As visualised in fig. 3f, the ZomB protein caused a shift in the BOX containing DNA even at the lowest concentration of protein used, demonstrating its high affinity for this substrate. There was also a shift in mobility of the control, *ply*-encoding DNA, but this occurred at higher concentrations of ZomB and resulted in the formation of less distinctly shifted DNA bands, which is indicative of non-specific binding. The binding activity of ZomB was not specific to this BOX containing region as we found that it also bound with high affinity to a second BOX containing region of DNA situated approximately 3kb upstream of the *ply* gene (supp. fig. 1). These results are consistent with ZomB displaying binding specificity for BOX elements. Although we have still much to learn about the *in vivo* biochemical functions of ZomB, we believe it likely that its nuclease activity may unwind the secondary structures that result from the BOX repeats, thereby relieving the effects these have on the transcription of neighbouring genes.

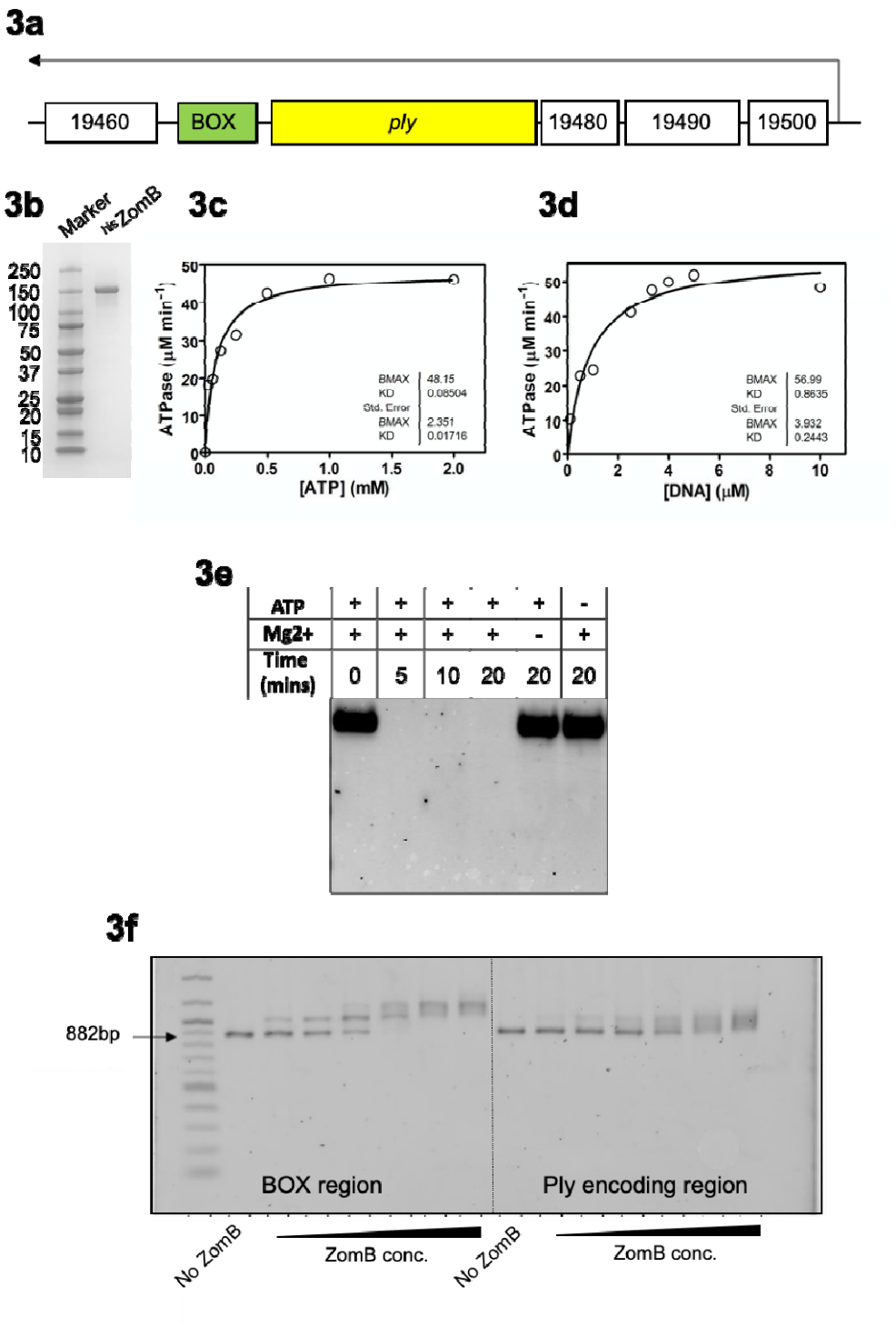


Fig. 3: The ZomB protein specifically binds to the BOX region within the Ply encoding operon. (**a**) cartoon of the *ply* encoding operon with the ply gene, the SPN23F- locus tags of the neighbouring genes, and promoter and transcript length (black arrow) indicated. (**b**) An SDS-PAGE gel showing the purified his tagged ZomB protein. (**c**) Steady-state ATPase activity of ZomB (50 nM) was measured at saturating ssDNA concentration to determine the

Michaelis-Menten parameters. (**d**) Steady-state ATPase activity of ZomB is strongly stimulated by ssDNA with an apparent dissociation constant of approximately 1 µM ntds. (**e**) Nuclease assays were performed with linear DNA demonstrating that it was degraded by the ZomB protein in the presence of both ATP and divalent cations. (**f**) EMSAs demonstrating the specificity of binding of a BOX element by the ZomB protein. An equivalent length of DNA from with the *ply* coding region was used as a control. The concentrations of ZomB used were 2.5, 5, 10, 20, 40 and 80nM.

Through the application of a functional genomics approach to a large collection of sequenced *S. pneumoniae* isolates, we have identified over a hundred novel putative affectors of Ply activity (supp. tables 1-3). Of these associated loci we have determined the molecular detail of the interaction between the ZomB protein encoded on ICE*Sp*23FST81 and the BOX repeat region within the *ply* operon. This however does not explain why so many other BOX containing regions were associated with Ply activity. One potential explanation we are exploring is whether the sequence variability within the associated BOX regions affects the binding affinity of ZomB to them, which would affect the availability of ZomB for binding to the BOX region within the *ply* operon. It will also be interesting to see with future work how ZomB contributes to the lifecycle of the ICE*Sp*23FST81, as this MGE does not contain any BOX-like regions, suggesting that if ZomB regulates the expression of any genes on the ICE, it uses a different sequence motif to achieve this.

Through their carriage of cargo genes, MGEs confer significant advantages to their bacterial hosts that enhance the success of both. ICE*Sp*23FST81 confers resistance to both chloramphenicol and tetracycline, which provides a clear benefit upon exposure to antibiotics. However, for many bacterial species antibiotic resistance often incurs a fitness cost in the absence of the antibiotic, which can be offset, for example, by reducing the energetically costly production of toxins[17,18]. Here, in stark contrast, we observe the simultaneous acquisition of increased resistance with increased toxin production. Future work will explore how these combined costs are compensated from the perspective of both the host cell and MGE, to result in such a globally successful pathogen.

Acknowledgements

ES was a BBSRC SWBio DTP funded PhD student. R.C.M is a Wellcome Trust funded Investigator. D.J.W. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (Grant 101237/Z/13/B) and is supported by a Big Data Institute Robertson Fellowship. We would also like to thank Prof. Tim Mitchell for helpful discussions during the preparation of this manuscript.

Materials and Methods

Bacterial strains and growth conditions.

Clinical isolates used in this study (listed in supp. table 4) had been previously sequenced[13] and all belonged to the PMEN1 clone of which *S. pneumoniae* ATCC 700669 is the reference strain. Strains were grown for 16-24 hours in 5% CO_2 at 37°C, on either brain-heart infusion (BHI) blood agar plates containing 5% defibrinated horse blood, or in BHI broth without blood.

Construction of deletion mutants in *S. pneumoniae*.

Genes were deleted in *S. pneumoniae* using linear PCR products as described previously[32]. In brief, a stitch PCR approach using the primers listed in table 1 were used to generate a single PCR product consisting of 1kb to either side if the gene to be deleted, with a gene encoding resistance to erythromycin (*ermAM* - amplified from plasmid pVA838) in the centre. To transform the bacteria with this PCR product, the wild type bacteria were grown overnight in 5ml BHI broth, and 2ml of this culture was used to inoculate a pre-warmed tube of 20ml BHI broth. This was incubated for a further 2 hours, and 0.5ml was added to 9.5ml pre-warmed BHI broth and incubated for 30 minutes. 1ml of this sub-culture was transferred to sterile tubes, 10µl of 10µg/ml competence-stimulating peptide-2 (CSP-2) was added each, and these were incubated for 10-15 minutes. To 200µl aliquots of this either the PCR product, or sterile water was added and the mixture incubated for a further 2 hours. This was then added to molten BHI agar (10ml) containing 5% defibrinated horse blood, allowed to set

in a petri dish and incubated for 2 hours. Plates were then overlaid with a further 10ml molten agar containing 5% defibrinated horse blood with erythromycin (2µg/ml) to selected for successfully transformed cells. Plates were incubated for up to five days at 37°C, or until colonies appeared within the agar.

Amplification site	**Primer Sequence**
ply (LHS) forward	5' CCCTTGCTCTGGTTAAAAAAAGAAGC 3'
ply (LHS) reverse	5' ATATTTTTGTTCATATTTGCCATCTTCTACC 3'
ery-ply forward	5' GGTAGAAGATGGCAAATATGAACAAAAATATAAAA 3'
ery-ply reverse	5' CTACCTGAGGTTATTTCCTCCCGTT 3'
ply (RHS) forward	5' GAGGAAATAACCTCAGGTAGAAGATAAG 3'
ply (RHS) reverse	5' GATCACCTTTTTTAGCTGCTACATAG 3'
zomB (LHS) forward	5' TGCCCACTATTTTTATCTAGTTGCTTACC 3'
zomB (LHS) reverse	5' ATTTTTGTTCATTGTTGTCATCGTTTTACCTC 3'
ery-zomB forward	5' CGATGACAACAATGAACAAAAATATAA 3'
ery-zomB reverse	5' CTTTTCCGGATTCTTATTTCCTCCC 3'
zomB (RHS) forward	5' GGAGGAAATAAGAATCCGGAAAAG 3'
zomB (RHS) reverse	5' AATTAATTCCTGAATACAAGTTAACAAAATAG 3'

Quantification of Ply activity

Strains were grown overnight in BHI broth, then sub-cultured at a 1:1 ratio into fresh broth and grown for one hour until OD_{600} of 0.4-0.7 was reached. These cultures were then serially diluted 4-fold in a 96 well plate containing BSA assay buffer (0.05g bovine serum albumin and 77mg dithiothreitol dissolved in 50ml sterile phosphate buffered saline. To each well 50µl of triple washed sheep red blood cells were added and the plates were incubated for one hour in 5% CO2 at 37°C. The plates were then centrifuged at 2000rpm for 10 minutes at room temperature to separate the intact cells from the soluble lysed material. The

supernatants from each well were transferred to a fresh 96-well plate and absorbance values read at 596nm were obtained using a FLUOstar Omega microplate reader.

GWAS

The initial Genome-wide associations between single nucleotide polymorphisms (SNP) and bacterial toxicity were determined by means of linear regression. We performed a singular value decomposition of the SNP data and used the first four principal components, which together explained around 50% of the variance, in the linear regression model to adjust for population structure. Subsequent GWAS analyses of the data by pyseer and BugWas were performed as previously described[14,15].

Quantification of Ply production

The bacteria were grown overnight in 15ml BHI broth and harvested by centrifugation. Proteins were extracted from the cell pellets of by resuspending it in 1.5ml of whole cell lysate buffer (0.01% SDS, 0.1% sodium deoxycholate and 0.015M sodium citrate) followed by an incubation for 30 minutes at 37°C. The individual proteins were separated on an SDS-PAGE gels and Western blotting of this conducted using anti-pneumolysin antibody followed by goat anti-mouse IgG HRP was used as the secondary antibody. The HRP signal was detected using the Immun-Blot Opti-4CN Colourimetric kit from BioRad.

Quantification of transcription of the *ply* gene.

Bacteria were grown overnight in 5ml BHI broth and RNA was extracted using a Zymo Quick-RNA Fungal/Bacterial Miniprep kit following the manufacturer's protocol. Turbo DNase digest kit was used to remove any contaminating DNA. The quality and quantity of the RNA was determined using a Nanodrop and the RNA was reverse transcribed to cDNA using a qScript cDNA synthesis kit, as per the manufacturer's protocol. RNA samples were standardised such that 100ng was added to each cDNA synthesis reaction. cDNA was

subsequently quantified using a Nanodrop and where necessary was standardised again to ensure the same quantity in each sample. qPCR was performed using a Mic qPCR cycler (Bio Molecular Systems) and reactions were set up using a SensiFAST SYBR No-ROX master mix. Three technical repeats were conducted for each cDNA Sample and the data anlaysed using the $2^{(Ct\ ply - Ct\ recA)}$ method. The primer pairs were used where *recA* was used as the control housekeeping gene:

ply FW: GGCTGATTTCGCTGAACAAG

ply RV: CCTTGAGTTGTTCCATGCTG

recA FW: ATCGGAGATAGCCATGTTGG

recA RV: ATAGAGGCGCCAAGTTTACG

Expression and purification of ZomB.

The *zomB* gene was cloned into the expression plasmid pET15b, and expressed in E. coli strain BL21(DE3) with a 6x histidine tag at its N-terminus. The *E. coli* cells were grown at 37°C in LB broth containing 100µg/ml ampicillin to an OD(600) of around 0.5, whereupon cells weretemperature acclimatised to 18°C and gene expression was induced for 12 hours via the addition of 1mM IPTG. Cells were harvested by centrifugation for 20minutes at 6000rpm and resuspended in buffer containing 50mM Tris pH 7.5,150mM NaCl, 10% sucrose and 0.1mM PMSF. Cells were lysed by sonication in buffer containing 1mM TCEP, 500mM NaCl, 20mM imidazole and a proteaseinhibitor cocktail, and cell debris was removed by centrifugation for 30 minutes at21,500rpm. Cell lysate was loaded onto a 5ml HisTrap Nickel binding columnwhich had been equilibrated in HisB buffer (20mM Tris pH7.5, 500mM NaCl,1mM TCEP, 5% glycerol) plus 20mM imidazole. Proteins possessing His-tagswere eluted with a gradient from 20mM to 500mM imidazole over 32 minutes with a flow rate of 3ml/min. Protein elution was monitored by measuring absorbanceat 280nm. Fractions containing the most protein were pooled and applied to a HiTrap Heparin affinity column equilibrated with HepQ/B buffer (20mM TrispH7.5, 1mM TCEP) plus 100mM NaCl. Heparin-

binding proteins were eluted witha gradient from 100mM to 1M NaCl over 30 minutes, monitored by absorbance at280nm. 2ml fractions containing the highest protein concentrations were pooledand applied to a MonoQ anion exchange chromatography column equilibratedwith HepQ/B buffer plus 100mM NaCl. Proteins were eluted with a gradient from100mM to 1M NaCl over 30 minutes, with protein absorbance monitored at 280nm, and 0.3ml fractions were collected. Fractions containing the highest concentrations of eluted protein were pooled and stored in HepQ/B buffer plus 330nM NaCl, corresponding to the salt concentration at which the protein was eluted from the column. Nanodrop A280 and the ZomB protein's predicted extinction coefficient (118390mol/L/cm) were used to determine that ZomB had been purified to a concentration of 7.0μM.

Biochemical characterisation of ZomB activity.

ATPase activity was measured by coupling the hydrolysis of ATP to the oxidation of NADH which gives a change in absorbance at 340 nm. Reactions were performed in a buffer containing 20 mM Tris-Cl pH 8.0, 50 mM NaCl, 2 mM DTT, 1 mM $MgCl_2$, 50 U/mL lactate dehydrogenase, 50 U/mL pyruvate dehydrogenase, 1 mM PEP and 100 µg/mL NADH. Rates of ATP hydrolysis were measured over 1 min at 25℃ and the ssDNA substrate used was Poly(dT). For calculation of K_{DNA} (defined as the concentration of DNA at which ATP hydrolysis is half-maximal), the ATP concentration was fixed at 2 mM. The Michaelis-Menten plot was performed at saturating DNA concentration which is defined as 10x the K_{DNA} value. The concentration of ZomB was 50 nM in these assays unless indicated otherwise.

Electrophoretic mobility shift assays.

The intergenic region between *ply* and the neighbouring gene SPN23F19460, containing BOX repeat regions was amplified (using the following primers; BoxF: GAGAGGAGAATGCTTGCGAC and BoxR: TAGGAATCTCCTTTTTTCACATTTTAATCTTTC). A region of the *ply* gene of equivalent size was also amplified (using plyF: ATGGCAAATAAAGCAGTAAATGACTTTATAC and

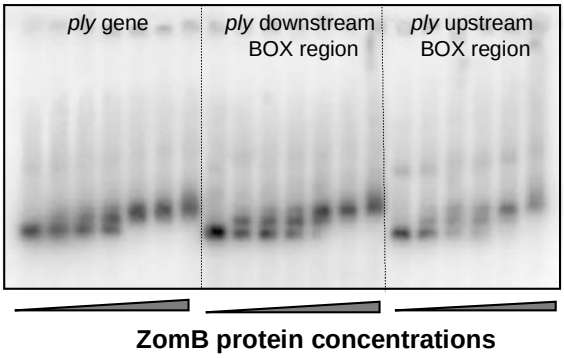
plyR: GCCCCCTAAAATAACCGCCTTC). The purified ZomB protein was added in a range of concentrations (5μM, 2.5μM, 1.25μM, 0.625μM, 0.3125μM and 0.15625μM) to 10nM of the PCR products in a buffer containing 20mM Tris (pH 8), 200mM sodium chloride, 1mM Tris(2-carboxyethyl) phosphine (TCEP) and 10% glycerol. Samples were then run on a 1.5% agarose gel in 1X TAE buffer for 110 minutes at 90V. Following this the gel was stained in TAE containing 1X SYBR Safe DNA gel stain for 30 minutes, and bands were visualised using a Typhoon FLA 9500.

REFERENCES

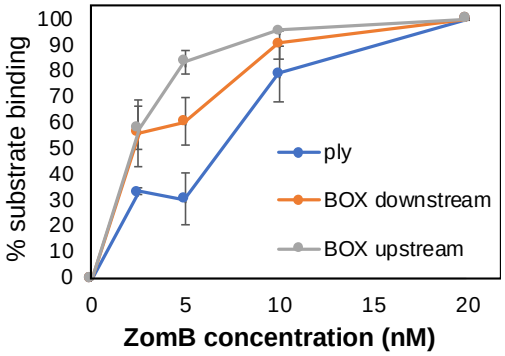
1. Deng, X. *et al*. Whole-genome sequencing reveals the origin and rapid evolution of an emerging outbreak strain of *Streptococcus pneumoniae* 12F. *Clin. Infect. Dis.* **62**, 1126-1132. (2016).
2. Ioachimescu, O.C., Ioachimescu, A.G. & Iannini, P.B. Severity scoring in community acquired pneumonia caused by *Streptococcus pneumoniae*: a 5-year experience. *Int. J. Antimicrob. Agents* **24**, 485-490 (2004).
3. O'Brien, K.L. *et al.* Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* **374,** 893-902 (2009).
4. Canvin, J.R. *et al.* The role of pneumolysin and autolysin in the pathology of pneumonia and septicaemia in mice infected with a type 2 pneumococcus. *J. Infect. Dis.* **172***,* 119-123 (1995).
5. Hirst, R.A. *et al.* The role of pneumolysin in pneumococcal pneumonia and meningitis. *Clin. Exp. Immunol.* **138**, 195-201 (2004).
6. Hirst, R.A. *et al. Streptococcus pneumoniae* deficient in pneumolysin or autolysin has reduced virulence in meningitis. *J. of Infect. Dis.* **197**, 744-751 (2008).
7. Mitchell, A.M. & Mitchell, T.J. *Streptococcus pneumoniae*: virulence factors and variation. *Clin.l Microbiol. Infect. Dis.* **16**, 411-418 (2010).
8. Croucher, N.J. *et al*. Role of Conjugative Elements in the Evolution of the Multidrug-Resistant Pandemic Clone *Streptococcus pneumoniae* Spain23F ST81. *J Bacteriol.* **191**, 1480-9 (2009).
9. Mitchell, T.J. *et al.* Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Mol. Microbiol.* **5**, 1883-8 (1991).
10. Paton, J.C., Rowan-Kelly, B. & Ferrante, A. Activation of human complement by the pneumococcal toxin pneumolysin. *Infect. Immun.* **43**, 1085-7 (1984).
11. Hotomi, M. *et al.* Pneumolysin plays a key role at the initial step of establishing pneumococcal nasal colonization. *Folia Microbiologica*. **61**, 375-83 (2016).
12. Zafar, M.A., Wang, Y., Hamaguchi, S. & Weiser, J.N. Host-to-Host Transmission of Streptococcus pneumoniae Is Driven by Its Inflammatory Toxin, Pneumolysin. *Cell Host Microbe.* **21**, 73-83 (2017).
13. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science*. **331**, 430-4 (2011).
14. Earle, S.G. *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat. Microbiol*. **1**, 16041 (2016).

15. Lees, J.A. *et al.* pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics*. **34**, 4310-2 (2018).
16. Martin, B. *et al.* A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. *Nucleic Acids Res*. **20**, 3479–83 (1992).
17. Rudkin, J.K. *et al*. Methicillin resistance reduces the virulence of healthcare-associated methicillin-resistant *Staphylococcus aureus* by interfering with the *agr* quorum sensing system. *J Infect Dis* **205**, 798–806 (2012).
18. Collins, J. *et al*. Offsetting virulence and antibiotic resistance costs by MRSA. *ISME J.* **4**, 577–84.
19. Hraiech, S. *et al*. Impaired virulence and fitness of a colistin-resistant clinical isolate of *Acinetobacter baumannii* in a rat model of pneumonia. *Antimicrob. Agents Chemother.* **57,** 5120–1 (2013).
20. Linares, J.F., *et al.* Overexpression of the multidrug efflux pumps MexCD-OprJ and MexEF-OprN is associated with a reduction of type III secretion in *Pseudomonas aeruginosa*. *J. Bacteriol.* **187**, 1384–91 (2005).
21. Azoulay-Dupuis, E. *et al.* Relationship between Capsular Type, Penicillin Susceptibility, and Virulence of Human *Streptococcus pneumoniae* Isolates in Mice. *Antimicrob. Agents Chemother*. **44**, 1575–7 (2000).
22. Rieux, V., Carbon, C. & Azoulay-Dupuis, E. Complex relationship Between Acquisition of Beta-Lactam Resistance and Loss of Virulence in *Streptococcus Pneumoniae*. *J Infect Dis.* **184**, 66-72 (2001).
23. Greene, N.G., Narciso, A.R., Filipe, S.R. & Camilli, A. Peptidoglycan Branched Stem Peptides Contribute to Streptococcus pneumoniae Virulence by Inhibiting Pneumolysin Release. *PLoS Pathog*. **11**, e1004996 (2015).
24. Zhang, J., Kasciukovic, T. & White, M.G. The CRISPR Associated Protein Cas4 Is a 5′ to 3′ DNA Exonuclease with an Iron-Sulfur Cluster. *PLoS One*. **7**, e47232 (2012).
25. Sampson, T.R. *et al.* A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*. **497**, 254-7 (2013).
26. Heidrich, N. *et al.* The CRISPR/Cas system in *Neisseria meningitidis* affects bacterial adhesion to human nasopharyngeal epithelial cells. *RNA Biol*. **16**, 390-396 (2019).
27. Ma, K. *et al.* cas9 Enhances Bacterial Virulence by Repressing the regR Transcriptional Regulator in *Streptococcus agalactiae*. *Infect Immun*. **86**, e00552-17 (2018).
28. Kimaro Mlacha, S.Z., *et al.* Phenotypic, genomic, and transcriptional characterization of *Streptococcus pneumoniae* interacting with human pharyngeal cells. *BMC Genomics*. **14**, 383 (2013).
29. Croucher, N.J., Vernikos, G.S., Parkhill, J. & Bentley, S.D. Identification, variation and transcription of pneumococcal repeat sequences. *BMC Genomics.* **12**, 120 (2011).
30. Knutsen, E. *et al.* BOX Elements Modulate Gene Expression in *Streptococcus Pneumoniae*: Impact on the Fine-Tuning of Competence Development. *J Bacteriol.* **188**, 8307-12 (2006).
31. Gilhooly, N.S., Gwynn, E.J. & Dillingham, M.S. Superfamily 1 Helicases. *Front Biosci.* **5**, 206-16 (2013).
32. Chen, L., Ge, X. & Xu, P. Identifying essential *Streptococcus sanguinis* genes using genome-wide deletion mutation. *Methods Mol. Biol*. **1279**, 15-23 (2015).

Supp 1a



Supp 1b



Supplementary Figure 1: ZomB binds to both BOX regions that flank the *ply* gene. EMSA was performed as described previously, but lower concentrations of both ZomB 2.5-80nM and DNA (50nM), and with the DNA radiolabelled gamma-32P-ATP to facilitate visualisation. A third 882bp PCR product was amplified for this experiment containing the BOX repeat regions upstream of ply, using the following primer sequences:
FW: CAGTAAGATACAGACTGCAAGGC
RV: TGAAAAAGAAACCATTCAGGGGTATTCGG
50nM of each of the three PCR product was added to a reaction containing 2µl phosphorous-32 (gamma-32P-ATP), 2µl T4 polynucleotide kinase (PNK) and 5µl PNK buffer,made up to 50µl with water. This reaction mix was incubated at 37°C for 1 hour, then at 65°C for 20 minutes to heat inactivate the PNK enzyme. Radiolabelled DNA was purified using Microspin S-400 columns as per the manufacturers protocol. 2nM DNA was used in an EMSA exactly as described above, using the protein concentrations from 2.5-80nM, with the exception that the gel was not stained with SYBR safe, but squashed for three hours before being dried using a vacuum-heated gel dryer (BioRad). Bands were visualised as before using a Typhoon FLA 9500.