Author/s:
Vergel L. Torres, V;Heinz, E;Stubenrauch, CJ;Wilksch, JJ;Cao, H;Yang, J;Clements, A;Dunstan, RA;Alcock, F;Webb, CT;Dougan, G;Strugnell, RA;Hay, ID;Lithgow, T

Title:
An investigation into the Omp85 protein BamK in hypervirulent Klebsiella pneumoniae, and its role in outer membrane biogenesis

Date:
2018-09-01

Citation:
Vergel L. Torres, V., Heinz, E., Stubenrauch, C. J., Wilksch, J. J., Cao, H., Yang, J., Clements, A., Dunstan, R. A., Alcock, F., Webb, C. T., Dougan, G., Strugnell, R. A., Hay, I. D. & Lithgow, T. (2018). An investigation into the Omp85 protein BamK in hypervirulent Klebsiella pneumoniae, and its role in outer membrane biogenesis. Molecular Microbiology, 109 (5), pp.584-599. https://doi.org/10.1111/mmi.13990.

Received Date:

Revised Date:

Accepted Date:

Article Type: Research Article

An investigation into the Omp85 protein BamK in hypervirulent *Klebsiella pneumoniae*, and its role in outer membrane biogenesis.

Von Vergel L. Torres[1], Eva Heinz[1,2], Christopher J. Stubenrauch[1], Jonathan J. Wilksch[1,3], Hanwei Cao[3], Ji Yang[3], Abigail Clements[1,4], Rhys A. Dunstan[1,2], Felicity Alcock[1,5], Chaille T. Webb[1], Gordon Dougan[2], Richard A. Strugnell[3], Iain D. Hay[1] & Trevor Lithgow[1,*]

1 - Infection & Immunity Program, Biomedicine Discovery Institute and Department of Microbiology, Monash University, Clayton 3800, Australia

2 - Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, United Kingdom

3 - Department of Microbiology & Immunology, University of Melbourne, Parkville 3052, Australia

4 - MRC Centre for Molecular Bacteriology and Infection, Department of Life Sciences, Imperial College, London, UK

5 - Department of Biochemistry, Oxford University, Oxford UK

* - To whom correspondence should be addressed: trevor.lithgow@monash.edu

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mmi.13990

Keywords: BAM complex, BamA, hypervirulence, cell wall, beta-barrel proteins

ABSTRACT

Members of the Omp85 protein superfamily have important roles in Gram-negative bacteria, with the archetypal protein BamA being ubiquitous given its essential function in the assembly of outer membrane proteins. In some bacterial lineages, additional members of the family exist and, in most of these cases, the function of the protein is unknown. We detected one of these Omp85 proteins in the pathogen *Klebsiella pneumoniae* B5055, and refer to the protein as BamK. Here we show that *bamK* is a conserved element in the core genome of *Klebsiella*, and its expression rescues a loss-of-function Δ*bamA* mutant. We developed an *E. coli* model system to measure and compare the specific activity of BamA and BamK in the assembly reaction for the critical substrate LptD, and find that BamK is as efficient as BamA in assembling the native LptDE complex. Comparative structural analysis revealed that the major distinction between BamK and BamA is in the external facing surface of the protein, and we discuss how such changes may contribute to a mechanism for resistance against infection by bacteriophage.

INTRODUCTION

The genus *Klebsiella* are Gram-negative bacteria, widely distributed in the environment, most commonly in association with soil and plants (Bagley, 2015). In just a few decades, the species *Klebsiella pneumoniae* has evolved from this lifestyle in the environment to become a common and deadly cause of infection in humans (Paczosa & Mecsas, 2016). While initially only seen associated with chronically unwell people *e.g.* liver abscesses in chronic alcoholics (Carpenter, 1990), *Klebsiella* are adept at lateral gene transfer and this, and other genetic means such as plasmid-mediated gene acquisition, has seen the evolution of *K. pneumoniae* strains that thrive in hospital settings (Ramirez *et al.*, 2014, Wyres & Holt, 2016, Navon-Venezia *et al.*, 2017). *Klebsiella* establishes itself in environments, including sinks and other wet areas in hospitals as well as human tissues, through the expression of surface factors that include a mucoid capsule and an assortment of adherent fimbriae (Paczosa & Mecsas, 2016). These factors assist the bacteria to grow and survive within organized communities known as

biofilms, with biofilms of *K. pneumoniae* being central to many chronic infections, particularly those associated with indwelling medical devices (Stickler, 2008, Hall-Stoodley *et al.*, 2004).

Antimicrobial resistance is a major concern with *K. pneumoniae*, with invasive and blood-borne infections resistant to many antibiotics (Wyres & Holt, 2016, Lee *et al.*, 2016a, Campos *et al.*, 2016, Bi *et al.*, 2017, Calfee, 2017, Ejaz *et al.*, 2017). Amongst the notifications to the Centers for Disease Control and Prevention for clinically-diagnosed carbapenem-resistant Enterobacteriaceae, an increasing proportion is found to be *K. pneumoniae* that are resistant to all available antimicrobial drugs (Chen *et al.*, 2017). In addition to drug-resistance, hypervirulent variants of *K. pneumoniae* have been emerging over the past years (Lee *et al.*, 2017). This hypervirulence is provided in part by a hypermucoviscous phenotype, caused by extensive secretion of polysaccharides that remain attached to the bacterial cell surface as an extracellular capsule. For example, infection of mice with 10^2 cfu of the *K. pneumoniae* strain B5055 kills all mice within 5 days, leading to its designation as hypervirulent (Kumar & Chhibber, 2011). Whereas B5055 thrives in murine serum, an isogenic mutant lacking the capsule secretion apparatus is totally inactivated by serum within an hour (Clements *et al.*, 2008). The Wza capsule secretion pore appears to self-assemble in the outer membrane (Dunstan *et al.*, 2015), whereas most other outer membrane proteins are assembled by a highly sophisticated β-barrel assembly machinery (BAM) composed of the core BAM complex (BAM_{ABCDE}) and the translocation and assembly module (the TAM) (Webb *et al.*, 2012a, Plummer & Fleming, 2016, Albenne & Ieva, 2017, Noinaj *et al.*, 2017). Each of these protein complexes is constituted around a membrane protein of the Omp85 protein family (BamA and TamA), and both are involved in the assembly of complex structures such as the adhesive fimbriae (Stubenrauch *et al.*, 2016, Stubenrauch *et al.*, 2017) required by *K. pneumoniae* for biofilm formation (Klemm & Schembri, 2000, Wilksch *et al.*, 2011, Khater *et al.*, 2015, Tan *et al.*, 2015).

A recent survey of the Omp85 family showed 10 sub-families that have the diagnostic β-barrel domain linked to either POTRA domains of distinct number and sequence characteristics, or to other domains that range from predicted proteases through to no N-terminal domains in the noNterm sub-family (Heinz & Lithgow, 2014). BamA has five N-terminal POTRA domains which facilitate interactions with the other Bam subunits, while TamA has three N-terminal POTRA domains that assist its interaction with a distinct TamB

partner protein (Selkrig *et al.*, 2015, Shen *et al.*, 2014). Analysis of the genome of the *K. pneumoniae* strain B5055 revealed Omp85 family members additional to the genes encoding BamA and TamA, including a form without POTRA domains (noNterm, BN49_0007), and a previously undescribed Omp85 protein that we refer to as BamK. We demonstrate that the gene encoding BamK (BN49_4981) is present in the core genome of *Klebsiella*. Like BamA, BamK has five POTRA domains attached to a β-barrel domain with distinguishing sequence characteristics that include alterations to three of the inter-strand loops, changes that would alter the domed surface exposed to the extracellular environment. Biochemical assays show that BamK is equally effective as BamA in its ability to catalyze the assembly of β-barrel proteins into the outer membrane. Under standard laboratory conditions, the gene encoding BamK is silenced, and we discuss the possible regulatory impact of a second form of the BAM complex in *Klebsiella* biology.

RESULTS

BamK can functionally replace BamA in K. pneumoniae

Analysis of the B5055 genome revealed that it encodes two proteins belonging to the BamA sub-family of Omp85 proteins (Heinz & Lithgow, 2014). BN49_4147 (annotated in the B5055 genome as *yaeT*) is designated here as *bamA*, given its genomic context is identical to that of the *E. coli bamA* locus (Fig. 1A). The second gene, BN49_4981, we designate as *bamK* (*Klebsiella*). Disruption of the *bamK* gene had no effect on growth in rich (LB) or minimal (M9) media, or in DMEM media which is often used as a proxy for "in host" growth conditions (Fig. S1A). A direct assessment of the virulence phenotype of B5055 in a mouse model of infection showed that replication of *K. pneumoniae*, with or without BamK, was equivalent in C57BL6 mice (Fig. S1B).

To examine relative expression levels of the *bamA* and *bamK* genes, qPCR assays were established. The expression level of *bamA* does not vary across the temperature range 30°C to 42°C, and *bamK* was not expressed to any measurable level under these conditions (Fig. 1B), or in DMEM (Fig. S1C). The promoter region upstream of *bamK* (Fig. S2A) is unusual in that the -10 element for RNA polymerase binding is (i) located at some distance relative to the predicted *bamK* start site (Fig. 1C), and (ii) is predicted to form a stem-loop structure. To address whether this region is a simple switch repressing expression of the gene, GFP reporter constructs were used (Fig. S2B). Deletion of the stem-loop region did not activate transcription from the *bamK* promoter region under the conditions of these experiments.

While this putative element did not function as a simple switch, the palindromic sequence and surrounding features inthe putative promoter is conserved upstream of the *bamK* locus in other *Klebsiella* type strains: MGH78578, NTUH-K2044 and 342 (Shu et al 2009, Ramos et al 2014). The protein sequences of BamK too are conserved in these diverse type strains (Fig. S3), inconsistent with an alternative scenario where the *bamK* sequence represents a pseudogene. We have been unable to determine what factor(s) would signal for expression of this cryptic gene.

To determine whether BamK can function in place of BamA, the B5055 strain Δ*bamA*::*bamK* was engineered to have the coding region for BamK replace the coding region for BamA, at the *bamA* locus of B5055 (Fig. S2C). Sequencing through the locus of the Δ*bamA*::*bamK* strain confirmed the absence of the coding region in the *bamA* gene, the integration of the *bamK* open-reading frame, and the removal of the kanamycin resistance cassette. To selectively detect BamK, the monoclonal antibody MAbK was raised to a unique sequence in the juncture of the POTRA3 and POTRA4 domains (Fig. S4). The Δ*bamA*::*bamK* replacement strain is viable and grows at the same rate as the parental B5055 strain (Fig. S5A). Outer membranes were prepared from B5055 and the Δ*bamA*::*bamK* strain, analyzed by SDS-PAGE and staining with Coomassie blue. The outer membrane protein profile for the Δ*bamA*::*bamK* strain was indistinguishable from that of the corresponding wild-type (Fig. 1D). Using blue-native polyacrylamide gel electrophoresis (BN-PAGE) and immunoblotting for analysis of the membranes, BamK co-migrates with the other subunits of the BAM complex: BamB, BamC, BamD and BamE (Fig. 1E, Fig. S5B). According to nomenclature adopted elsewhere (Bakelar et al 2016), this would correspond to the BAM_{KBCDE} complex.

Kp*BamA and* Kp*BamK gene replacement strains of* E. coli

To evaluate the specific activity of BamK with respect to BamA, we established a cellular model system to measure the rate of outer membrane protein assembly. The essential *bamA* gene in *E. coli* BL21 Star™ (DE3) was replaced by either BN49_4147 (strain Δ*bamA*::Kp*bamA*) or BN49_4981 (strain Δ*bamA*::Kp*bamK*). Both *E. coli* gene replacement strains were viable, grew at similar rates to the parental strain (Fig. S5C), and their outer membrane protein profiles were indistinguishable from each other (Fig. 2A). In these profiles, the most abundant species are the porins, trimeric β-barrel proteins that account for the semi-permeable nature of the outer membrane (Koebnik et al 2000). BN-PAGE of the two strains showed that the proteins *Kp*BamA and *Kp*BamK are each able to form a BAM

complex with the BamB, BamC, BamD and BamE subunits of *E. coli* (Fig. 2B). This is consistent with the high degree of sequence similarity between BamK and BamA in the periplasmic turns and POTRA domains (Fig. S3), the structural elements that make the critical contacts with the other Bam subunits (Bakelar et al 2016; Gu et al 2016; Han et al 2016). Deletion of either *bamB* or *bamC* shifted the mobility of BamK on the native gels, confirming that it is a component of a BAM_{KBCDE} complex (Fig. S5E).

The assembly of trimeric porins can be measured with a pulse-chase assay based on semi-native PAGE to separate the assembled trimers from the monomeric forms of unassembled protein (Heinz *et al.*, 2016, Stubenrauch *et al.*, 2016). To monitor the rate of assembly of the trimeric porin, translation of [^{35}S]-labeled PhoE precursor was induced in each strain and the assembly of the porin trimer was determined by semi-native PAGE and densitometry. In the semi-native gel electrophoresis the monomeric PhoE species is better focused than the oligomeric species, so the exponential decay of the monomer (as it assembles into an oligomer) was calculated (Stubenrauch et al 2016) for the time-course of PhoE assembly in *E. coli* Δ*bamA*::*KpbamA* (Fig. 2C) and Δ*bamA*::*KpbamK* (Fig. 2D). While we cannot rule out subtle differences in the earliest time-point measurements, there was no significant difference in the observed rate constants for the PhoE assembly reaction (Fig. 2E), confirming that BamK can efficiently catalyze outer membrane protein assembly in the absence of BamA.

The assembly of the LptDE complex (Fig. 3A) has been shown to be highly dependent on the function of the BAM complex (Lee *et al.*, 2016b), and an assay was established to measure the co-assembly of LptD and LptE (Fig. S6). Expression and translation of [^{35}S]-labeled LptD and [^{35}S]-labeled LptE was induced, and the assembly of the LptDE complex was determined by semi-native PAGE (Fig. 3B, Fig. 3C). Under these electrophoresis conditions, the LptDE dimer displays heat modifiability, whereby it migrates faster than would be expected based on size alone. This stability (heat modifiability) could be a result of the extensive interactions LptE makes within the LptD lumen (Ruiz *et al.*, 2010, Botos *et al.*, 2016). Confirming its identity as the LptDE complex, formation of this species depends on the presence of both LptD and LptE (Fig. S6). There was no significant difference in the rate of LptDE formation, or the observed rate constants for the LptDE assembly reaction, mediated by the BAM_{ABCDE} complex or the BAM_{KBCDE} complex (Fig. 3D, Fig. 3E).

***BamK is encoded in the core genome of* K. pneumoniae**

We used a global dataset of diverse isolates of *Klebsiella* spp. (Holt *et al.*, 2015) and our comparative sequence analysis demonstrated that both *bamA* and *bamK* are conserved in the *K. pneumoniae* species complex (comprised of *K. pneumoniae*, *K. quasipneumoniae* and *K. varicola*; Fig. 4). BamK was also found in several close relatives of *Klebsiella*, including species that are annotated as *Raoultella* spp. and *Enterobacter* spp. (Fig. S7, Table S1). Given the current revisions of *Klebsiella* taxonomy (Brisse *et al.*, 2014, Holt *et al.*, 2015, Li *et al.*, 2016), these species may be closely related to *Klebsiella* spp. or indeed become revised to be *Klebsiella* spp. Either way, these observations denote that BamK is encoded in the core genome of *K. pneumoniae*.

We also note the presence of a second group of BamA-paralogues (Fig. S7, orange branches): like BamK, these are present in addition to BamA in the respective organisms. However, these sequences do not form a monophyletic group with BamK. Thus, organisms such as *Haemophilus* spp. that are distantly related to *Klebsiella* spp. also have BamA-paralogues. Given that most Enterobactericae, such as *Escherichia* spp. and *Salmonella* spp. do not have second isoforms of BamA, the occurrence in *Pasturellaceae* must represent an independent evolution of a second BamA paralogue. We conclude that BamK is a specific trait for several related groups of the *Enterobacteriaceae*, such as *Raoultella* spp. and *Enterobacter* spp. and is an element of the core genome in *Klebsiella* spp.

Extracellular loops of BamA and BamK

The crystal structure of BamA from *E. coli* has been determined (Fig. 5A; Bakelar *et al.*, 2016, Gu *et al.*, 2016, Han *et al.*, 2016) and the sequence of BamA from *K. pneumoniae* maps readily onto these crystal coordinates. The extracellular loops of the β-barrel domain of BamA are important in contributing to the dome presented on the bacterial cell, and to help drive the dynamic changes in structure of BamA within the confines of the BAM complex (Bakelar *et al.*, 2016, Gu *et al.*, 2016, Han *et al.*, 2016). A major distinguishing feature between BamK and BamA concerns three of these loops: loop 4, loop 6 and loop 7. The sequences corresponding to these loops are highly conserved in BamK (Fig. S3): loops 4 and 7 are longer and loop 6 is substantially shorter in BamK, relative to BamA (Fig. 5B). This will impact on the surface features displayed on the outer face of the dome of BamK.

Differences in the lengths of these same three loops of BamA have been noted once before, in a previous study that described *E. coli* suppressors of bacteriophage infection (Smith *et al.*,

2007). Under pressure from Shiga toxin-encoding (Stx) phage infection, mutations were recovered in *bamA*, mutations that mimic the loop sequences in BamK (Fig. 5B, Fig. S3B). Structural modelling of this BamA(Φ) suggests substantive changes in the dome of the β-barrel domains of the two forms of BamA (Fig. 5C). While mere speculation, it may be that the evolution of BamK has been selected in response to phage (or other factors) in the environment. Despite its hypermucovisocous phenotype, B5055 is subject to killing by bacteriophage, and such phage have been mooted as alternative therapies for treating burn wound infections by *K. pneumoniae* (Kumari *et al.*, 2011).

DISCUSSION

Members of the Omp85 family of proteins have evolved to mediate the assembly of β-barrel proteins into bacterial outer membranes. Previous structural and phylogenetic analyses (Heinz & Lithgow, 2014) suggested a model for the evolution of TamA through a gene duplication of *bamA* followed by an evolutionary process of sub-functionalization. This process of gradual change in one copy of the gene provides the means to ultimately encode a protein (TamA) that performs its function in a significantly different way (Shen *et al.*, 2014, Selkrig *et al.*, 2015, Bamert *et al.*, 2017), to facilitate the process of outer membrane protein assembly in collaboration with BamA (Albenne & Ieva, 2017). In principle, *bamK* might represent an early event on an equivalent trajectory towards a further sub-functionalization. At the present time however, several features of BamK suggest that it can be considered an isoform of BamA. Structurally, the β-strands and inter-strand turns remain highly conserved (Fig. 5B, Fig. S3), as do most elements in the POTRA domains. This in turn suggested that BamK would interact with the same set of partner proteins as BamA, a suggestion that was validated in biochemical assays here with the engineered strains of *E. coli* (Δ*bamA*::*KpbamA* and Δ*bamA*::*KpbamK*) and in the *K. pneumoniae* Δ*bamA*::*bamK* strain: BamK forms a BAM complex with the partner proteins BamB, BamC, BamD and BamE.

While BamA is a crucial feature for membrane biogenesis, it is also a liability under some circumstances, with three classes of agents known to bind selectively to the extracellular dome of BamA and impact negatively on bacterial survival. Firstly, infection of *E. coli* with Stx bacteriophage relies on an initial, species-specific recognition of BamA involving sites in extracellular loops 4, 6 and 7 (Smith *et al.*, 2007). Secondly, the inhibitor of the CdiAB contact-inhibition system binds to BamA in a species-specific recognition that thereafter

prevents cell growth and division (Aoki *et al.*, 2008). Thirdly, a peptidomimetic compound designed to mimic natural defensin-like molecules binds BamA and inhibits growth of *E. coli* (Urfer *et al.*, 2016).

These scenarios provoke speculation on a potential neo-functionalization process that might be active in *Klebsiella* and other bacterial lineages. The inter-strand loops in BamK have evolved and, for loops 4, 6, and 7 both sequence diversity and length differences impact to change the dome surface of BamA presented at the bacterial cell surface (Fig. 5C). Whether this "new function" provides a means for a phage-resistant form of the BAM complex in *Klebsiella*, consistent with previous work on *E. coli* (Smith *et al.*, 2007), or for some other reason, the observation that the dome facing the extracellular *milieu* has been the most specific site of evolution in BamK is striking. In the *E. coli* scenario, the surface loops of BamA are recognized by Stx phage as a receptor, in interactions mediated through the conserved tail spike protein of the bacteriophage. Under selection with Stx phage, phage-resistant mutants were recovered (Smith *et al.*, 2007), with these *E. coli* mutants having surface loops showing similar sequence characteristics to the surface loops in BamK (Fig. 5B). Whether or not it relates to phage-resistance, according to the comparative sequence analyses presented herein, equivalent scenarios of sub-functionalization in *bamA* seems to have independently evolved in other bacterial lineages outside the *Enterobacteriaceae*. This speaks to a substantial selective pressure on diverse bacteria to evolve distinct isoforms of BamA. The processes of sub- and neo-functionalization of duplicate genes is recognized as a major driver of evolution (Andersson *et al.*, 2015).

The mechanisms behind host-pathogen interactions mediated by *K. pneumoniae* are largely driven by the outer surface features of the bacterium: its LPS, fimbrial adhesins and capsule (Krachler, 2016). Directly or indirectly, the BAM complex impacts on the display of these virulence factors, and the discovery of *bamK* suggests a potential for adaptability of the BAM complex in *Klebsiella*. LPS display on the cell surface depends on the activity of the LptDE complex, which is assembled by the BAM complex. The type I and type III fimbrial adhesins are extended through usher proteins (Klemm & Schembri, 2000, Wilksch *et al.*, 2011, Khater *et al.*, 2015), which are themselves dependent on the BAM complex and the TAM for efficient assembly (Stubenrauch *et al.*, 2016, Stubenrauch *et al.*, 2017, Heinz *et al.*, 2016). While the Wza secretion pore for capsular polysaccharide is not assembled by the BAM complex (Dunstan *et al.*, 2015), the capsule has a key attachment factor Wzi that is a β-barrel

protein (Bushell *et al.*, 2013). The extent to which the BAM complex thereby controls virulence in *Klebsiella* spp. has been underscored from host-pathogen assessments of a Δ*bamB* mutant of *K. pneumoniae* (Krachler, 2016, Hsieh *et al.*, 2016). A better understanding of how these aspects of biogenesis cooperate and are regulated is a first step towards targeting them to control bacterial replication in the course of blood stream and other infections by *K. pneumoniae*.

EXPERIMENTAL PROCEDURES

Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are described in Table S3. The parental strain for *K. pneumoniae* is B5055 (*K2*:O1 *serotype*), and for *E. coli* is BL21 Star™ (DE3) (Invitrogen). Unless otherwise stated, bacteria were routinely grown in Luria-Bertani (LB) medium overnight at 37°C with orbital shaking. When appropriate, media were supplemented with antibiotics at the following concentrations: ampicillin, 100 µg/mL; kanamycin, 50 µg/mL; chloramphenicol 34 µg/mL.

In order to examine bacterial growth in vitro, bacterial strains were grown overnight in LB broth (M9 minimal media, or DMEM) with orbital shaking at 200 rpm, diluted to a starting optical density at 600 nm (OD_{600}) of 0.05 into fresh LB broth (M9 minimal media, or DMEM), placed into 96-well plates (Falcon-BD), and shaken at 200 rpm for 24 hours. Values for the OD_{600} of each sample were recorded at regular intervals using the Tecan Spark™ 10M, and all strains were assayed in three independent biological replicates in technical triplicates.

Prior to isolating membrane fractions, B5055 or BL21 Star™ (DE3) were grown to mid-log phase at 37°C with orbital shaking at 200 rpm. Cells were harvested by centrifugation (5000 × *g* for *E. coli*, 15000 × *g* for B5055 strains, 10 min, 4 °C) and membranes isolated as previously described (Clements *et al.*, 2009). Total membranes were stored in aliquots at -80 °C until required.

Construction of deletion and knock-in strains

PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) or Taq DNA Polymerase (Roche, Switzerland). Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. Synthetic

oligonucleotides for PCR and sequencing (Table S4) were obtained from IDT (Singapore). Knock-out and knock-in mutants, in which target genes were deleted or replaced by allelic exchange with a kanamycin resistance-encoding gene, were constructed in *K. pneumoniae* B5055 using the "gene doctoring" technique (Lee *et al.*, 2009). All primers used are listed in Table S4. 'Donor' plasmids carrying the desired mutation were constructed as follows. The kanaymcin gene was amplified from pKD4 using primers kanF and kanR. The resulting product included flanking fragment length polymorphism (FLP) recombinase target (FRT) sites to permit subsequent kanamycin cassette excision through use of a helper plasmid. Using either *K. pneumoniae* B5055 genomic DNA as the template, or *E. coli* BL21 Star™ (DE3) genomic DNA as the template, approximately 0.5 kb regions flanking the upstream and downstream sequence of the target gene were PCR amplified. The DNA fragments were joined using the Gibson assembly® Master Mix (New England Biolabs, Ipswich, MA), where the kanamycin cassette was flanked by the upstream and downstream target gene sequences (for knock ins, *bamA* or *bamK* open-reading frames followed by the kanamycin cassette) into a PCR amplified ISce-I-flanked pGEM-T Easy (Promega) backbone to yield donor plasmids, which were then sequenced.

Plasmid pACBSR carries genes encoding I-SceI endonuclease and lambda Red recombinase under inducible control by L-arabinose. Donor plasmids and pACBSR were transformed into electrocompetent B5055 cells (0.1 cm gap-width cuvette; 200 ohms, 25 µF, 1.8 kV) and selected on LB agar containing kanamycin and chloramphenicol. A single co-transformant was inoculated into 1 mL LB containing 0.2% L-arabinose (Sigma-Aldrich, St. Louis, MO) and chloramphenicol and grown in a shaking incubator at 37°C for 16 h. Cell dilutions were grown on LB agar containing kanamycin, and resultant colonies were screened by colony PCR using primers flanking the targeted region and within the kanamycin resistance gene. The loss of pACBSR was induced by 0.2% L-arabinose without selection. When required, the kanamycin resistance gene was excised via the FRT sites using the FLP helper plasmid pCP20. All mutations were confirmed by PCR analysis, using primers flanking the targeted regions.

Gene expression analysis

To determine the effect of temperature on *bamA* or *bamK* expression, RNA was extracted from log phase *K. pneumoniae* B5055 wildtype cultures exposed to pre-heated LB media (30 °C, 37 °C or 42 °C) for 1 hour using the RNeasy minikit (Qiagen). Synthesis of cDNA used

70 ng of random hexamers (Invitrogen) and Superscript II Reverse Transcription kit (Invitrogen), all according to the manufacturer's guidelines. qPCR was performed using SYBR Green Master (Roche) with primer pairs used at 0.125 µM each; Table S4. Cycling conditions using the LightCycler480 instrument (Roche) were as follows: 95°C for 10 min followed by 35 cycles of 95 °C for 30 sec, 55 °C for 60 sec and 72 °C for 15 sec. For determining relative transcript levels for *bamA and bamK*, values were normalised to the expression of housekeeping gene *rpoD*. Primer specificity was determined by melting curve analysis using open source qPCR data analyser LinReg (Ruijter *et al.*, 2009). All qPCR sample reactions were performed in triplicate.

Construction of gfp reporter fusions and GFP fluorescence measurements

DNA fragments 500bp directly upstream of the translational start site of *bamA* or *bamK* were amplified from *K. pneumoniae* B5055 genomic DNA and were cloned into reporter plasmid pPROBE′-gfp[tagless] digested with BamHI and EcoRI. The vector is a broad-host range reporter plasmid with a "promoterless" open-reading frame encoding the green fluorescent protein (GFP; Miller *et al.*, 2000). For the *bamK* promoter sequence with a modified stem-loop region (non-palindrome), nucleotide gene blocks (gBlocks, from IDT) were synthesized and cloned into pPROBE′-gfp[tagless] as above. Resultant gfp reporter fusion plasmids were transformed into *K. pneumoniae* B505 via electroporation.

Overnight cultures of the reporter strains were diluted to a starting OD_{600} of 0.05 into fresh LB broth and incubated for 3 h at 37 °C on an orbital incubator shaker (180 rpm). From the mid-logarithmic reporter strain cultures, 1 mL aliquots were then harvested (15000 × g for 10 min) and washed three with PBS. Washed cells were then resuspended in PBS back to 1 mL and GFP fluorescence emission was measured using filter set at 488 nm excitation and 512 nm using Spark™ 10M multimode plate reader (Tecan). All GFP measurements were performed in triplicate and were normalized to its cell density (OD_{600}).

Protein structure prediction

For the graphic in Fig. S4, the *K. pneumoniae* BamK and *E. coli* BamA structural models were built from amino acid sequence data using Protein Homology/analogY Recognition Engine (PHYRE) (Kelley & Sternberg, 2009, Kelley *et al.*, 2015). The models were further evaluated using I-TASSER (Iterative Threading Assembly Refinement) (Zhang, 2008, Roy *et al.*, 2010) to provide validation to the predicted structures. Images, including

superimpositions, of the models and crystal structures were generated using Pymol (Version 1.8). Using the equivalent modelling approach, a structural comparison of the *E. coli* proteins BamA and BamA(Φ) was undertaken. The sequence of BamA(Φ) was reconstructed from published protein sequence data (Smith *et al.*, 2007).

Pulse-Chase Assay

Radioactive labelling of proteins and pulse-chase assays were performed as previously described (Stubenrauch *et al.*, 2016), with some modifications. Briefly, overnight cultures of *E. coli* strains harboring pKS02 (*fimD*-expressing plasmid) were diluted 1:50 with fresh LB and incubated to an OD_{600} of 0.60 at 37 °C with shaking. Cells were subjected to centrifugation (4600 × *g*, 10 min, 4 °C), washed and resuspended in M9-S media. Following 30-minute incubation at 37 °C with shaking, cells were diluted with an equal volume of 40 % vol/vol glycerol, aliquoted and snap frozen in liquid nitrogen. Aliquots from the same batch were considered to be technical replicates, and aliquots from different, independent batches were considered to be biological replicates.

When required, aliquots were thawed on ice, subjected to centrifugation (3000 × *g*, 5 min, 4 °C) and resuspended in M9-S media. Following addition of 0.5 mg/mL rifampicin, cells were incubated for 1 hour at 37 °C, 400 rpm (3 mm orbit), 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) was added and cells were incubated for a further 5 min. Samples were 'pulsed' by addition of 22 μCi mL^{-1} [^{35}S]-methionine and [^{35}S]-cysteine (NEG072, Perkin Elmer) for 60 s without shaking, subjected to centrifugation (3000 × *g*, 5 min, 4 °C), and then 'chased' by resuspension in M9+S media for up to 32 min (the chase temperature was 37 °C (static) for both PhoE and LptDE assembly assays).

To track PhoE or LptDE assembly, aliquots were removed to semi-native (SN) sample buffer (Stubenrauch *et al.*, 2016) at 10 seconds, 2, 4, 8, 16 and 32 min. Samples were incubated for 10 minutes at 37 °C or at 100 °C as a control to denature any assembled oligomers. Samples were analyzed by SN-PAGE (see Stubenrauch *et al.*, 2016) using 4–16% gradient gels. Proteins were transferred to 0.45 μm nitrocellulose membranes using standard techniques, membranes were air-dried and radiation was captured using a storage phosphor screen (GE Health Sciences) and detected using Typhoon Trio (at 320 nm).

From at least three biological replicate gels (n=3), PhoE monomeric or LptDE oligomeric band densities were measured using ImageQuant™ TL software version 7.0 (GE Health Sciences) with the local average background correction tool. To be able to compare biological replicates, 'normalized density' values were calculated by dividing the density measurements of the six timepoints by the greatest density measurement among those timepoints. To calculate the observed rate constant, the 'normalized density' values were subsequently determined using the one-phase decay nonlinear regression tool from GraphPad Prism software version 7.02 (GraphPad software, Inc.).

Animal infections

Mice (6-week-old, male BALB/c) were infected by intranasal delivery with 8-9 × 10^3 CFU of B5055 wildtype or the isogenic Δ*bamK* mutant strain. After 4, 24, 72 and 96 hours post-infection, mice were euthanized and bacterial counts from lungs and liver determined. All animal experiments were approved by The University of Melbourne Animal Ethics Committee and were conducted in accordance with the Prevention of Cruelty to Animals Act (1986) and the Australian National Health and Medical Research Council Code of Practice for the Care and Use of Animals for Scientific Purposes (1997).

Sequence analysis

The dataset for *Klebsiella pneumoniae* (Table S1) was assembled and annotated as described previously (Holt *et al.*, 2015), and the pan- and core-genome clusters generated using the software package roary (Page *et al.*, 2015) with a cutoff of id90, and the resulting core gene alignment was used for tree calculation (Fig. S7). All genes containing a PF01103 domain (the canonical domain for the Omp85 beta-barrel) according to the annotation with prokka (Seemann, 2014) were retrieved; and all genes in clusters (retrieved from the roary output) containing a PF01103 gene were considered putative Omp85-enconding genes. Representative sequences of the clusters were analysed manually to exclude the other two *Klebsiella* Omp85 genes (Heinz & Lithgow, 2014); only genes with domain architectures similar to BamA were used to calculate a tree to separate BamA from other copies (BamK). For the tree to distinguish BamA from BamK sequences (Fig. S8), the encoded amino acid sequences were aligned using muscle (version 3.8.31; Edgar, 2004), and informative sites were chosen using gblocks (Castresana, 2000) as implemented in seaview (version 4; (Gouy

et al., 2010)) allowing all more lenient settings. The tree was calculated using RAxML (version 7.8.6; (Stamatakis, 2006)) with 100 bootstraps and the LG matrix.

The core gene tree was constructed based on the roary output of the core gene alignment, by selecting only sites encoding single nucleotide polymorphisms (snps) with SNP-sites (http://mgen.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000056), the tree was calculated using RaxML (version 7.8.6; (Stamatakis, 2006) with 100 bootstraps and the GTR gamma model. The presence/absence of BamA/K as inferred from a BamA/K tree as described above (Fig. S8), as well as the phylogroups as described in (Holt *et al.*, 2015), were plotted onto the tree (Fig. S7).

For the dataset describing the BamK distribution across other species (Table S2, Fig. 4), both BamA and BamK from *K. pneumoniae* MGH78578 (UniProt ID A6T4X9 and A6TGX3, respectively) were used as input into blast with UniProt (UniProt, 2015) searching only microbial genomes. The first 100 hits for both these search runs were used, and the sequences and species joined as a dataset as follows: for each sequence, assigned to a taxonomically unique strain, in the BamK as well as the BamA dataset, where the respective strain was not also represented in the other dataset as well after the first blast search (i.e. a BamK sequence derived from an organism where the corresponding BamA was not in the dataset already, or vice versa), a manual search was performed, and the respective BamA or BamK (if present) was added to the dataset, to give a full set of BamA and BamK sequences for the organisms investigated (Table S2). In the case of highly fragmented sequences, both the BamA and the BamK sequence were removed from the dataset for the respective organism. The sequences were aligned using muscle (Edgar, 2004) as implemented in seaview (Gouy *et al.*, 2010), and to extract the barrel region, the BamA sequence of *K. pneumoniae* MGH78578 (UniProt ID A6T4X9) was used as input for the structure prediction server phyre2 (Kelley *et al.*, 2015), and the barrel domain extracted according to the structure (Noinaj *et al.*, 2013); informative sites were subsequently extracted using gblocks as described above. The tree (Fig. 4) based on the amino acid sequences was calculated with RaxML using the LG matrix and 100 bootstrap replicates. All alignment and tree files are available on figshare [https://figshare.com/s/a37171b4fa4a4d27b7de; link will be set public upon acceptance].

ACKNOWLEDGEMENTS

We thank Dominika Elmlund for expert advice on the transcription activation studies. We acknowledge the staff of the Monash Antibody Technology Facility for monoclonal antibodies. Research was supported by Program Grant 1092262 from the National Health and Medical Research Council of Australia (NHMRC). T.L. is an Australian Research Council Laureate Fellow (FL130100038).

AUTHOR CONTRIBUTIONS

VT, EH, RAD, CJS, FHA, AC, JJW, HC, CTW designed and carried out analysis. EH, CJS, JJW, CTW, JY, IDH, TL provided expertise to analyses. EH, IDH, TL supervised experimental work and evaluated data. VT, GD, RAS, IDH, TL wrote the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

PLAIN LANGUAGE SUMMARY

BamA is a protein in the outer membrane that controls the essential process for the assembly of other outer membrane proteins. We show that the important human pathogen *Klebsiella pneumoniae* has two types of BamA, an evolutionary insight into the natural pressure placed on this essential process in bacterial cell biology.

REFERENCES

Albenne, C. & R. Ieva, (2017) Job contenders: roles of the beta-barrel assembly machinery and the translocation and assembly module in autotransporter secretion. Mol Microbiol 106: 505-517.

Andersson, D.I., J. Jerlstrom-Hultqvist & J. Nasvall, (2015) Evolution of new functions de novo and from preexisting genes. Cold Spring Harb Perspect Biol 7.

Aoki, S.K., J.C. Malinverni, K. Jacoby, B. Thomas, R. Pamma, B.N. Trinh, S. Remers, J. Webb, B.A. Braaten, T.J. Silhavy & D.A. Low, (2008) Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. Mol Microbiol 70: 323-340.

Bagley, S.T., (2015) Habitat Association of *Klebsiella* Species. Infection Control 6: 52-58.

Bakelar, J., S.K. Buchanan & N. Noinaj, (2016) The structure of the beta-barrel assembly machinery complex. Science 351: 180-186.

Bamert, R.S., K. Lundquist, H. Hwang, C.T. Webb, T. Shiota, C.J. Stubenrauch, M.J. Belousoff, R.J.A. Goode, R.B. Schittenhelm, R. Zimmerman, M. Jung, J.C. Gumbart & T. Lithgow, (2017) Structural basis for substrate selection by the translocation and assembly module of the beta-barrel assembly machinery. Mol Microbiol 106: 142-156.

Bi, W., H. Liu, R.A. Dunstan, B. Li, V.V.L. Torres, J. Cao, L. Chen, J.J. Wilksch, R.A. Strugnell, T. Lithgow & T. Zhou, (2017) Extensively Drug-Resistant *Klebsiella pneumoniae* Causing Nosocomial Bloodstream Infections in China: Molecular Investigation of Antibiotic Resistance Determinants, Informing Therapy, and Clinical Outcomes. Front Microbiol 8: 1230.

Botos, I., N. Majdalani, S.J. Mayclin, J.G. McCarthy, K. Lundquist, D. Wojtowicz, T.J. Barnard, J.C. Gumbart & S.K. Buchanan, (2016) Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. Structure 24: 965-976.

Brisse, S., V. Passet & P.A. Grimont, (2014) Description of *Klebsiella quasipneumoniae* sp. nov., isolated from human infections, with two subspecies, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella variicola*. Int J Syst Evol Microbiol 64: 3146-3152.

Bushell, S.R., I.L. Mainprize, M.A. Wear, H. Lou, C. Whitfield & J.H. Naismith, (2013) Wzi is an outer membrane lectin that underpins group 1 capsule assembly in *Escherichia coli*. Structure 21: 844-853.

Calfee, D.P., (2017) Recent advances in the understanding and management of *Klebsiella pneumoniae*. F1000Res 6: 1760.

Campos, A.C., J. Albiero, A.B. Ecker, C.M. Kuroda, L.E. Meirelles, A. Polato, M.C. Tognim, M.A. Wingeter & J.J. Teixeira, (2016) Outbreak of *Klebsiella pneumoniae*

carbapenemase-producing *K pneumoniae*: A systematic review. Am J Infect Control 44: 1374-1380.

Carpenter, J.L., (1990) *Klebsiella* Pulmonary Infections: Occurrence at One Medical Center and Review. Reviews of Infectious Diseases 12: 672-682.

Chen, L., R. Todd, J. Kiehlbauch, M. Walters & A. Kallen, (2017) Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing *Klebsiella pneumoniae* - Washoe County, Nevada, 2016. MMWR Morb Mortal Wkly Rep 66: 33.

Clements, A., D. Bursac, X. Gatsos, A.J. Perry, S. Civciristov, N. Celik, V.A. Likic, S. Poggio, C. Jacobs-Wagner, R.A. Strugnell & T. Lithgow, (2009) The reducible complexity of a mitochondrial molecular machine. Proc Natl Acad Sci U S A 106: 15791-15795.

Dunstan, R.A., I.D. Hay, J.J. Wilksch, R.B. Schittenhelm, A.W. Purcell, J. Clark, A. Costin, G. Ramm, R.A. Strugnell & T. Lithgow, (2015) Assembly of the secretion pores GspD, Wza and CsgG into bacterial outer membranes does not require the Omp85 proteins BamA or TamA. Mol Microbiol 97: 616-629.

Edgar, R.C., (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

Ejaz, H., N. Wang, J.J. Wilksch, A.J. Page, H. Cao, S. Gujaran, J.A. Keane, T. Lithgow, I. Ul-Haq, G. Dougan, R.A. Strugnell & E. Heinz, (2017) Phylogenetic Analysis of *Klebsiella pneumoniae* from Hospitalized Children, Pakistan. Emerg Infect Dis 23: 1872-1875.

Gouy, M., S. Guindon & O. Gascuel, (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221-224.

Gu, Y., H. Li, H. Dong, Y. Zeng, Z. Zhang, N.G. Paterson, P.J. Stansfeld, Z. Wang, Y. Zhang, W. Wang & C. Dong, (2016) Structural basis of outer membrane protein insertion by the BAM complex. Nature 531: 64-69.

Hall-Stoodley, L., J.W. Costerton & P. Stoodley, (2004) Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2: 95-108.

Han, L., J. Zheng, Y. Wang, X. Yang, Y. Liu, C. Sun, B. Cao, H. Zhou, D. Ni, J. Lou, Y. Zhao & Y. Huang, (2016) Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. Nat Struct Mol Biol 23: 192-196.

Heinz, E. & T. Lithgow, (2014) A comprehensive analysis of the Omp85/TpsB protein superfamily structural diversity, taxonomic occurrence, and evolution. Front Microbiol 5: 370.

Heinz, E., C.J. Stubenrauch, R. Grinter, N.P. Croft, A.W. Purcell, R.A. Strugnell, G. Dougan & T. Lithgow, (2016) Conserved Features in the Structure, Mechanism, and Biogenesis of the Inverse Autotransporter Protein Family. Genome Biol Evol 8: 1690-1705.

Holt, K.E., H. Wertheim, R.N. Zadoks, S. Baker, C.A. Whitehouse, D. Dance, A. Jenney, T.R. Connor, L.Y. Hsu, J. Severin, S. Brisse, H. Cao, J. Wilksch, C. Gorrie, M.B. Schultz, D.J. Edwards, K.V. Nguyen, T.V. Nguyen, T.T. Dao, M. Mensink, V.L. Minh, N.T. Nhu, C. Schultsz, K. Kuntaman, P.N. Newton, C.E. Moore, R.A. Strugnell & N.R. Thomson, (2015) Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci U S A 112: E3574-3581.

Hsieh, P.F., C.R. Hsu, C.T. Chen, T.L. Lin & J.T. Wang, (2016) The *Klebsiella pneumoniae* YfgL (BamB) lipoprotein contributes to outer membrane protein biogenesis, type-1 fimbriae expression, anti-phagocytosis, and in vivo virulence. Virulence 7: 587-601.

Kelley, L.A., S. Mezulis, C.M. Yates, M.N. Wass & M.J. Sternberg, (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845-858.

Kelley, L.A. & M.J. Sternberg, (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363-371.

Khater, F., D. Balestrino, N. Charbonnel, J.F. Dufayard, S. Brisse & C. Forestier, (2015) In silico analysis of usher encoding genes in *Klebsiella pneumoniae* and characterization of their role in adhesion and colonization. PLoS ONE 10: e0116215.

Klemm, P. & M.A. Schembri, (2000) Bacterial adhesins: function and structure. International Journal of Medical Microbiology 290: 27-35.Koebnik, R., K.P. Locher & P. Van Gelder (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol. 37: 239-53.

Krachler, A.M., (2016) BamB and outer membrane biogenesis - The Achilles' heel for targeting *Klebsiella* infections? Virulence 7: 508-511.

Kumar, V. & S. Chhibber, (2011) Acute lung inflammation in *Klebsiella pneumoniae* B5055-induced pneumonia and sepsis in BALB/c mice: a comparative study. *Inflammation* 34: 452-462.

Kumari, S., K. Harjai & S. Chhibber, (2011) Bacteriophage versus antimicrobial agents for the treatment of murine burn wound infection caused by *Klebsiella pneumoniae* B5055. J Med Microbiol 60: 205-210.

Lee, C.R., J.H. Lee, K.S. Park, J.H. Jeon, Y.B. Kim, C.J. Cha, B.C. Jeong & S.H. Lee, (2017) Antimicrobial Resistance of Hypervirulent *Klebsiella pneumoniae*: Epidemiology, Hypervirulence-Associated Determinants, and Resistance Mechanisms. Front Cell Infect Microbiol 7: 483.

Lee, C.R., J.H. Lee, K.S. Park, Y.B. Kim, B.C. Jeong & S.H. Lee, (2016a) Global Dissemination of Carbapenemase-Producing *Klebsiella pneumoniae*: Epidemiology, Genetic Context, Treatment Options, and Detection Methods. Front Microbiol 7: 895.

Lee, D.J., L.E. Bingle, K. Heurlier, M.J. Pallen, C.W. Penn, S.J. Busby & J.L. Hobman, (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. BMC Microbiol 9: 252.

Lee, J., M. Xue, J.S. Wzorek, T. Wu, M. Grabowicz, L.S. Gronenberg, H.A. Sutterlin, R.M. Davis, N. Ruiz, T.J. Silhavy & D.E. Kahne, (2016b) Characterization of a stalled complex on the beta-barrel assembly machine. Proc Natl Acad Sci U S A 113: 8717-8722.

Li, C.Y., Y.L. Zhou, J. Ji & C.T. Gu, (2016) *Klebsiella alba* is a later heterotypic synonym of *Klebsiella quasipneumoniae* subsp. *similipneumoniae*. Int J Syst Evol Microbiol 66: 2406-2408.

Miller, W.G., Leveau, J.H.J., and Lindow, S.E. (2000) Improved *gfp* and *inaZ* broad-host- range promoter- pr obe vectors. *Mol Plant Microbe Interact* 13: 1243–1250.

Navon-Venezia, S., K. Kondratyeva & A. Carattoli, (2017) *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. FEMS Microbiol Rev 41: 252-275.

Noinaj, N., J.C. Gumbart & S.K. Buchanan, (2017) The beta-barrel assembly machinery in motion. Nat Rev Microbiol 15: 197-204.

Noinaj, N., A.J. Kuszak, J.C. Gumbart, P. Lukacik, H. Chang, N.C. Easley, T. Lithgow & S.K. Buchanan, (2013) Structural insight into the biogenesis of beta-barrel membrane proteins. Nature 501: 385-390.

Paczosa, M.K. & J. Mecsas, (2016) *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. Microbiol Mol Biol Rev 80: 629-661.

Page, A.J., C.A. Cummins, M. Hunt, V.K. Wong, S. Reuter, M.T. Holden, M. Fookes, D. Falush, J.A. Keane & J. Parkhill, (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31: 3691-3693.

Plummer, A.M. & K.G. Fleming, (2016) From Chaperones to the Membrane with a BAM! Trends Biochem Sci 41: 872-882.

Ramirez, M.S., G.M. Traglia, D.L. Lin, T. Tran & M.E. Tolmasky, (2014) Plasmid-Mediated Antibiotic Resistance and Virulence in Gram-negatives: the *Klebsiella pneumoniae* Paradigm. Microbiology spectrum 2: 1-15.

Roy, A., A. Kucukural & Y. Zhang, (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725-738.

Ruijter, J.M., C. Ramakers, W.M. Hoogaars, Y. Karlen, O. Bakker, M.J. van den Hoff & A.F. Moorman, (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 37: e45.

Ruiz, N., S.S. Chng, A. Hiniker, D. Kahne & T.J. Silhavy, (2010) Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. Proc Natl Acad Sci U S A 107: 12245-12250.

Seemann, T., (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068-2069.

Selkrig, J., M.J. Belousoff, S.J. Headey, E. Heinz, T. Shiota, H.H. Shen, S.A. Beckham, R.S. Bamert, M.D. Phan, M.A. Schembri, M.C. Wilce, M.J. Scanlon, R.A. Strugnell & T. Lithgow, (2015) Conserved features in TamA enable interaction with TamB to drive the activity of the translocation and assembly module. Sci Rep 5: 12905.

Shen, H.H., D.L. Leyton, T. Shiota, M.J. Belousoff, N. Noinaj, J. Lu, S.A. Holt, K. Tan, J. Selkrig, C.T. Webb, S.K. Buchanan, L.L. Martin & T. Lithgow, (2014) Reconstitution of a nanomachine driving the assembly of proteins into bacterial outer membranes. Nat Commun 5: 5078.

Smith, D.L., C.E. James, M.J. Sergeant, Y. Yaxian, J.R. Saunders, A.J. McCarthy & H.E. Allison, (2007) Short-tailed stx phages exploit the conserved YaeT protein to disseminate Shiga toxin genes among enterobacteria. J Bacteriol 189: 7223-7233.

Stamatakis, A., (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

Stickler, D.J., (2008) Bacterial biofilms in patients with indwelling urinary catheters. Nat Clin Pract Urol 5: 598-608.

Stubenrauch, C., M.J. Belousoff, I.D. Hay, H.H. Shen, J. Lillington, K.L. Tuck, K.M. Peters, M.D. Phan, A.W. Lo, M.A. Schembri, R.A. Strugnell, G. Waksman & T. Lithgow, (2016) Effective assembly of fimbriae in *Escherichia coli* depends on the translocation assembly module nanomachine. Nat Microbiol 1: 16064.

Stubenrauch, C.J., G. Dougan, T. Lithgow & E. Heinz, (2017) Constraints on lateral gene transfer in promoting fimbrial usher protein diversity and function. Open Biol 7.

Tan, J.W., J.J. Wilksch, D.M. Hocking, N. Wang, Y.N. Srikhanta, M. Tauschek, T. Lithgow, R.M. Robins-Browne, J. Yang & R.A. Strugnell, (2015) Positive autoregulation of mrkHI by the cyclic di-GMP-dependent MrkH protein in the biofilm regulatory circuit of *Klebsiella pneumoniae*. J Bacteriol 197: 1659-1667.

UniProt, C., (2015) UniProt: a hub for protein information. Nucleic Acids Res 43: D204-212.

Urfer, M., J. Bogdanovic, F. Lo Monte, K. Moehle, K. Zerbe, U. Omasits, C.H. Ahrens, G. Pessi, L. Eberl & J.A. Robinson, (2016) A Peptidomimetic Antibiotic Targets Outer Membrane Proteins and Disrupts Selectively the Outer Membrane in *Escherichia coli*. J Biol Chem 291: 1921-1932.

Vlahovicek, K., (2003) DNA analysis servers: plot.it, bend.it, model.it and IS. Nucleic Acids Research 31: 3686-3687.

Webb, C.T., E. Heinz & T. Lithgow, (2012a) Evolution of the beta-barrel assembly machinery. Trends Microbiol 20: 612-620.

Webb C.T., Selkrig J., Perry A.J., Noinaj N., Buchanan S.K. & Lithgow T. (2012b) Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. J Mol Biol. 422:545-555

Wilksch, J.J., J. Yang, A. Clements, J.L. Gabbe, K.R. Short, H. Cao, R. Cavaliere, C.E. James, C.B. Whitchurch, M.A. Schembri, M.L. Chuah, Z.X. Liang, O.L. Wijburg, A.W. Jenney, T. Lithgow & R.A. Strugnell, (2011) MrkH, a novel c-di-GMP-dependent transcriptional activator, controls *Klebsiella pneumoniae* biofilm formation by regulating type 3 fimbriae expression. PLoS Pathog 7: e1002204.

Wyres, K.L. & K.E. Holt, (2016) *Klebsiella pneumoniae* Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol 24: 944-956.

Zhang, Y., (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

FIGURE LEGENDS

Figure 1. The gene *bamK* encodes a novel Omp85 family protein in *Klebsiella pneumoniae* B5055. (A) Gene synteny in the *bamA* locus of the *Klebsiella* strain B5055 and *E. coli* are compared and found equivalent for BN49_4147. In a distinct gene context, BN49_4981 is herein designated *bamK*. The coloring denotes conserved genes even where the annotations may differ between species e.g. the *K. pneumoniae* gene *yaeT* is homologous to the gene *bamA* in *E. coli*. In comparison, genomic context display of gene sequence data for *K. pneumoniae* subsp. *pneumoniae* B5055 around the *bamK* locus. (B) Transcript levels for *bamA* and *bamK* were determined through qPCR at three growth temperatures: 30°C, 37°C or 42°C. Data is expressed relative to control transcripts from the constituently-expressed *rpoD*. (C) *In silico* analysis of intrinsic curvature of the 5' regulatory region of *bamK* was predicted with bend.it® (Vlahovicek, 2003) and regions with >5 degrees per helical turn of DNA (in red) correspond to highly curved regions of DNA. The 200bp region of the genome immediately upstream from the start of the *bamK* open-reading frame is summarized as an inset (full sequence details in Fig. S2). (D) Outer membranes (100 μg protein) prepared from B5055 (wild-type) and the Δ*bamA*::*bamK* mutant were subjected to SDS-PAGE, and Coomassie blue staining of the gel. The migration positions for the molecular weight standards are indicated. (E) Outer membranes (200 μg protein) were prepared from B5055 (wild-type) and the Δ*bamA*::*bamK* mutant and analysed by BN-PAGE and immunoblotting with the indicated antiserum. Diagnostically, the BAM complex containing BamK (BAM_{KBCDE}) migrates marginally slower than the BAM_{ABCDE} complex. The full BN-PAGE data is provided in Fig. S5B.

Figure 2. Gene replacement mutants of *E. coli* expressing *Kp*BamA or *Kp*BamK. (A) Outer membranes (100 μg protein) isolated from *E. coli* strains Δ*bamA*::*KpbamA* or Δ*bamA*::*KpbamK* were subject to SDS-PAGE and Coomassie blue staining. Migration of the molecular weight standards is indicated. (B) Outer membranes (200 μg protein) were prepared from *E. coli* strain Δ*bamA*::*KpbamA* and *E. coli* strain Δ*bamA*::*KpbamK* and analysed by BN-PAGE and immunoblotting with the indicated antiserum. Diagnostically, the BAM complex containing BamK (BAM_{KBCDE}) migrates marginally slower than the BAM_{ABCDE} complex. The full BN-PAGE data is provided in Fig. S5D. (C) The rate of assembly of [^{35}S]-PhoE into mature oligomers was assessed in the *E. coli* strain Δ*bamA*::*KpbamA*. Cells harboring pKS07 (a plasmid encoding PhoE) were starved of methionine and cysteine, pulse-labeled with [^{35}S]-methionine/[^{35}S]-cysteine for 40 seconds,

followed by a 60-minute chase with unlabeled methionine and cysteine. Aliquots were taken at 10 sec, 2, 4, 8, 16 and 32 minutes. Samples were incubated at either 37 °C (time-course) or 100 °C (control sample at 32 min time-point) before analysis by semi-native SDS-PAGE, storage phosphor-imaging and densitometry. The position of the 39 kDa monomer and slower migrating oligomer are indicated. (D) The rate of assembly of [^{35}S]-PhoE into mature oligomers was assessed in the *E. coli* strain Δ*bamA*::*KpbamK*, as described above. (E) Histogram of the observed rate constants for reduction of the PhoE monomer. Error bars correspond to SEM of fit (n=3). The statistical significance was determined by an unpaired two-tailed Student's t test (ns, not significant).

Figure 3. *Kp*BamA and *Kp*BamK mediates the assembly of the LptD/E complex in *E. coli*. (A) Structure of LptDE complex (pdb:4RHB). In the correctly assembled LptDE, the lipoprotein subunit LptE (red) sits as a core stabilizing the β-barrel protein LptD (blue) within the plane of the outer membrane. (B) The rate of assembly of [^{35}S]-LptD and [^{35}S]-LptE into the LptDE complex was assessed in the *E. coli* strain Δ*bamA*::*KpbamA*. Cells harboring pCJS44 (a pET-Duet plasmid encoding both LptD and LptE) were starved of methionine and cysteine, pulse-labeled with [^{35}S]-methionine/[^{35}S]-cysteine for 40 seconds, followed by a 60-minute chase with unlabeled methionine and cysteine. Aliquots were taken at 10 sec, 2, 4, 8, 16 and 32 minutes. Samples were incubated at either 37 °C (time-course) or 100 °C (control sample at 32 min time-point) before analysis by semi-native SDS-PAGE, storage phosphor-imaging and densitometry. The position of the 20 kDa LptE monomer, 95 kDa LptD monomer and the dimeric LptDE complex are indicated. The formation of the LptDE complex is dependent on the co-expression of both LptD and LptE (Fig. S6). (C) The rate of assembly of [^{35}S]-LptD and [^{35}S]-LptE into the LptDE complex was assessed in the *E. coli* strain Δ*bamA*::*KpbamK*, as described above. (D) LptDE dimer assembly was calculated by plotting the normalised density of each dimeric band over time fitted to a curve following an exponential growth equation. The accumulation of the dimeric band over time was assumed to be in proportion to the oligomerization of LptD and LptE monomers into a functional dimer. (E) Histogram of the observed rate constants for appearance of the LptDE dimer. Error bars correspond to SEM of fit (n=3). The statistical significance was determined by an unpaired two-tailed Student's t test (ns, not significant).

Figure 4. BamK is encoded in the core genome of *Klebsiella pneumoniae*. The tree is based on the core gene alignment of the global *Klebsiella* collection (Holt et al 2015), the

outer ring indicates the species and subspecies of *Klebsiella*. The same species-specific coloring is used for the branches of the tree. The ubiquitous presence of *bamA* genes is indicated in the red (inner) circle, and *bamK* genes are shown in blue (middle circle). BamA and BamK are found throughout the entire *Klebsiella pneumoniae* complex diversity; incomplete open reading frames interrupted through either frame shifts or contig breaks/gaps in the alignment are indicated as shown in the legend; however, no pattern of continuous loss can be observed, and the incomplete sequences would more likely be derived from assembly errors.

Figure 5. Extracellular loops distinguish BamA and BamK on the surface of *K. pneumoniae*. (A) Topology of the BAM complex, highlighting the structure of *E. coli* BamA (pdb: 5ekq) in the context of the partner lipoprotein subunits of the BAM complex and showing the disposition of the inter-strand loops 4, 6 and 7. OM – outer membrane, IM – inner (cytoplasmic) membrane. (B) Multiple sequence alignment with BamA from *E. coli* (*Ec*BamA) highlights the similarities and differences in *Kp*BamA and *Kp*BamK, with comparison to the modification sites found in phage-resistant *Ec*BamA(Φ) (Smith et al., 2007). A multiple sequence alignment detailing the full sequences of *Kp*BamA and *Kp*BamK from several type strains of *K. pneumoniae* is presented in Fig. S3. (C) Superposition of *Ec*BamA (red) structure (pdb: 5ekq) and the model of *Ec*BamA(Φ) generated by phyre2 (green), highlighting the differences in the loops contributing to the surface dome (boxed). Inset: an expanded view highlights how the increased length in loops 4 and 7 raise them to the surface of *Ec*BamA(Φ), accommodated by the diminished size of loop 6 in *Ec*BamA(Φ) relative to *Ec*BamA. The extent of changes to the extracellular dome can be appreciated from the top-down view, shown as a space-filling representation to take into account side-chains on each of the loops.

SUPPLEMENT

Supplementary Figure S1. BamK is not essential for viability or virulence in *Klebsiella pneumoniae*. (A) Growth curves are shown for B5055 (wild-type) and the isogenic Δ*bamK* mutant strain as monitored on rich (LB), minimal (M9) or "host-like" (DMEM) medium. (B) Groups of five BALB/c mice were infected with 8-9 $\times$ 10^3 CFU *K. pneumoniae* B5055 wildtype, or the isogenic Δ*bamK* mutant strain, or the complemented mutant (Δ*bamK* C') by

intranasal administration. After 4, 24, 72 and 96 h post-infection, the number of bacteria in the lungs and liver were enumerated: the colonization of the liver has begun in a few individuals at ~24 hours, and is substantial in all individuals by 72 hours. The mean is shown for each group. (C) Transcript levels for *bamA* and *bamK* were determined through qPCR, using RNA samples extracted from log phase *K. pneumoniae* B5055 cultured in DMEM. Data is expressed relative to transcript levels from the constituently-expressed *rpoD.*

Supplementary Figure S2. **The 5'-untranslated region of *bamK*.** (A) Sequence (1000bp) upstream of the *bamK* open-reading frame. The putative -35 and -10 element of the predicted σ^{70} promoter is labeled in red, a palindromic sequence overlapping the -10 element is underlined. The predicted ribosome-binding site (RBS) is shown in blue and the ATG start codon is indicated in italics. (B) The predicted stem-loop structure ("native sequence") that would be possible from the palindromic sequence upstream of *bamK*, in comparison to the modified sequence ("non-palindrome") which would not form a stem-loop structure. Promoter activity analysis in *K. pneumoniae* B5055 carrying the transcriptional fusions pPROBE'-gfp[bamA], pPROBE'-gfp[bamK] or pPROBE'-gfp[bamK non-palindrome]. Data are presented on a scatter plot (bar denotes mean of three biological replicates). (C) Schematic showing the replacement of the *bamA* coding sequence with the *bamK* coding sequence to produce gene replacement strain *K. pneumoniae* B5055 Δ*bamA*::*bamK*. A similar strategy was used to achieve gene replacement strains in *E. coli* BL21 Star™ (DE3) (Δ*bamA*::Kp*bamA* and Δ*bamA*::Kp*bamK*).

Supplementary Figure S3. Conserved sequence features of BamK. (A) Multiple sequence alignment of a select set of BamA and BamK proteins sequences. Residues are colored according to side-chain properties to guide assessment of conservative and non-conservative substitutions. The boundaries for the five POTRA domains are indicated, based on the crystal structure of BamA from *E. coli*. The positions of the transmembrane β-strands in the β-barrel domain of BamA from *E. coli* are indicated by arrows and are numbered (β1, β2 etc). Red lines indicate the extracellular loop regions, blue dots indicate the sequences in the periplasmic turns. (B) Schematic of predicted domain structure of BamK.

Supplementary Figure S4. Monoclonal antibodies that selectively recognize BamK. (A) Schematic outlining the screening strategy for the production of monoclonal antibodies (MAb) to BamK and the subsequent cloning of the MAbK clones 1-5 (MK1-MK5). At left, a

structural overlay of the modelled structures for *Kp*BamA (red) and *Kp*BamK (blue). This and multiple sequence alignments suggested a peptide at the junction of POTRA3 and POTRA4 (IALNEGERYRVDRT$_{271}$) as being diagnostic of BamK. (B) Total cell extracts (100 μg protein) were prepared from *K. pneumoniae* strains B5055 and B5055*bamA*::*bamK* and analysed by SDS-PAGE and immunoblotting. Replicate blots were probed with the monoclonal antibodies MAbK.1, MAbK.2, MAbK.3, MAbK.4 or MAbK.5 (MK1-MK5). The migration position of molecular weight standards is indicated. (C) Total cell extracts (200 μg protein) were prepared from *K. pneumoniae* strains B5055 (wild-type) and Δ*bamA*::*bamK* and analysed by BN-PAGE and immunoblotting. Replicate blots were probed with the monoclonal antibodies MAbK.1, MAbK.2, MAbK.3, MAbK.4 or MAbK.5 (MK1-MK5). The migration position of molecular weight standards is indicated. Clone MAbK3 was immortalized for large scale production of the monoclonal antibody, as it recognizes both denatured and native conformations of BamK.

Supplementary Figure S5. BamK can functionally replace BamA in either *K. pneumoniae* or in *E. coli.* (A) Growth curves for the Δ*bamA::bamK* replacement strain cultured in LB medium. (B) Outer membranes (200 μg protein) were subject to BN-PAGE and immunoblotting with the indicated antiserum. (C) Growth curves for *E. coli* strains Δ*bamA*::Kp*bamA* and Δ*bamA*::Kp*bamK*. The three strains were cultured in LB medium. (D) Outer membranes (200 μg protein) were prepared from *E. coli* BL21 Star™ (DE3), *E. coli* strain Δ*bamA*::Kp*bamA* and *E. coli* strain Δ*bamA*::Kp*bamK*. Membrane fractions were analysed by BN-PAGE and immunoblotting with the indicated antiserum. The BAM complex containing BamK (BAM$_{KBCDE}$) migrates diagnostically more slowly that the BAM$_{ABCDE}$ complex. (E) Outer membranes (200 μg protein) were prepared from *E. coli* strain Δ*bamA*::Kp*bamK* or *E. coli* strain Δ*bamA*::Kp*bamK*,Δ*bamB* or *E. coli* strain Δ*bamA*::Kp*bamK*,Δ*bamC*. Membrane fractions were analysed by BN-PAGE and immunoblotting with the indicated antiserum. The migration positions for the sub-complexes generated when BamB or BamC is absent (Webb et al 2012b) are indicated. For example, in the absence of BamC, most of the BAM$_{AB}$ module is released from the other components (Webb et al 2012b). Alternatively, in the absence of BamB, most of the BAM complex is recovered as a BAM$_{ACDE}$ form. Irrespective of the precise identities of each species generated, the analysis shows a diagnostic change in the mobility of BamK on BN-PAGE.

Supplementary Figure S6. Formation of the LptD/E complex in *E. coli* depends on co-expression of both LptD and LptE. (A) The rate of assembly of [^{35}S]-LptD was assessed in *E. coli* strains Δ*bamA*::*KpbamA* and Δ*bamA*::*KpbamK*. Cells harboring pCJS42 (encoding LptD) were starved of methionine and cysteine, pulse-labeled with ^{35}S-methionine/^{35}S-cysteine for 40 seconds, followed by a 60-minute chase with unlabeled methionine and cysteine. Aliquots were taken at 10 sec, 2, 4, 8, 16 and 32 minutes. Samples were incubated at either 37 °C or 95 °C before analysis by semi-native SDS-PAGE, phosphorimaging and densitometry. The position of the 95 kDa LptD monomer is indicated. (B) The rate of assembly of [^{35}S]-LptE was assessed in *E. coli* strains Δ*bamA*::*KpbamA* and *E. coli* strains Δ*bamA*::*KpbamK*. Cells harboring pCJS43 (encoding LptE) were assayed as described above. The position of the 20 kDa LptE monomer is indicated. (C) The rate of assembly of [^{35}S]-LptD and [^{35}S]-LptE into the LptDE complex as assessed in *E. coli* strains Δ*bamA*::*KpbamA* and *E. coli* strains Δ*bamA*::*KpbamK*. Cells harboring pCJS44 (a pET-Duet plasmid encoding both LptD and LptE) were assayed as described above. The position of the 20 kDa LptE monomer, 95 kDa LptD monomer and the LptDE complex are indicated.

Supplementary Figure S7. The distribution of BamK across bacterial species. The tree was calculated using RaxML as described in the methods section. In the representation of the analysis, the *Klebsiella*-type BamK sequences (blue) are segregated from BamA (and other BamA-related) sequences (red). The taxonomic grouping of the further species for which BamA sequences were analyzed is indicated by colors in a legend (inset). Furthermore, the branch colors illustrate members of the *Pasteurellaceae* (orange) or *Enterobacteriaceae* (green), with this also highlighted alongside the sequence entries. Zooming in on the Figure reveals the Uniprot identifiers for each sequence. All sequence entries are further documented in Table S2.

Supplementary Figure S8. BamK is present, along with BamA, across species and subspecies of *Klebsiella pneumoniae*. Phylogenetic tree of the *K. pneumoniae* amino acid sequences (Table S1) to distinguish BamA and BamK. As indicated in Fig. 4, the tree is based on the core gene alignment of the global *Klebsiella* collection (Holt et al 2015), the outer stripe of color-coding indicates the species and subspecies of the *Klebsiella pneumoniae* complex. The ubiquitous presence of *bamA* genes is indicated in the red (inner) stripe, and *bamK* genes are indicated with blue. BamA and BamK are found throughout the entire

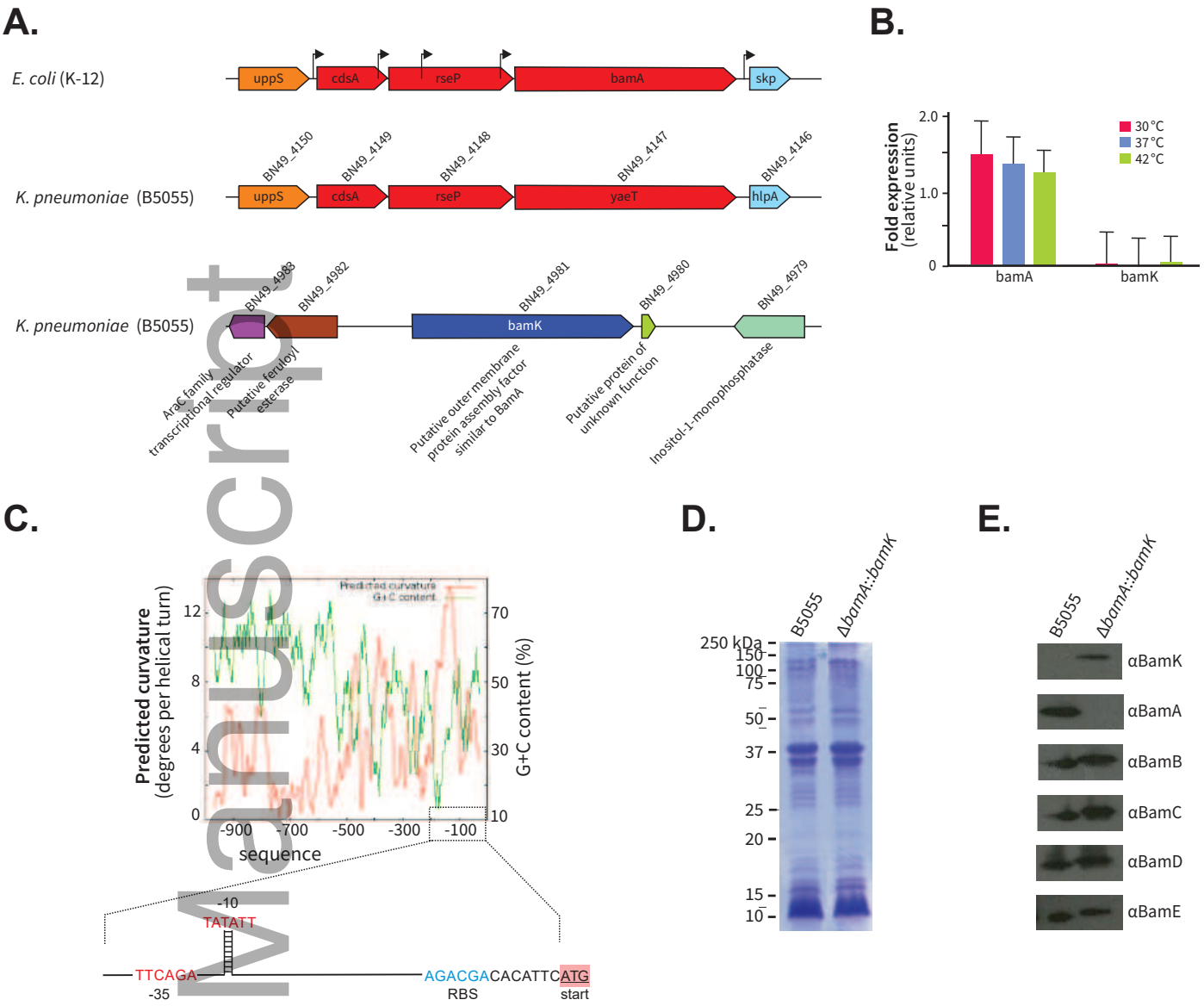
Klebsiella pneumoniae complex diversity. Zooming in on the Figure reveals the Uniprot identifiers for each sequence.

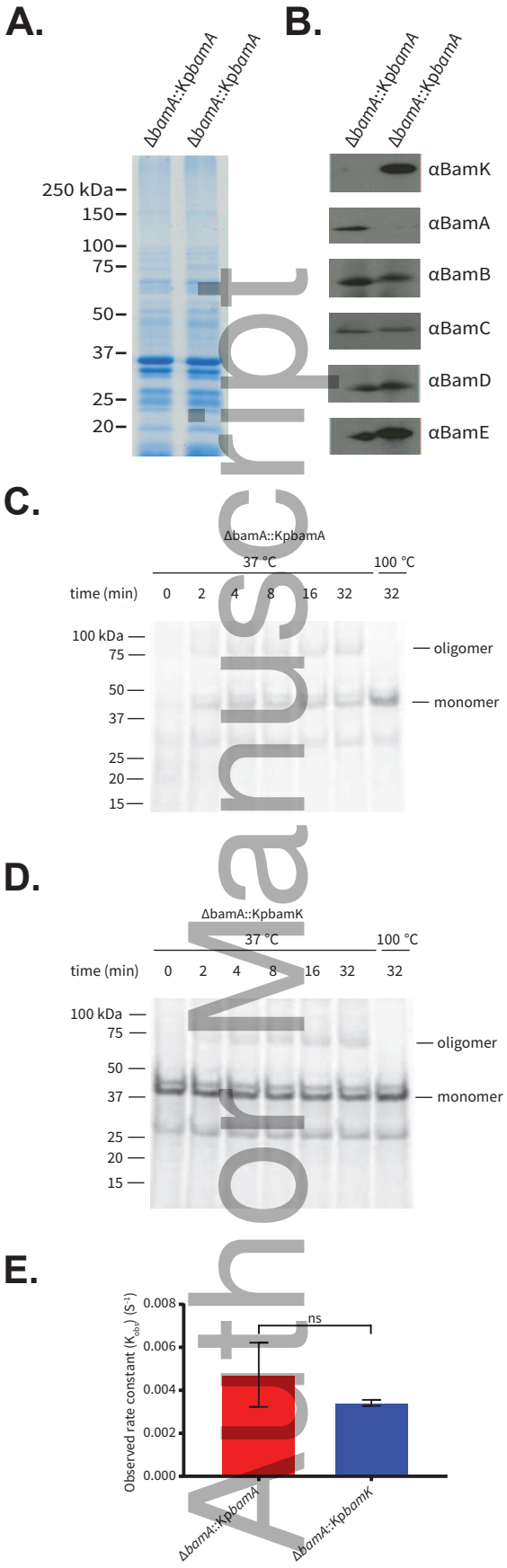
Supplementary Table S1. *K. pneumoniae* strains and accession numbers, and their respective BamA and BamK sequence identifiers.

Supplementary Table S2. BamA and BamK accession numbers.

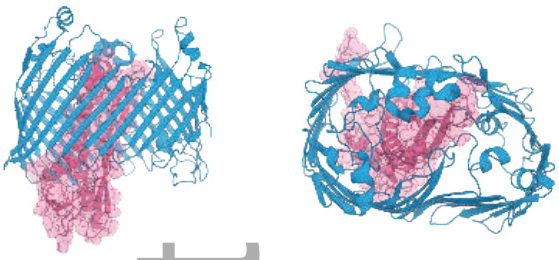
Supplementary Table S3. Strains and plasmids.

Supplementary Table S4. Oligonucleotide primer sequences.

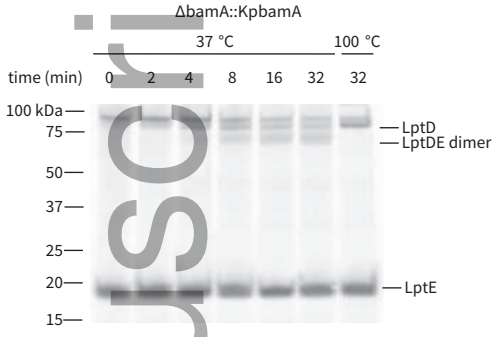




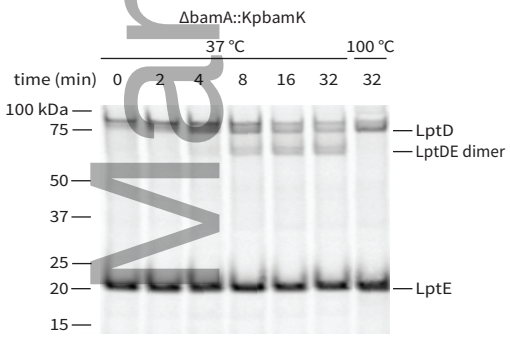
A.



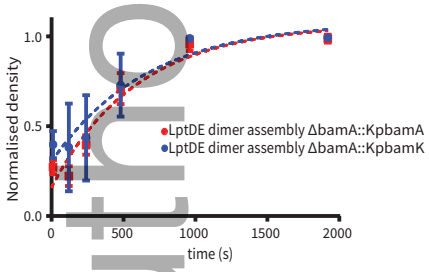
B.



C.



D.



E.