Deep transcriptome profiling of clinical *Klebsiella pneumoniae* isolates reveals strain- and sequence type-specific adaptation

Running title: Comparative transcriptomics of *Klebsiella pneumoniae*

Sebastian Bruchmann^{1,2}, Uthayakumar Muthukumarasamy^{1,2}, Sarah Pohl^{1,2}, Matthias Preusse¹, Agata Bielecka^{1,2}, Tanja Nicolai¹, Isabell Hamann³, Roger Hillert³, Axel Kola⁴, Petra Gastmeier⁴, Denitsa Eckweiler^{1,2}, Susanne Häussler^{1,2*}

¹ Department of Molecular Bacteriology, Helmholtz Centre for Infection Research, Braunschweig, Germany

² Institute for Molecular Bacteriology, Twincore, Centre for Clinical and Experimental Infection Research, A Joint Venture of the Helmholtz Centre for Infection Research and the Hannover Medical School, Hannover, Germany

³ Medizinisches Labor Ostsachsen, Mikrobiologie, Görlitz, Germany

⁴ Institute of Hygiene and Environmental Medicine, Charité University Medicine Berlin, Berlin, Germany.

*Corresponding author:

Susanne Häussler

susanne.haeussler@helmholtz-hzi.de

Helmholtz Centre for Infection Research, Inhoffenstrasse 7, Braunschweig, DE 38124

Phone: +49 531 6181 3000

Fax: +49 531 6181 3099

Summary

Health care-associated infections by multidrug-resistant bacteria constitute one of the greatest challenges to modern medicine. Bacterial pathogens devise various mechanisms to withstand the activity of a wide range of antimicrobial compounds, among which the acquisition of carbapenemases is one of the most concerning. In *Klebsiella pneumoniae* the dissemination of the carbapenemase KPC is tightly connected to the global spread of certain clonal lineages. Although antibiotic resistance is a key driver for the global distribution of epidemic high-risk clones, there seem to be other adaptive traits that may explain their success. Here, we exploited the power of deep transcriptome profiling (RNA-seq) to shed light on the transcriptomic landscape of 37 clinical *K. pneumoniae* isolates of diverse phylogenetic origins. We identified a large set of 3346 genes which was expressed in all isolates. While the core-transcriptome profiles varied substantially between groups of different sequence types, they were more homogenous among isolates of the same sequence type. We furthermore linked the detailed information on differentially expressed genes with the clinically relevant phenotypes of biofilm formation and bacterial virulence. This allowed for the identification of a diminished expression of biofilm-specific genes within the low biofilm producing ST258 isolates as a sequence type-specific trait.

Introduction

Adaptation of bacteria to complex and changing environments is well reflected by global changes in gene expression profiles, which may become fixed by adaptive mutations to facilitate survival in challenging habitats (Cooper et al., 2003). Thus, to study microbial pathogenesis and to fully understand bacterial adaptation strategies it is crucial to explore the genomic make-up of a pathogen and to analyze how this influences the transcriptional landscape. One of the most crucial adaptations a bacterial pathogen has to accomplish is the adaptation to the activity of antimicrobial compounds. In the last decade, multidrug-resistant bacterial pathogens have been isolated from patient material at constantly increasing rates (Tavares et al., 2013). This poses a serious threat to human health and leads to alarming limitations of treatment options, especially against the so called ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species) (Rice, 2008). *K. pneumoniae*, one of these high-risk pathogens, is a ubiquitous Gram-negative bacterial member of the Enterobacteriaceae family, closely related to *Escherichia coli*. While *E. coli* is primarily a commensal organism, *K. pneumoniae* is colonizing mucosal surfaces of humans and livestock, but also inhabits natural environments (Podschun and Ullmann, 1998). *K. pneumoniae* is responsible for opportunistic community- and hospital-acquired human infections mainly of the urinary- and respiratory-tract. Together with *E. coli* it is the leading cause of bacteremia in the UK (Livermore et al., 2008) and causes infectious diseases with high mortality rates which exceed 50 % for severe systemic infections (Podschun and Ullmann, 1998). *K. pneumoniae* has acquired diverse mechanisms to withstand the activity of a wide range of

antimicrobial compounds (Walsh, 2000; Walsh, 2003; Alekshun and Levy, 2007), among which the acquisition of the *Klebsiella pneumoniae* carbapenemase (KPC), an enzyme which degrades virtually all classes of β -lactams, is one of the most concerning (Nordmann et al., 2009; Daikos and Markogiannakis, 2011; Rapp and Urban, 2012). Studies showed an increase in mortality rates to more than 70 % when hospitalized patients are infected with KPC-positive isolates (Marchaim et al., 2008; Borer et al., 2009).

In the last years, numerous studies focused on the epidemiology, phylogeny and the genomic structure of *K. pneumoniae* (Chen et al., 2014b; Holt et al., 2015), yet little is known about the global transcriptome profile of this important bacterial pathogen. Only a limited number of transcriptional studies on *K. pneumoniae* have been performed. However, they deliver valuable and important insights into the transcriptional architecture and regulatory aspects of *K. pneumoniae* (Kim et al., 2012; Seo et al., 2012), its adaptation to changing environments (Li et al., 2014) and the development of colistin resistance (Wright et al., 2014a).

High throughput sequencing of messenger RNAs (RNA-seq) offers the ability to study gene expression on a global scale and to analyze qualitative and quantitative gene expression in an unbiased manner (Croucher and Thomson, 2010; Sorek and Cossart, 2010; van Vliet, 2010).

In this study, we used RNA-seq to determine the conservation and variation of gene expression profiles across 37 clinical *K. pneumoniae* isolates of various phylogenetic origins – including ten isolates of the highly successful epidemic clone ST258. We found that the core-transcriptome of the *K. pneumoniae* isolates comprises as many as 3346 genes and that global core gene expression profiles are similar among, but distinct between the various *K. pneumoniae* sequence types. Our analysis of how differentially expressed genes correlate to clinically relevant phenotypes such as biofilm formation and bacterial virulence uncovered *K.*

pneumoniae sequence type-specific traits which might facilitate successful spreading or survival of the epidemic clonal lineage in the hospital setting.

Results

Phylogenetic distribution of *K. pneumoniae* clinical isolates included in this study

With the aim to gain detailed insights into the variation of the *K. pneumoniae* transcriptional landscape, we performed deep transcriptome sequencing. 34 clinical isolates were sampled from various infection sites (blood, urine, wound- and rectal swabs) (supplementary table S1) from out-patients as well as in-patients of 10 German hospitals. The median patient's age was 65 years with a range from 20 to 90 years. Additionally to these 34 clinical isolates, the three previously completely sequenced *K. pneumoniae* clinical strains (MGH 78578 (Ogawa et al., 2005), JH1 and 1162281 (Kumar et al., 2011)) were also included in this RNA-seq study. The *K. pneumoniae* isolates were grown under standard laboratory conditions in rich medium at 37 °C until late logarithmic growth phase. After RNA extraction and ribosomal RNA depletion, barcoded transcriptome sequencing (Dotsch et al., 2012) was performed on an Illumina HiSeq 2500. This generated a total of 787 million paired-end reads, each with a length of 100 base pairs. Up to 99 % of the reads (13.1 to 36.6 million reads per isolate with a median of 20 million reads, supplementary table S1) could be mapped to a non-redundant gene list of the *K. pneumoniae* pan-genome. This pan-genome was generated on the basis the fully sequenced genome of the clinical strain 1084 (sequence type 23) (Lin et al., 2012) complemented with publicly available DNA sequence data of 10 *K. pneumoniae* isolates (see Materials and Methods for details) to gain comprehensive and unbiased insights into the

transcriptional landscape of this pathogen. To exclude any bias in mapping efficiency due to the choice of the core genome reference strain, we also created a pan-genome based on a different sequence type (MGH 78578; ST38). The comparison of mapping efficiency revealed an average difference of just 2,000 reads (or 0.001%) when using the two alternative pan-genomes.

Since deep transcriptome sequencing delivers not only quantitative data on the gene expression profile but also high quality sequence data of transcribed coding sequences (Wang et al., 2009), we used the information on complementary (c)DNA sequence variations among the strains to analyze their phylogenetic relationship. Figure 1 depicts the phylogenetic distribution of our set of 37 clinical isolates as well as of the 11 previously completely sequenced reference strains. The tree is based on the complete sequence of 404 genes corresponding to 314,561 nucleotide positions, all of which were covered by at least five sequencing reads in our pool of clinical isolates. Overall 35,796 variable sites were detected among the strains. In addition, to this (c)DNA based phylogenetic reconstruction, we performed multilocus sequence typing (MLST) (Diancourt et al., 2005) and *wzi* typing (Brisse et al., 2013) for all isolates (the results are also shown in figure 1). The *Klebsiella* MLST scheme consists of the partial sequences of eight house-keeping genes: *gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB* and *tonB*. The nucleotide sequences of *gapA*, *infB*, *mdh*, *pgi*, and *rpoB* were extracted from the RNA-seq data. Since the coverage of *phoE* and *tonB* did not match our strict quality criteria, Sanger sequencing was performed for these two genes as already described (Diancourt et al., 2005). The full sequence of the *wzi* gene, encoding the outer membrane protein Wzi, which is involved in cell surface capsule attachment, was also sequenced by Sanger sequencing. Wzi is highly conserved in distinct capsular (K) types (Brisse et al., 2013; D'Andrea et al., 2014; Wright et al., 2014b). However, Wyres and

colleagues recently highlighted the vast number of variant capsular types within the clonal group (CG) 258 cluster (Wyres et al., 2015), indicating the limited utility of *wzi* for sub-typing purposes.

Our analysis revealed a broad phylogenetic distribution of the *K. pneumoniae* isolates included in this study. We detected 12 different multilocus sequence types (ST), one of which (isolate 2_7, MLST pattern: 2-2-2-1-1-1-91) has not been described before and also 13 different *wzi* types, two of which have not been described yet.

13 of the 37 isolates were of ST101 origin. ST101 isolates are distributed world-wide and are known to carry beta-lactamase OXA-48 and the extended spectrum beta-lactamase CTX-M-15 (Potron et al., 2013). Furthermore, they have been reported to be involved in clinical outbreaks in Europe (Pitart et al., 2011; Poulou et al., 2013). 11 of these 13 ST101 isolates have been isolated from the same hospital within a period of five months. They showed no sequence variation in the phylogenetic analysis indicating that these isolates are likely to have their origin in a single reservoir.

Additionally, we identified ten and two isolates, belonging to the broadly distributed multidrug-resistant high-risk clones ST258 and ST512 (Woodford et al., 2011), respectively. Both sequence types are members of the clonal group CG258, sharing seven out of eight MLST alleles (Breurec et al., 2013). Isolates of the sequence type 258 are reported to be involved in several European and North American hospital outbreaks (Kitchel et al., 2009; Grundmann et al., 2010) and are the primary reason for the spread of KPC (Woodford et al., 2011). Here, the *K. pneumoniae* carbapenemase KPC-2 was identified in all ten ST258 isolates, whereas both ST512 isolates carried KPC-3. Of note, all other 25 sequenced clinical isolates were found to be KPC negative. Recent studies described the heterogeneous nature of ST258 isolates with two large clades ST258 I and ST258 II (Chen et al., 2014a; D'Andrea et

al., 2014; Deleo et al., 2014; Wyres et al., 2015). All ST258 isolates of this study were members of clade ST258 I (described as ST258a in (Wright et al., 2014b)) due to the *wzi* allele *wzi*-29 (K41). The two ST512 isolates shared the same *wzi*-154 allele (Diago-Navarro et al., 2014) with strains NJST258_1 and NJST258_2 of clade ST258 II and KP13 (ST442), indicating that the ST512 isolates evolved from an ST258 clade II ancestor as suggested previously (D'Andrea et al., 2014).

Another successful and multidrug-resistant clone is ST15 (Damjanova et al., 2008; Liapis et al., 2014; Rodrigues et al., 2014). Our collection contained four isolates of sequence type ST15 which split into two distinct phylogenetic groups based on 9 variable sites within the sequence of the 404 conserved genes. This separation was also supported by *wzi* sequencing, since two isolates (isolate 1_16 and isolate 2_2) shared allele *wzi*-24 (K24) (clade ST15 I) and two isolates (isolate 1_8 and isolate 2_10) shared *wzi*-93 (K60) (clade ST15 II). *wzi*-24 (clade I) has been identified previously in ST15 as well as ST59 isolates (Brisse et al., 2013) and *wzi*-93 (clade II) has been identified in a New Delhi metallo-beta-lactamase positive isolate from Nepal (Stoesser et al., 2014). The remaining seven isolates were of diverse phylogenetic origin.

The *K. pneumoniae* pan-genome

K. pneumoniae, like almost all other Gram-negative bacteria, shows a high variation in genome content (Holt et al., 2015). The size of the completely sequenced and annotated 11 genomes, which were used as a reference in this study, varies from 5.2 to 6.1 Mbp encoding for 4887 to 5577 genes per genome (table 1). Thus, in order to map a maximum of RNA sequencing reads from our collection of clinical isolates, we generated a *K. pneumoniae* pan-genome. This pan-genome is based on the 11 *K. pneumoniae* genomes which contain overall

57,312 chromosomal genes, with an average of 5,210 genes per strain. The non-redundant *K. pneumoniae* pan-genome contained 7859 genes (supplementary table S2). 3336 of those genes were shared by all 11 reference strains (“core-genome”) and 4523 genes were absent in at least one of the strains. Among the latter, 1598 genes were identified in only one of the reference genomes (“singletons”). Supplementary figure S1 depicts the development of the size of the pan-genome, core-genome and singletons by sequentially adding the genomic information of one genome to that of the others. Based on extrapolated data, it is expected that the size of these groups would change by less than 2 % if genomic information of another genome would be added, less than 1 % by adding information of 20 genomes, and less than 0.5 % by adding 40 genomes in total.

To further explore the genomic content of the pan-genome, we extracted for each gene the gene ontology (GO) information and simultaneously deduced the gene function using the COG database (Tatusov et al., 2000), where possible. As expected, the GO term analysis revealed an enrichment of house-keeping genes in the core-genome, which are essential to maintain cellular function and integrity (supplementary figure S2).

The transcriptional landscape of *K. pneumoniae*

The distribution of normalized reads per kilo base (nRPK) (Dotsch et al., 2012), after mapping all RNA sequencing reads against the non-redundant gene list of the *K. pneumoniae* pan-genome, was found to be continuous and unimodal (figure 2A), with a median nRPK of 4.22 and a maximum of 24.01. Depending on the isolate, between 4744 and 5378 genes exhibited an expression value above the sensitivity limit of 2.20 nRPK (see Materials and Methods section for definition), thus demonstrating a very high number of transcribed genes per isolate under the chosen conditions. Of all 7859 genes in the pan-genome, we identified a

large set of 3346 genes to be commonly expressed in all isolates above the threshold level. This core-transcriptome (highlighted in green in figure 2B) accounted for 62 % to 71 % of all transcribed genes within one isolate and largely overlapped with the core-genome (2515 genes, 75 % of the core-genome was commonly transcribed). Due to this large overlap between core-genome and core-transcriptome, it was not surprising that the core-transcriptome likewise consisted mostly of genes with house-keeping functions, as revealed by a GO term enrichment analysis (supplementary figure S2).

On the other end of the scale we found 719 genes within the pan-genome that exhibited maximal expression values below the sensitivity limit (highlighted in red in figure 2B). Those genes were not expressed in any clinical isolate under standard laboratory conditions. Interestingly, only eight of them belonged to the core-genome, whereas 473 (almost 66 %) of these never expressed genes were singletons and occurred only in single reference genomes. With the aim to map RNA sequencing reads also to potentially non-chromosomally encoded genes, we performed a *de novo* assembly of all reads from the 37 isolates that did not map to the reference pan-genome. We found a total of 1482 genes with an average of 251 genes per isolate and a range from 57 to 419 that did not map to our reference pan-genome. 1258 (85 %) of these genes had homologs with at least 80 % sequence identity to genes that have been previously identified in *Klebsiella spp.*. 150 of the remaining 187 genes had homologs in other members of the Enterobacteriaceae family. While the majority of 60 % (900 out of 1486 genes) was encoding for hypothetical (including putative and predicted) proteins, many genes were found to be related to antibiotic and metal resistance, integrase/transposase, phage-related and involved in plasmid integrity. The complete list of genes potentially encoded on plasmids, integrons or other mobile elements is shown in supplementary table S3.

The *K. pneumoniae* gene expression variance

To gain further insights into the expression variance, we calculated the differential expression of each gene in any possible pairwise comparison of two of the 37 isolates (overall 666 comparisons). Pairwise gene comparison data were considered only if a gene in both isolates reached expression values above the sensitivity limit of 2.20 nRPK. The expression variance of one gene was defined as the fraction (between 0 and 1) of pairwise comparisons showing at least a two-fold differential expression as determined by DESeq (Anders and Huber, 2010). Thus, for example, an expression variance of 0.1 means that 10% of the pairwise comparisons showed at least a two-fold difference in expression values. We obtained information about the expression variance of overall 5696 genes. Ribosomal RNA coding genes were excluded from this analysis due to the influence of rRNA removal during library preparation.

571 genes were found to be very stably expressed with an expression variance of 0.1 or less (supplementary table S2). These included 28 genes which were never differentially expressed according to the DESeq analysis. On the other hand, 575 genes exhibited a large gene expression variance, between 0.58 and 0.83. A GO term analysis (supplementary figure S2) revealed enrichment of genes involved in translation, replication, cell cycle control, cell wall synthesis and coenzyme transport in the most stably expressed genes; whereas only genes involved in the class of energy production and conservation were highly enriched in the set of variably expressed genes. The stably expressed genes were slightly, but significantly higher expressed (average nRPK value of 6.82) compared to the variably expressed genes (average nRPK value of 6.05; two-tailed Student's t-test, P-value < 0.001). The core-transcriptome exhibited significantly lower expression variance (0.31) as compared to the set of genes that did not belong to the core-transcriptome (0.41); two-tailed Student's t-test, P-value < 0.001.

Figure 3 shows a scatter plot of the expression variance of the core-transcriptome as compared to the individual minimal (gray dots) and maximal (black dots) gene expression values. We observed a rather homogenous expression pattern across the core-transcriptome with smoothly increasing distance between the average minimal (bottom line) and maximal (top line) gene expression. This was even more pronounced in the set of genes not belonging to the core-transcriptome (not shown).

Correlation between phylogenetic background and the transcriptional profile

We next assessed whether and how the genetic background of the various clinical isolates impacts on global gene expression profiles. We performed a hierarchical clustering based on the Spearman rank correlation of all 37 *K. pneumoniae* isolates according to the overall similarity of the expression profile of all 7859 genes in the pan-genome (figure 4A). Hierarchical clustering revealed three major sub-groups: the first included the two ST512 isolates together with all ten ST258 isolates, the second group included all except one ST101 isolates and the third group included the four ST15 isolates together with various other sequence types. A principal component analysis (PCA) (figure 4C) also demonstrates a clear separation of the transcriptional profiles of all major phylogenetic lineages within our set of isolates. This clustering might have been expected since the genomic composition i.e. the presence of distinct sets of accessory genes of the various MLST sequence types strongly influences the clustering.

To evaluate the impact of the transcription of accessory genes, we next analyzed the clustering of the 37 clinical isolates based on variations within genes of the core-transcriptome. Remarkably, clustering of the expression profiles based solely on the core-transcriptome (figure 4B) still revealed concordance with the phylogenetic clustering.

Although the separation into distinct phylogenetic groups - becoming especially apparent in the PCA (figure 4D) - was not as strong as observed for the hierarchical clustering based on the pan-genome, the phylogenetic groups could still be separated.

Expression of small, non-coding RNAs

Our comprehensive RNA-seq dataset enabled the investigation of the expression of small, non-coding RNAs (sRNAs), which are known to be involved in central regulatory pathways in a vast variety of bacterial pathogens (Papenfort and Vogel, 2010; Caldelari et al., 2013). 51 sRNAs have previously been annotated in the 11 *K. pneumoniae* reference genomes and are therefore included in the pan-genome used in this study. Another set of 30 sRNAs were predicted using sRNAscanner (Sridhar et al., 2010) with default parameters and were additionally included in the pan-genome.

Out of these 81 sRNAs, 34 (including seven of the predicted sRNAs) were commonly expressed in all isolates (see figure 5 and supplementary table S4) and only seven RNAs (six of them predicted by sRNAscanner) were never expressed in any isolate under the chosen culture conditions. The average expression as well as the expression variance of all sRNAs varied greatly. The most stable (and also highest) expressed small RNA was the transfer-messenger RNA *ssrA*, whereas 24 sRNAs were regarded as highly variably expressed genes with an expression variance of 0.58 to 0.79. Among them, *cyaR* (*ryeE*), *glmZ* (*ryiA*), and *gcvB* exhibited the highest average expression levels. The expression of the Hfq-binding regulatory RNA CyaR (RyeE) is tightly controlled by the cyclic AMP receptor protein Crp under high cyclic adenosine monophosphate (cyclic-AMP) levels and represses the translation of the OmpX porin in *Salmonella enterica* serovar Typhimurium (Papenfort et al., 2008) and *E. coli* (De Lay and Gottesman, 2009). CyaR has been linked to carbon metabolism regulation (Wright et al.,

2013) and was found to be highly expressed in *Yersinia pestis* murine lung infections (Yan et al., 2013). The sRNA GlmZ (RyiA) is an activator of *glmS*, encoding the glucosamine 6-phosphate (GlcN-6-P) synthase GlmS. GlcN-6-P is a precursor molecule of peptidoglycan and lipopolysaccharide biosynthesis, both are essential components of cell walls and outer membranes of Gram-negative bacteria (Reichenbach et al., 2008). Another highly expressed sRNA was GcvB, which is a conserved Hfq-dependent small RNA involved in the regulation of amino acid uptake and biosynthesis (Urbanowski et al., 2000; Miyakoshi et al., 2015). GcvB is highly abundant in nutrient-rich growth conditions (Sharma et al., 2007) and plays a central role in amino acid homeostasis (Miyakoshi et al., 2015).

Virulence of *K. pneumoniae* in *Galleria* wax moth larvae is independent of the phylogenetic background

Since the overall expression profiles of the core-transcriptome seemed to be associated with distinct *K. pneumoniae* sequence types, we wondered whether this global expression profile determines bacterial behavior which might contribute to the success of global clonal lineages. The larvae of the greater wax moth *Galleria mellonella* are widely used as a model to study the virulence of pathogens (Pustelny et al., 2013; Alghoribi et al., 2014; Giannouli et al., 2014), including *K. pneumoniae* (Insua et al., 2013; Wand et al., 2013; Diago-Navarro et al., 2014; McLaughlin et al., 2014). Here, we used the *G. mellonella* infection model to test for the virulence phenotype of the various sequence type isolates. Healthy *G. mellonella* larvae were injected with an average dose of 7.5×10^5 cfu *K. pneumoniae* and the larval survival was monitored for 72 hours at 37 °C (figure 6). Virulence differed greatly not only among the various MLST groups but also within them. Furthermore, isolates with the same capsular type, which was previously shown to play an important role in *K. pneumoniae*

virulence (Yu et al., 2007), did not exhibit similar virulence profiles. Our data indicate that virulence-associated traits are independent from the phylogenetic background, and therefore are rather isolate specific.

We found one isolate (isolate 3_8 of ST101) that exhibited a very low virulence. When we analyzed the growth behavior of all isolates in rich medium and in minimal medium supplemented with two different carbon sources (glucose and mannose, respectively; supplementary table S5), this isolate exhibited a severe growth defect, which might explain its low virulence potential (Diago-Navarro et al., 2014). For all other isolates there was no correlation between growth and pathogenicity.

To identify possible novel virulence associated genetic determinants, we compared the transcriptomes of 11 highly virulent isolates (≥ 90 % dead larvae after 72 hours, red diamonds in figure 6) with 12 lowly virulent isolates (between 10 % and 50 % dead larvae, yellow circles in figure 6) and performed transcriptome-wide associations to search for differentially transcribed genes within these two subsets. However, we could not identify any statistically significant differences in gene expression profiles among the two sub-sets.

To gain an overview about the transcription of genes previously associated with bacterial pathogenicity, we specifically quantified transcription of genes that have been listed in the virulence factor data base VFDB (Chen et al., 2012) and the *K. pneumoniae* BIGSdb database (Bialek-Davenet et al., 2014). 86 of those genes belonged to the *K. pneumoniae* pan-genome used in this study, the expression of which is shown in figure 7 and supplementary table S6. Clearly, the overall expression of the virulence-associated genes was low. This might be due to the fact that the transcriptional profiles were recorded under standard laboratory conditions and genes that might contribute to the differential in virulence in the *G. mellonella* infection model might not be expressed under rich medium conditions.

Type 1- (*fimD*) and type 3- (*mrkABCDJIH*) fimbriae gene clusters were highly and almost universally expressed, whereas CFA/I- (*yagVWXYZykgK*) fimbriae genes were - albeit widespread - expressed at lower levels. Besides fimbriae genes, genes encoding for a diverse array of siderophores were commonly expressed in the clinical isolates, whereas the microcin E492 gene cluster was expressed in only one isolate of ST405. Interestingly, expression of the yersiniabactin siderophore gene cluster (*fyuA* to *ybtS*) was linked to the phylogenetic origin as it was expressed in isolates of ST35 and ST405, both ST15 *wzi*-24 isolates and all ST101 isolates. However, there was no general correlation of virulence gene expression and the virulence phenotype as measured in the *G. mellonella* infection model.

Since virulence among the isolates could also be variable due to the presence of accessory genes on plasmids or integrons acquired by horizontal gene transfer (Darmon and Leach, 2014), we extracted all reads that did not map to the 11 strain-*K. pneumoniae*-pan-genome and compared the accessory transcriptome to VFDB and BIGSdb, both of which do not only contain chromosomal genes (supplementary table S7). However, again, none of these genes were differentially expressed in the group of highly virulent isolates versus the lowly virulent ones (data not shown).

ST258 isolates produce significantly less biofilm as compared to other MLST sequence types

In addition to virulence, the capability to form biofilms is a clinically relevant bacterial phenotype. We next analyzed the capability of each clinical isolate to produce biofilms by the use of a crystal violet assay (figure 8A). Interestingly, as opposed to the virulence phenotype, the capability to produce biofilms seemed to be associated with the affiliation to a specific sequence type (figure 8B). As shown previously (Naparstek et al., 2014), biofilm formation

was significantly lower in ST258 isolates as compared to other isolates, with a median OD₅₉₀ of 0.06 versus 0.28 for all other isolates, respectively (two-tailed Student's t-test, P-value < 0.01). The isolate with the highest absorbance in the crystal violet assay was the one isolate of ST101 (isolate 3_8) that already exhibited a non-virulent phenotype in the *Galleria* assay and showed a severe growth defect in minimal media. Among all other isolates, growth was not associated with high or low biofilm formation.

We classified the clinical isolates according to their ability to form biofilm into high producers (OD₅₉₀ > 0.187, 11 isolates (all but one ST101 isolate were not included due to their probable clonal nature)) and low producers (OD₅₉₀ < 0.187, 15 isolates) and analyzed whether we can detect group specific gene expression profiles. We could identify 90 genes whose expression was significantly different in the high biofilm forming group as compared to the low biofilm forming group (supplementary table S8). The list of 46 genes, which were higher expressed in the high biofilm producing group (figure 9 and table 2), included *mrkA* encoding the major pilin subunit of type 3 fimbriae (Allen et al., 1991), which has been shown previously to be crucial for binding to abiotic surfaces (Langstraat et al., 2001). The remaining genes of the *mrk* operon, *mrkB* to *mrkF*, were also found to be higher expressed, although with a corrected P-value higher than 0.05. MrkB, MrkC and MrkF are chaperones, outer membrane usher and minor pilin subunit, respectively (Huang et al., 2009). MrkD is the adhesive subunit of type 3 fimbriae and has been shown to facilitate binding to extracellular matrix proteins (Jagnow and Clegg, 2003). Another fimbrial cluster of three genes (D364_17530, D364_17540 and D364_17550) was significantly higher expressed in the biofilm forming isolates. Besides fimbrial genes, the genes *acsA*, *bcsB* and *bcsC* were expressed at higher levels. They all are involved in cellulose synthesis, an extracellular polysaccharide found in bacterial biofilms (Zogaj et al., 2003). It has been shown that the cellulose synthase BcsB is

activated via cyclic di-GMP (Morgan et al., 2014), a bacterial second messenger well known for its control of biofilm formation (Romling et al., 2013). Another operon which was higher expressed in the biofilm forming isolates was the fumarate reductase operon. Although only *frdD* is listed, *frdA*, *frdB* and *frdC* were also expressed on higher levels, but just below our strict threshold. It has been shown that inhibition of the fumarate reductase in *Porphyromonas gingivalis* significantly inhibited biofilm formation (Dashper et al., 2010). Furthermore, the outer membrane protein *ompW* was expressed at higher levels in the biofilm forming isolates. It has been shown that the *ompW* homolog in *P. aeruginosa* *oprG* is highly expressed in biofilms (Mikkelsen et al., 2007) and an *oprG* deletion mutant produced lower biofilm volumes as compared to the wild-type (Ritter et al., 2012).

Although we did not demonstrate causality, our correlation analysis suggests that the differential expression of these genes contribute to the observed low-biofilm phenotype in the ST258 isolates.

Discussion

The alarming increase of infections caused by multidrug-resistant pathogens poses a serious threat to human health, the global economy and to society in general (Cinel and Dellinger, 2007; Giske et al., 2008; World Health Organization, 2012). To successfully combat infections it is crucial to understand the molecular processes of bacterial adaptation to infection-relevant habitats; especially, since the effectiveness of proven antibiotics to control bacterial infections is diminishing and novel treatment-strategies and -targets are urgently needed. Over the past decade, there were multiple reports on the epidemic dissemination of multidrug-resistant Enterobacteriaceae, among which the carbapenemase-producing

organisms have been some of the most concerning (Nordmann et al., 2009). The spread of certain epidemic clones that are involved in numerous hospital outbreaks worldwide is emerging and a major concern (Woodford et al., 2011). The global success of these clonal lineages of multidrug-resistant pathogens is expected to be determined by a complex interplay between pathogenicity, epidemicity, and antibiotic resistance. While the impact of antibiotic resistance mechanisms on the global success of the bacterial pathogens seems to be established, other specific adaptive traits that may explain the success of epidemic high-risk clones remain unexplored (Chen et al., 2014b). Understanding the reasons for the success of these clones could be crucial for designing specific treatment and infection control strategies.

In this cross-sectional study, we took advantage of deep transcriptome sequencing (RNA-seq) to gain an unbiased view on the global transcriptional landscape of clinical *K. pneumoniae* isolates. Besides valuable transcriptomic information, RNA-seq allows for the extraction of the genomic sequences on the single nucleotide level of sufficiently transcribed genes, which, together with multilocus sequence typing (Diancourt et al., 2005) and capsular typing based on the *wzi* gene sequence (Brisse et al., 2013; D'Andrea et al., 2014), provided detailed insights into the phylogenetic relationship of the 37 clinical *K. pneumoniae* isolates studied in this work. Our phylogenetic data revealed that most of the clinical isolates belong to the two sequence types ST101 and ST258 - including the closely related isolates of ST512. Both groups comprise world-wide distributed, often multidrug-resistant epidemic clones with a high risk potential. Of note, while 12 out of 13 ST101 isolates in this study seemed to originate from a common reservoir, the ST258 isolates were less closely related and were obtained from various hospitals.

In this study, we recorded high-resolution transcriptome data of clinical *K. pneumoniae* isolates. We combined the transcriptome data with an analysis of infection relevant phenotypes such as of biofilm formation and virulence in a *Galleria* infection model and explored potential biological parameters that may explain the success of these high-risk clones. By using a pan-genome as a reference assembled from 11 previously published and publicly available genome sequences, we observed that a large set of 3346 genes was expressed in all isolates. These genes showed great overlap (75 %) with the core-genome consisting of 3336 genes and, like the core-genome, contains mostly of genes with central house-keeping functions. Most strikingly, we found clustering of isolates based on differential expression profiles of those commonly expressed genes. These results are in stark contrast to those obtained from *Pseudomonas aeruginosa* (Pohl et al., 2014) and *Escherichia coli* (Bielecki et al., 2014). In *E. coli* at least the phylogenetic background was clearly reflected in the overall gene expression profiles, which included accessory gene sets, however there were no subgroup-specific differences in core gene expression profiles (Bielecki et al., 2014). In *P. aeruginosa* even the acquisition of accessory genes seemed to be isolate specific and was independent of the phylogenetic background (Pohl et al., 2014).

Since *K. pneumoniae* ST258 isolates showed striking sequence type specific differences in their global expression profiles, we sought to correlate the expression profiles with the infection relevant phenotypes of virulence and biofilm formation. In a *Galleria* infection model, virulence has been shown to correlate to resistance against human serum (Diago-Navarro et al., 2014) and was linked to the occurrence of the K1 and K2 capsular type (Insua et al., 2013). None of our isolates were of K1 or K2 type. Nevertheless, we observed differences in the virulence of the isolates between the various sequence types, but also within isolates of the same sequence type. These results clearly indicate that virulence - as

measured under the chosen experimental conditions in the *Galleria* infection model - is not associated with a particular sequence type. The same variation in virulence was previously described for ST258 isolates containing the *wzi*-154 allele (which is also shared by ST512 isolates) in a *Galleria* infection model, macrophage killing assay and human serum resistance assay (Diago-Navarro et al., 2014). A recent publication demonstrated an association of the expression of the regulator of the mucoid phenotype encoded by *rmpA*, the expression of siderophores and further iron metabolism genes with community-acquired invasive infections in humans (Holt et al., 2015).

We found an overall low expression of virulence associated genes in our clinical isolates and were unable to correlate their expression levels with a virulence phenotype. The analysis of *ex vivo* transcriptomes recorded during the course of infection (Toledo-Arana et al., 2009; Mandlik et al., 2011; Bielecki et al., 2014; Szafranska et al., 2014) might provide more valuable information on how this pathogen adapts to the host and whether this adaptation might differ between the various sequence types.

Biofilm formation experiments showed that all ST258 isolates formed only poor biofilms, while in most of the other clinical strains biofilm formation was more common (Naparstek et al., 2014). Classifying the strains into biofilm-proficient and -deficient isolates revealed differential expression of 90 genes, some of which have previously been linked to biofilm formation; 41 of those belonged to the core-transcriptome. Our results on the low expression of biofilm-associated genes in the ST258 isolates imply that the high-risk clone ST258 has adopted poor biofilm formation as a common trait, and this trait is determined by distinct changes in the transcriptome. This finding might be unexpected since the capability to form biofilms has been associated with more successful survival in the clinical environment (Vickery et al., 2012). Nevertheless, the discovery that the ST258 isolates share

a non-biofilm-specific gene expression profile within their core-transcriptome strongly implicates that this trait confers to an advantageous phenotype that might be linked to the establishment and maintenance of an infection within the human host but it might also be linked to a more successful spreading or survival of this clonal lineage in the hospital setting. Further studies are inevitable to explore a possible correlation of the incidence of the ST258 *K. pneumoniae* sequence type with distinct patient or hospital environmental settings that might privilege biofilm-deficient isolates.

Materials and Methods

Bacterial isolates

Clinical *K. pneumoniae* isolates were sampled from ten different German hospitals (n=31; indicated with A-J in supplementary table S1) and one medical practice (Dr. Schanz, Salzgitter, Germany, n=3). Three already published *K. pneumoniae* isolates (MGH 78578 (http://www.ncbi.nlm.nih.gov/nuccore/NC_009648.1), JH1 and 1162281 (Kumar et al., 2011)) were also included in this study.

Bacterial culture conditions, RNA extraction and Illumina-based RNA sequencing

Bacteria were cultured in Luria–Bertani (LB) at 37 °C and vigorous shaking with 180 rpm until late exponential growth phase to an optical density at 600 nm (OD₆₀₀) of 2.0. RNA extraction and RNA-seq library preparation were performed as previously described (Dotsch et al., 2012) with an additional step of tobacco acid pyrophosphatase (TAP, epicenter) treatment to convert 5'-triphosphate RNA into 5'-monophosphate RNA before adapter ligation. Messenger

RNAs were enriched after RNA extraction using the MICROBExpress™ bacterial mRNA enrichment kit (Ambion) according to the manufacturer's instructions. Up to 18 libraries were pooled and treated with duplex-specific nuclease (DSN) for additional rRNA removal. Sequencing was performed on an Illumina HiSeq 2500 generating paired-end reads each of 100 base pairs length, resulting in 13.1 to 36.6 million reads per isolate (with a median of 20 million reads per isolate).

Generation of the *K. pneumoniae* pan-genome

In this study we used 11 fully sequenced and annotated *K. pneumoniae* strains whose sequence information is publicly available in GenBank/EMBL. These 11 *K. pneumoniae* genomes contain 57,312 genes in total, with an average of 5,210 genes per strain (for more detailed information see table 1).

With the aim to generate a non-redundant gene set, all gene sequences were blasted against each other using BLASTN, discarding hits having < 90 % length and 90 % sequence identity. Only genes having reciprocal homologs in all 11 genomes were considered as “core”; otherwise, they were classified as “accessory”. Accessory genes having homologs in 8 to 10 out of the 11 *K. pneumoniae* genomes were manually re-evaluated. The set of core genes detected in the reciprocal blast search comprised 3,270 genes with additional 41 genes assigned to the core-genome after re-evaluation. Furthermore, small RNAs in *K. pneumoniae* subsp. MGH 78578 were predicted by using sRNAsScanner (Sridhar et al., 2010) with default parameters. A total of 30 sRNAs were included in the final gene set. To generate a single reference sequence file, sequences of core-genom genes were selected from the sequence type 23 strain “1084” (Lin et al., 2012) and the accessory genes and the singletons were added from each of the remaining 10 genomes considered in this study. The origin of each

sequence is indicated by the locus ID in tables S1 (see also table 1). The pan-genome was concatenated by adding the 50 bp long genomic sequence upstream- and downstream of each gene and a separator of 100 “N”s between all genes. This resulted in a total pan-genome of 7.92 Mbp, of which 6.35 Mbp were coding sequences. The non-redundant pan-genome had a size of 7589 genes with 3336 genes belonging to the core-genome (i.e. genes occurring in every genome) and 1598 unique genes (singletons). Supplementary figure S1 shows the development of the size of the pan-genome, core-genome and the singletons in all possible combinations of the 11 genomes. Based on the averages, it shows an exponential expansion for all 3 groups with the following formulae: pan-genome: $f(x) = 4958.2 x^{0.1807}$, $R^2=1.0$; core-genome: $f(x) = 4891.5 x^{0.16}$, $R^2=0.998$ and singletons: $f(x) = 1135.7 x^{0.1481}$, $R^2=0.996$). With x representing the number of completely sequenced genomes.

Mapping and gene expression profiling

Raw sequencing reads were de-multiplexed and trimmed by a custom Perl script, adapter and barcode sequences were removed using the fastq-mcf script included in the ea-utils package (<https://code.google.com/p/ea-utils/>). Reads with more than one mismatch in the adapter sequence were discarded. Processed reads were mapped against the *K. pneumoniae* pan-genome with stampy using bwaoptions -q10 (Lunter and Goodson, 2011). Absolute quantification of gene expression was performed as described by Dötsch *et al.* (Dotsch et al., 2012). Read counts were normalized to generate normalized reads per kilobase of gene sequence (nRPK (Dotsch et al., 2012)) values according to the following equation:

$$nRPK = \log_2 \left(\frac{\frac{1000}{l_i} * RPG}{F_j} + 1 \right) \quad \text{where } l_i \text{ is the length in bp of gene } i, RPG \text{ is the absolute}$$

count of reads of gene i and F_j is the size factor calculated by DESeq (Anders and Huber,

2010) of isolate *j*. This normalization method delivers more robust data as e.g. RPKM when analyzing highly expressed genes (Dotsch et al., 2012).

A gene was considered as expressed when the corresponding nRPK value was higher than the threshold $nRPK_0$ (Dotsch et al., 2015), whereby $nRPK_0$ is defined as:

$$nRPK_0 = \log_2 \left(\frac{\frac{1000}{l_{med}} * 1}{F_{min}} + 1 \right) = \log_2 \left(\frac{\frac{1000}{700} * 1}{0.398} + 1 \right) \approx 2.20 \quad \text{where } l_{med} \text{ is the median length in}$$

bp of all genes in the pan-genome and F_{min} is the smallest size factor of all RNA-seq libraries (i.e. the lowest sequencing depth) as determined in DESeq. Gene expression profiles were further analyzed in R using the packages scatterplot3D and gplots.

The variation in gene expression was analyzed in R with DESeq using pairwise comparisons in all possible combinations of two isolates, resulting in 666 comparisons

$$\left(\sum_{K=1}^n k, \text{ with } n=36 \text{ being the number of comparisons for one isolate} \right) \quad \text{In these pairwise}$$

comparisons, a gene was regarded as differentially expressed when the absolute value of log2 fold change was above one. The value of variation is expressed as the fraction of comparisons in which the gene is differentially regulated vs all comparisons for that specific gene. Genes lower as the 10th percentile or higher as the 90th percentile were regarded as very stable and highly variable expressed, respectively.

***De novo* assembly of accessory genes**

All reads not mapping to the *K. pneumoniae* pan-genome were used as input for a *de novo* transcriptome assembly with OASES (Schulz et al., 2012). After testing k-mers from 17 to 41, the assembly was performed with k-mer 33 and a minimal transcript length of 250 bp. The

assembled transcripts were blasted against the microbial genome database (Uchiyama, 2003; Uchiyama et al., 2013) using a minimal hit length of 100 bp and a minimum sequence similarity of 70 %. To generate a non-redundant list, the sequences of all positive hits were extracted and blasted against each other, discarding hits having < 80 % length and 80 % sequence identity. All unmapped reads were mapped against this list using stampy. Only genes with coverage of at least 70 % were regarded as true positive hits resulting in a list of 1482 genes. To screen for the presence of known virulence associated transcripts, the assemblies were compared against the virulence factors database (VFDB) (Chen et al., 2012) as well as the virulence database from *K. pneumoniae* BIGSdb (Bialek-Davenet et al., 2014) using BLASTn.

Phylogenetic relationship

For each isolate a consensus sequence was generated using of the *mpileup* option in the SAMtools package (Li et al., 2009). Only protein coding sequences from the core-genome transcribed in every isolate with minimal coverage of 5 reads per position were used. The coverage was determined with the “depth” command in SAMtools and the sequences of 404 genes fulfilling our criteria were extracted using a Perl script. We extracted the corresponding orthologous gene sequences from 11 publicly available *K. pneumoniae* genomes and included them in the alignment generation. All sequences were aligned with Clustal Omega (Sievers et al., 2011; Sievers and Higgins, 2014) to generate an alignment of 314,561 bp length. All positions containing gaps and missing data were eliminated. The consensus Neighbor Joining tree was drawn with MEGA 6.0 (Tamura et al., 2013) using the maximum composite likelihood model and 1000 bootstrapping replications.

Gene Ontology term enrichment

We downloaded the current UniProt Gene Ontology (GO) knowledge database (as of 07/09/14) (Dimmer et al., 2012). Using custom Perl scripts, we mapped the gene locus IDs (in KEGG format) to their UniProt identifiers and extracted the relevant GO IDs from the flat files. Significantly enriched or depleted categories were tested by one-sided hypergeometric tests in R and the retrieved P-values were adjusted by the Benjamini-Hochberg correction to control the false-discovery rate (FDR).

Transcriptome-wide association of differentially expressed genes for infection relevant phenotypes

Bacterial isolates were grouped according to a particular phenotype (biofilm or virulence) and compared to identify differentially expressed genes in one group vs the other group using Wilcoxon's rank-sum test. P-values were adjusted by the Benjamini-Hochberg correction to control the false-discovery rate (FDR) of potential hits. Genes were regarded as differentially expressed, when their median expression differed by at least 2 nRPK, the standard deviation was not higher than the median and their adjusted P-value was smaller than 0.05. Since nRPK values are log2-transformed (see "Mapping and gene expression profiling"), a difference of 2 nRPK corresponds to a 4-fold change in the gene expression.

Multilocus sequence typing (MLST)

To perform MLST (Diancourt et al., 2005), sequences from *gapA*, *infB*, *mdh*, *pgi* and *rpoB* were extracted from RNA-seq by SAMtools' *mpileup* option (Li et al., 2009). Due to low coverage, sequencing of *phoE* and *tonB* was performed by classical Sanger sequencing according to (Brisse et al., 2009) using universal sequence primers and an annealing temperature of 50 °C. The following primers were used: *phoE*:F:604.1:oF

(GTTTTCCCAGTCACGACGTTGTAACTACCGCAACACCGACTTCTTCGG), phoE:R:604.2:oR
 (TTGTGAGCGGATAACAATTTCTGATCAGAACTGGTAGGTGAT), tonB:1F:oF
 (GTTTTCCCAGTCACGACGTTGTACTTTATACCTCGGTACATCAGGTT), tonB:2R:oR
 (TTGTGAGCGGATAACAATTTTCATTCGCCGGCTGRGCRGAGAG), ForwardSeq:oF
 (GTTTTCCCAGTCACGACGTTGTA), ReverseSeq:oR (TTGTGAGCGGATAACAATTTTC). Sequences
 were analyzed via <http://bigsdb.web.pasteur.fr/klebsiella/klebsiella.html>

wzi typing

A fragment of *wzi* was amplified by primers *wzi_for2* (GTG CCG CGA GCG CTT TCT ATC TTG GTA TTC C) and *wzi_rev* (GAG AGC CAC TGG TTC CAG AAY TTS ACC GC) adapted from (Brisse et al., 2013) and Sanger sequenced with primer *wzi_for2*. Sequences were analyzed and compared to published results via <http://bigsdb.web.pasteur.fr/klebsiella/klebsiella.html>.

KPC sequencing

An 872 bp long fragment from the *Klebsiella pneumoniae* carbapenemase coding gene *bla_{KPC}* was amplified with primers KPC_FP (ATG TCA CTG TAT CGC CGT CT) and KPC_RP (TTG ACG CCC AAT CCC TCG) from (McLaughlin et al., 2014) with an annealing temperature of 51 °C. Nucleotide sequences were obtained by Sanger sequencing using both primers and compared to known *bla_{KPC}* alleles in MEGA 6.0 (Tamura et al., 2013).

***Galleria mellonella* virulence assay**

Protocols for the *Galleria mellonella* virulence assay were adjusted from previous publications (Pustelny et al., 2013; McLaughlin et al., 2014). Bacterial isolates were grown in LB until mid-exponential growth phase. Cells were harvested at 8.000 rpm for 5 min and washed with sterile phosphate-buffered saline (10 mM PBS, pH 7.5). Cultures were adjusted

to an OD₆₀₀ of 0.5 and 10-fold serially diluted in PBS. For each isolate, ten randomly chosen, healthy *Galleria mellonella* larvae (fauna topics GmbH, Marbach, Germany) were inoculated with 20 µl of a 1:10 dilution containing app. 8×10^5 colony forming units (cfu) by injection into the haemocoel of the rear left proleg with a 100 µl Hamilton syringe and a 30-G needle. The larvae were placed in sterile Petri dishes and incubated in the dark at 37 °C. Mortality was monitored for 72 h. Larval death was assessed by the lack of movement of larvae in response to physical stimulation with a sterile inoculating loop. Ten larvae inoculated with 20 µl PBS and 10 larvae without any treatment were used as negative controls. Each experiment was performed in duplicates on different days. Experiments were repeated if more than one larva died in the controls. Colony counts were performed by plating appropriate 10-fold dilutions on LB agar plates.

Measurement of bacterial growth

Planktonic growth of *K. pneumoniae* isolates was monitored in LB, BM2 minimal medium (Overhage et al., 2008) [62 mM potassium phosphate buffer, pH 7, 7 mM (NH₄)₂SO₄, 2 mM MgSO₄, 10 µM FeSO₄] supplemented with 50mM glucose and BM2 supplemented with 50 mM mannose, respectively using an automated growth analysis system (Bioscreen C MBR, Oy Growth Curves Ab Ltd, Helsinki, Finland). Cells of overnight cultures grown in LB at 37 °C were harvested at 8.000 rpm for 5 min and washed with the desired medium. Cultures were diluted to an OD₆₀₀ of 0.02 and 200 µl were transferred into a 100-well honeycomb plate (four replicates per isolate). The plates were incubated at 37 °C with continuous shaking and 15 min measurement intervals for a period of 24 h. Maximal optical density and doubling time were determined using GrowthRates (Hall et al., 2014).

Crystal violet biofilm assay

Protocols to measure biofilm formation were adapted from previous publications (Maldonado N.C., 2007; Naparstek et al., 2014). An overnight culture grown in LB at 37 °C and 180 rpm was washed and diluted to an OD₆₀₀ of 0.02 in fresh LB. 8 wells of a 96-well microtiter plate (BD biosciences, Heidelberg, Germany) were inoculated with 100 µl and the plate was incubated at 37 °C in a humid atmosphere. After 24 h wells were washed three times with 200 µl sterile water and stained with 150 µl crystal violet staining solution (0.1 % m/v in water) for 30 min. Wells were washed three times with 200 µl water, crystal violet was extracted with 200 µl 95% ethanol for 30 min and absorbance was measured at 590 nm in an EnSpire Multimode Plate Reader (PerkinElmer, Waltham, Massachusetts USA). All experiments were performed with 8 individual repeats per measurement. Isolates were regarded as high biofilm producers if their OD₅₉₀ value was three times the OD₅₉₀ value of the negative control (Mordhorst et al., 2009).

Nucleotide sequence accession number

The RNA-seq data has been deposited at the National Center for Biotechnology Information sequence read archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under the accession no. SRP051240.

Acknowledgements

We acknowledge Klaus Hornischer for data management and Bianca Nouri for technical support. The *Klebsiella pneumoniae* strain MGH 78578 was a kind gift of Michael McClelland (University of California, Irvine, USA), strains JH1 and 1162281 were kind gifts of James R. Brown (GlaxoSmithKline, Collegeville, Pennsylvania, USA). We thank Robert Geffers and the

Genome Analytics Research Group (Helmholtz Centre for Infection Research, Braunschweig, Germany) for performing the Illumina sequencing.

This work was supported by a starter grant (RESISTOME 260276) from the European Research Council and by the President's Initiative and Networking Fund of the Helmholtz Association of German Research Centers (HGF) under contract number VH-GS-202. The funding agencies had no role in design, collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

References

- Alekshun, M.N., and Levy, S.B. (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128**: 1037-1050.
- Alghoribi, M.F., Gibreel, T.M., Dodgson, A.R., Beatson, S.A., and Upton, M. (2014) *Galleria mellonella* infection model demonstrates high lethality of ST69 and ST127 uropathogenic *E. coli*. *PLoS One* **9**: e101547.
- Allen, B.L., Gerlach, G.F., and Clegg, S. (1991) Nucleotide sequence and functions of mrk determinants necessary for expression of type 3 fimbriae in *Klebsiella pneumoniae*. *J Bacteriol* **173**: 916-920.
- Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.
- Bialek-Davenet, S., Criscuolo, A., Ailloud, F., Passet, V., Jones, L., Delannoy-Vieillard, A.S. et al. (2014) Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. *Emerg Infect Dis* **20**: 1812-1820.
- Bielecki, P., Muthukumarasamy, U., Eckweiler, D., Bielecka, A., Pohl, S., Schanz, A. et al. (2014) In vivo mRNA profiling of uropathogenic *Escherichia coli* from diverse phylogroups reveals common and group-specific gene expression profiles. *MBio* **5**: e01075-01014.
- Borer, A., Saidel-Odes, L., Riesenberger, K., Eskira, S., Peled, N., Nativ, R. et al. (2009) Attributable mortality rate for carbapenem-resistant *Klebsiella pneumoniae* bacteremia. *Infect Control Hosp Epidemiol* **30**: 972-976.
- Breurec, S., Guessennd, N., Timinouni, M., Le, T.A., Cao, V., Ngandjio, A. et al. (2013) *Klebsiella pneumoniae* resistant to third-generation cephalosporins in five African and two Vietnamese major towns: multiclonal population structure

with two major international clonal groups, CG15 and CG258. *Clin Microbiol Infect* **19**: 349-355.

Brisse, S., Fevre, C., Passet, V., Issenhuth-Jeanjean, S., Tournebize, R., Diancourt, L., and Grimont, P. (2009) Virulent clones of *Klebsiella pneumoniae*: identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS One* **4**: e4982.

Brisse, S., Passet, V., Haugaard, A.B., Babosan, A., Kassis-Chikhani, N., Struve, C., and Decre, D. (2013) wzi Gene sequencing, a rapid method for determination of capsular type for *Klebsiella* strains. *J Clin Microbiol* **51**: 4073-4078.

Caldelari, I., Chao, Y., Romby, P., and Vogel, J. (2013) RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harb Perspect Med* **3**: a010298.

Chen, L., Xiong, Z., Sun, L., Yang, J., and Jin, Q. (2012) VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res* **40**: D641-645.

Chen, L., Mathema, B., Pitout, J.D., DeLeo, F.R., and Kreiswirth, B.N. (2014a) Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. *MBio* **5**: e01355-01314.

Chen, L., Mathema, B., Chavda, K.D., DeLeo, F.R., Bonomo, R.A., and Kreiswirth, B.N. (2014b) Carbapenemase-producing *Klebsiella pneumoniae*: molecular and genetic decoding. *Trends Microbiol* **22**: 686-696.

Cinel, I., and Dellinger, R.P. (2007) Advances in pathogenesis and management of sepsis. *Curr Opin Infect Dis* **20**: 345-352.

Cooper, T.F., Rozen, D.E., and Lenski, R.E. (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichiacoli*. *Proc Natl Acad Sci U S A* **100**: 1072-1077.

Croucher, N.J., and Thomson, N.R. (2010) Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* **13**: 619-624.

D'Andrea, M.M., Amisano, F., Giani, T., Conte, V., Ciacci, N., Ambretti, S. et al. (2014) Diversity of capsular polysaccharide gene clusters in Kpc-producing *Klebsiella pneumoniae* clinical isolates of sequence type 258 involved in the Italian epidemic. *PLoS One* **9**: e96827.

Daikos, G.L., and Markogiannakis, A. (2011) Carbapenemase-producing *Klebsiella pneumoniae*: (when) might we still consider treating with carbapenems? *Clin Microbiol Infect* **17**: 1135-1141.

Damjanova, I., Toth, A., Paszti, J., Hajbel-Vekony, G., Jakab, M., Berta, J. et al. (2008) Expansion and countrywide dissemination of ST11, ST15 and ST147 ciprofloxacin-resistant CTX-M-15-type beta-lactamase-producing *Klebsiella pneumoniae* epidemic clones in Hungary in 2005--the new 'MRSA's'? *J Antimicrob Chemother* **62**: 978-985.

Darmon, E., and Leach, D.R. (2014) Bacterial genome instability. *Microbiol Mol Biol Rev* **78**: 1-39.

- Dashper, S., Ang, C.S., Liu, S.W., Paolini, R., Veith, P., and Reynolds, E. (2010) Inhibition of *Porphyromonas gingivalis* biofilm by oxantel. *Antimicrob Agents Chemother* **54**: 1311-1314.
- De Lay, N., and Gottesman, S. (2009) The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J Bacteriol* **191**: 461-476.
- Deleo, F.R., Chen, L., Porcella, S.F., Martens, C.A., Kobayashi, S.D., Porter, A.R. et al. (2014) Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc Natl Acad Sci U S A* **111**: 4988-4993.
- Diago-Navarro, E., Chen, L., Passet, V., Burack, S., Ulacia-Hernando, A., Kodiyanplakkal, R.P. et al. (2014) Carbapenem-resistant *Klebsiella pneumoniae* exhibit variability in capsular polysaccharide and capsule associated virulence traits. *J Infect Dis* **210**: 803-813.
- Diancourt, L., Passet, V., Verhoef, J., Grimont, P.A., and Brisse, S. (2005) Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J Clin Microbiol* **43**: 4178-4182.
- Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J. et al. (2012) The UniProt-GO Annotation database in 2011. *Nucleic Acids Res* **40**: D565-570.
- Dotsch, A., Eckweiler, D., Schniederjans, M., Zimmermann, A., Jensen, V., Scharfe, M. et al. (2012) The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. *PLoS One* **7**: e31092.
- Dotsch, A., Schniederjans, M., Khaledi, A., Hornischer, K., Schulz, S., Bielecka, A. et al. (2015) The *Pseudomonas aeruginosa* Transcriptional Landscape Is Shaped by Environmental Heterogeneity and Genetic Variation. *MBio* **6**.
- Giannouli, M., Palatucci, A.T., Rubino, V., Ruggiero, G., Romano, M., Triassi, M. et al. (2014) Use of larvae of the wax moth *Galleria mellonella* as an in vivo model to study the virulence of *Helicobacter pylori*. *BMC Microbiol* **14**: 228.
- Giske, C.G., Monnet, D.L., Cars, O., Carmeli, Y., and ReAct-Action on Antibiotic, R. (2008) Clinical and economic impact of common multidrug-resistant gram-negative bacilli. *Antimicrob Agents Chemother* **52**: 813-821.
- Grundmann, H., Livermore, D.M., Giske, C.G., Canton, R., Rossolini, G.M., Campos, J. et al. (2010) Carbapenem-non-susceptible Enterobacteriaceae in Europe: conclusions from a meeting of national experts. *Euro Surveill* **15**.
- Hall, B.G., Acar, H., Nandipati, A., and Barlow, M. (2014) Growth rates made easy. *Mol Biol Evol* **31**: 232-238.
- Holt, K.E., Wertheim, H., Zadoks, R.N., Baker, S., Whitehouse, C.A., Dance, D. et al. (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.

Huang, Y.J., Liao, H.W., Wu, C.C., and Peng, H.L. (2009) MrkF is a component of type 3 fimbriae in *Klebsiella pneumoniae*. *Res Microbiol* **160**: 71-79.

Hudson, C.M., Bent, Z.W., Meagher, R.J., and Williams, K.P. (2014) Resistance determinants and mobile genetic elements of an NDM-1-encoding *Klebsiella pneumoniae* strain. *PLoS One* **9**: e99209.

Insua, J.L., Llobet, E., Moranta, D., Perez-Gutierrez, C., Tomas, A., Garmendia, J., and Bengoechea, J.A. (2013) Modeling *Klebsiella pneumoniae* pathogenesis by infection of the wax moth *Galleria mellonella*. *Infect Immun* **81**: 3552-3565.

Jagnow, J., and Clegg, S. (2003) *Klebsiella pneumoniae* MrkD-mediated biofilm formation on extracellular matrix- and collagen-coated surfaces. *Microbiology* **149**: 2397-2405.

Kim, D., Hong, J.S., Qiu, Y., Nagarajan, H., Seo, J.H., Cho, B.K. et al. (2012) Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genet* **8**: e1002867.

Kitchel, B., Rasheed, J.K., Patel, J.B., Srinivasan, A., Navon-Venezia, S., Carmeli, Y. et al. (2009) Molecular epidemiology of KPC-producing *Klebsiella pneumoniae* isolates in the United States: clonal expansion of multilocus sequence type 258. *Antimicrob Agents Chemother* **53**: 3365-3370.

Kumar, V., Sun, P., Vamathevan, J., Li, Y., Ingraham, K., Palmer, L. et al. (2011) Comparative genomics of *Klebsiella pneumoniae* strains with different antibiotic resistance profiles. *Antimicrob Agents Chemother* **55**: 4267-4276.

Langstraat, J., Bohse, M., and Clegg, S. (2001) Type 3 fimbrial shaft (MrkA) of *Klebsiella pneumoniae*, but not the fimbrial adhesin (MrkD), facilitates biofilm formation. *Infect Immun* **69**: 5805-5812.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, J., Liu, F., Wang, Q., Ge, P., Woo, P.C., Yan, J. et al. (2014) Genomic and transcriptomic analysis of NDM-1 *Klebsiella pneumoniae* in spaceflight reveal mechanisms underlying environmental adaptability. *Sci Rep* **4**: 6216.

Liapis, E., Pantel, A., Robert, J., Nicolas-Chanoine, M.H., Cavalie, L., van der Mee-Marquet, N. et al. (2014) Molecular epidemiology of OXA-48-producing *Klebsiella pneumoniae* in France. *Clin Microbiol Infect*.

Lin, A.C., Liao, T.L., Lin, Y.C., Lai, Y.C., Lu, M.C., and Chen, Y.T. (2012) Complete genome sequence of *Klebsiella pneumoniae* 1084, a hypermucoviscosity-negative K1 clinical strain. *J Bacteriol* **194**: 6316.

Liu, P., Li, P., Jiang, X., Bi, D., Xie, Y., Tai, C. et al. (2012) Complete genome sequence of *Klebsiella pneumoniae* subsp. *pneumoniae* HS11286, a multidrug-resistant strain isolated from human sputum. *J Bacteriol* **194**: 1841-1842.

Livermore, D.M., Hope, R., Brick, G., Lillie, M., Reynolds, R., and Surveillance, B.W.P.o.R. (2008) Non-susceptibility trends among Enterobacteriaceae from bacteraemias in the UK and Ireland, 2001-06. *J Antimicrob Chemother* **62 Suppl 2**: ii41-54.

Lunter, G., and Goodson, M. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* **21**: 936-939.

Maldonado N.C., S.d.R.C., Cecilia M. and Nader-Macias M.E. (2007) A simple technique to detect *Klebsiella* biofilm-forming-strains. Inhibitory potential of *Lactobacillus fermentum* CRL 1058 whole cells and products. In *Communicating Current Research and Educational Topics and Trends in Applied Microbiology*. Mendez-Vilas, A. (ed): FORMATEX, pp. 52-59.

Mandlik, A., Livny, J., Robins, W.P., Ritchie, J.M., Mekalanos, J.J., and Waldor, M.K. (2011) RNA-Seq-based monitoring of infection-linked changes in *Vibrio cholerae* gene expression. *Cell Host Microbe* **10**: 165-174.

Marchaim, D., Navon-Venezia, S., Schwaber, M.J., and Carmeli, Y. (2008) Isolation of imipenem-resistant *Enterobacter* species: emergence of KPC-2 carbapenemase, molecular characterization, epidemiology, and outcomes. *Antimicrob Agents Chemother* **52**: 1413-1418.

McLaughlin, M.M., Advincula, M.R., Malczynski, M., Barajas, G., Qi, C., and Scheetz, M.H. (2014) Quantifying the clinical virulence of *Klebsiella pneumoniae* producing carbapenemase *Klebsiella pneumoniae* with a *Galleria mellonella* model and a pilot study to translate to patient outcomes. *BMC Infect Dis* **14**: 31.

Mikkelsen, H., Duck, Z., Lilley, K.S., and Welch, M. (2007) Interrelationships between colonies, biofilms, and planktonic cells of *Pseudomonas aeruginosa*. *J Bacteriol* **189**: 2411-2416.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J* **34**: 1478-1492.

Mordhorst, I.L., Claus, H., Ewers, C., Lappann, M., Schoen, C., Elias, J. et al. (2009) O-acetyltransferase gene *neuO* is segregated according to phylogenetic background and contributes to environmental desiccation resistance in *Escherichia coli* K1. *Environ Microbiol* **11**: 3154-3165.

Morgan, J.L., McNamara, J.T., and Zimmer, J. (2014) Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat Struct Mol Biol* **21**: 489-496.

Naparstek, L., Carmeli, Y., Navon-Venezia, S., and Banin, E. (2014) Biofilm formation and susceptibility to gentamicin and colistin of extremely drug-

resistant KPC-producing *Klebsiella pneumoniae*. *J Antimicrob Chemother* **69**: 1027-1034.

Nordmann, P., Cuzon, G., and Naas, T. (2009) The real threat of *Klebsiella pneumoniae* carbapenemase-producing bacteria. *Lancet Infect Dis* **9**: 228-236.

Ogawa, W., Li, D.W., Yu, P., Begum, A., Mizushima, T., Kuroda, T., and Tsuchiya, T. (2005) Multidrug resistance in *Klebsiella pneumoniae* MGH78578 and cloning of genes responsible for the resistance. *Biol Pharm Bull* **28**: 1505-1508.

Overhage, J., Bains, M., Brazas, M.D., and Hancock, R.E. (2008) Swarming of *Pseudomonas aeruginosa* is a complex adaptation leading to increased production of virulence factors and antibiotic resistance. *J Bacteriol* **190**: 2671-2679.

Papenfort, K., and Vogel, J. (2010) Regulatory RNA in bacterial pathogens. *Cell Host Microbe* **8**: 116-127.

Papenfort, K., Pfeiffer, V., Lucchini, S., Sonawane, A., Hinton, J.C., and Vogel, J. (2008) Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Mol Microbiol* **68**: 890-906.

Pitart, C., Sole, M., Roca, I., Fabrega, A., Vila, J., and Marco, F. (2011) First outbreak of a plasmid-mediated carbapenem-hydrolyzing OXA-48 beta-lactamase in *Klebsiella pneumoniae* in Spain. *Antimicrob Agents Chemother* **55**: 4398-4401.

Podschun, R., and Ullmann, U. (1998) *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* **11**: 589-603.

Pohl, S., Klockgether, J., Eckweiler, D., Khaledi, A., Schniederjans, M., Chouvarine, P. et al. (2014) The extensive set of accessory *Pseudomonas aeruginosa* genomic components. *FEMS Microbiol Lett* **356**: 235-241.

Potron, A., Poirel, L., Rondinaud, E., and Nordmann, P. (2013) Intercontinental spread of OXA-48 beta-lactamase-producing Enterobacteriaceae over a 11-year period, 2001 to 2011. *Euro Surveill* **18**.

Poulou, A., Voulgari, E., Vrioni, G., Koumaki, V., Xidopoulos, G., Chatzipantazi, V. et al. (2013) Outbreak caused by an ertapenem-resistant, CTX-M-15-producing *Klebsiella pneumoniae* sequence type 101 clone carrying an OmpK36 porin variant. *J Clin Microbiol* **51**: 3176-3182.

Pustelny, C., Brouwer, S., Musken, M., Bielecka, A., Dotsch, A., Nimtz, M., and Haussler, S. (2013) The peptide chain release factor methyltransferase PrmC is essential for pathogenicity and environmental adaptation of *Pseudomonas aeruginosa* PA14. *Environ Microbiol* **15**: 597-609.

Ramos, P.I., Picao, R.C., Almeida, L.G., Lima, N.C., Girardello, R., Vivan, A.C. et al. (2014) Comparative analysis of the complete genome of KPC-2-producing

Klebsiella pneumoniae Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. *BMC Genomics* **15**: 54.

Rapp, R.P., and Urban, C. (2012) Klebsiella pneumoniae carbapenemases in Enterobacteriaceae: history, evolution, and microbiology concerns. *Pharmacotherapy* **32**: 399-407.

Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E., and Gorke, B. (2008) The small RNA GlmY acts upstream of the sRNA GlmZ in the activation of glmS expression and is subject to regulation by polyadenylation in Escherichia coli. *Nucleic Acids Res* **36**: 2570-2580.

Rice, L.B. (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J Infect Dis* **197**: 1079-1081.

Ritter, A., Com, E., Bazire, A., Goncalves Mdos, S., Delage, L., Le Pennec, G. et al. (2012) Proteomic studies highlight outer-membrane proteins related to biofilm development in the marine bacterium Pseudoalteromonas sp. D41. *Proteomics* **12**: 3180-3192.

Rodrigues, C., Machado, E., Ramos, H., Peixe, L., and Novais, A. (2014) Expansion of ESBL-producing Klebsiella pneumoniae in hospitalized patients: A successful story of international clones (ST15, ST147, ST336) and epidemic plasmids (IncR, IncFII). *Int J Med Microbiol*.

Romling, U., Galperin, M.Y., and Gomelsky, M. (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**: 1-52.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**: 1086-1092.

Seo, J.H., Hong, J.S., Kim, D., Cho, B.K., Huang, T.W., Tsai, S.F. et al. (2012) Multiple-omic data analysis of Klebsiella pneumoniae MGH 78578 reveals its transcriptional architecture and regulatory features. *BMC Genomics* **13**: 679.

Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev* **21**: 2804-2817.

Shin, S.H., Kim, S., Kim, J.Y., Lee, S., Um, Y., Oh, M.K. et al. (2012) Complete genome sequence of the 2,3-butanediol-producing Klebsiella pneumoniae strain KCTC 2242. *J Bacteriol* **194**: 2736-2737.

Sievers, F., and Higgins, D.G. (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**: 105-116.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

Sorek, R., and Cossart, P. (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**: 9-16.

Sridhar, J., Sambaturu, N., Sabarinathan, R., Ou, H.Y., Deng, Z., Sekar, K. et al. (2010) sRNAsScanner: a computational tool for intergenic small RNA detection in bacterial genomes. *PLoS One* **5**: e11970.

Stoesser, N., Giess, A., Batty, E.M., Sheppard, A.E., Walker, A.S., Wilson, D.J. et al. (2014) Genome sequencing of an extended series of NDM-producing *Klebsiella pneumoniae* isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. *Antimicrob Agents Chemother* **58**: 7347-7357.

Szafranska, A.K., Oxley, A.P., Chaves-Moreno, D., Horst, S.A., Rosslenbroich, S., Peters, G. et al. (2014) High-Resolution Transcriptomic Analysis of the Adaptive Response of *Staphylococcus aureus* during Acute and Chronic Phases of Osteomyelitis. *MBio* **5**.

Tamura, K., Stecher, G., Peterson, D., Filipinski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33-36.

Tavares, L.S., Silva, C.S., de Souza, V.C., da Silva, V.L., Diniz, C.G., and Santos, M.O. (2013) Strategies and molecular tools to fight antimicrobial resistance: resistome, transcriptome, and antimicrobial peptides. *Front Microbiol* **4**: 412.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D. et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950-956.

Uchiyama, I. (2003) MBGD: microbial genome database for comparative analysis. *Nucleic Acids Res* **31**: 58-62.

Uchiyama, I., Mihara, M., Nishide, H., and Chiba, H. (2013) MBGD update 2013: the microbial genome database for exploring the diversity of microbial world. *Nucleic Acids Res* **41**: D631-635.

Urbanowski, M.L., Stauffer, L.T., and Stauffer, G.V. (2000) The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol Microbiol* **37**: 856-868.

van Vliet, A.H. (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol Lett* **302**: 1-7.

Vickery, K., Deva, A., Jacombs, A., Allan, J., Valente, P., and Gosbell, I.B. (2012) Presence of biofilm containing viable multiresistant organisms despite terminal cleaning on clinical surfaces in an intensive care unit. *J Hosp Infect* **80**: 52-55.

Walsh, C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406**: 775-781.

Walsh, C. (2003) *Antibiotics: Actions, Origins, Resistance*: ASM Press.

Wand, M.E., McCowen, J.W., Nugent, P.G., and Sutton, J.M. (2013) Complex interactions of *Klebsiella pneumoniae* with the host immune system in a *Galleria mellonella* infection model. *J Med Microbiol* **62**: 1790-1798.

Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63.

Woodford, N., Turton, J.F., and Livermore, D.M. (2011) Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev* **35**: 736-755.

World Health Organization (2012) *The evolving threat of antimicrobial resistance: options for action*.

Wright, M.S., Suzuki, Y., Jones, M.B., Marshall, S.H., Rudin, S.D., van Duin, D. et al. (2014a) Genomic and transcriptomic analyses of colistin-resistant clinical isolates of *Klebsiella pneumoniae* reveal multiple pathways of resistance. *Antimicrob Agents Chemother*.

Wright, M.S., Perez, F., Brinkac, L., Jacobs, M.R., Kaye, K., Cober, E. et al. (2014b) Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern U.S. hospitals. *Antimicrob Agents Chemother* **58**: 4961-4965.

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R. et al. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci U S A* **110**: E3487-3496.

Wu, K.M., Li, L.H., Yan, J.J., Tsao, N., Liao, T.L., Tsai, H.C. et al. (2009) Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. *J Bacteriol* **191**: 4492-4501.

Wyres, K.L., Gorrie, C., Edwards, D.J., Wertheim, H.F., Hsu, L.Y., Van Kinh, N. et al. (2015) Extensive capsule locus variation and large-scale genomic recombination within the *Klebsiella pneumoniae* clonal group 258. *Genome Biol Evol*.

Yan, Y., Su, S., Meng, X., Ji, X., Qu, Y., Liu, Z. et al. (2013) Determination of sRNA expressions by RNA-seq in *Yersinia pestis* grown in vitro and during infection. *PLoS One* **8**: e74495.

Yu, V.L., Hansen, D.S., Ko, W.C., Sagnimeni, A., Klugman, K.P., von Gottberg, A. et al. (2007) Virulence characteristics of *Klebsiella* and clinical manifestations of *K. pneumoniae* bloodstream infections. *Emerg Infect Dis* **13**: 986-993.

Zogaj, X., Bokranz, W., Nimtz, M., and Romling, U. (2003) Production of cellulose and curli fimbriae by members of the family Enterobacteriaceae isolated from the human gastrointestinal tract. *Infect Immun* **71**: 4151-4158.

Figure legends

Figure 1: Phylogenetic relationship based on the sequence of 404 core-transcriptome genes, multilocus sequence types and *wzi* types of all 37 clinical isolates and 11 reference strains.

The unrooted Neighbor Joining tree was constructed using the aligned sequences of 404 genes which were covered by at least 5 nucleotides at each position in all isolates. Bootstrapping values of 1000 replications are indicated at each branch. MLST and *wzi* type results are also listed for each isolate. *wzi* alleles without perfect matches are indicated by “*wzi*-like” and the nucleotide differences to the closest matching allele are shown in brackets. Isolates included in this study are written in bold letters and reference genomes are indicated by a – k (table 1). Isolates belonging to the same MLST sequence type are highlighted in the same color. The scale indicates the number of base substitutions per site. The insert in the upper left corner shows the same tree in a radiation style to demonstrate phylogenetic distances.

Figure 2: The *Klebsiella pneumoniae* transcriptional landscape.

A: Histogram showing the expression values (in nRPK) of all 7859 genes after mapping all RNA sequencing reads of the 37 clinical isolates against the non-redundant gene list of the *K. pneumoniae* pan-genome (which was based on the *K. pneumoniae* 1084 genome and additional 10 reference strains).

B: The median, maximal and minimal expression value (in nRPK) of each gene is plotted. Genes are ranked according to their median gene expression on the x-axis (blue line). The

orange line denotes the expression sensitivity limit ($nRPK_0=2.20$). All genes with a maximal expression value below the sensitivity limit (never expressed under the given growth conditions) are highlighted in red ($n=719$). All genes with a minimal gene expression value above the sensitivity limit (expressed in every isolate [core-transcriptome, $n = 3346$]) are highlighted in green.

Figure 3: Expression variance of the core-transcriptome.

The plot shows the maximal (black dots) and minimal (gray dots) expression of the core-transcriptome genes ($n=3346$) ranked according to their increasing expression variance. Black lines show the moving average of the minimal (bottom line) and the maximal (top line) gene expression obtained using 100 genes equally spaced on either side of the gene where the mean was calculated.

Figure 4: Phylogenetic relationship is reflected in the transcriptional profile.

The heat maps show the expression of the pan-genome (A) and core-transcriptome (B), respectively. Genes (vertical) are hierarchically clustered using Pearson correlation distances, and the isolates (horizontal) are clustered according to their Spearman rank correlation. The same datasets are shown in C and D as three dimensional principal component plots. Sequence types are highlighted in colors: ST512 (magenta), ST258 (purple), ST101 (green), ST15 (orange) and ST38 (MGH 78578) (blue).

Figure 5: Expression profile of small regulatory RNAs.

The heat map shows the expression of 51 described and 30 predicted small, non-coding RNAs in the set of clinical isolates. RNA genes (vertical) are arranged according to their average expression from highest (top) to lowest (bottom) expression. Samples (horizontal) are clustered according to Spearman rank correlation and sequence types are highlighted in colors: ST512 (magenta), ST258 (purple), ST101 (green), ST15 (red) and ST38 (MGH 78578) (blue). The histogram describes the correlation of colors to nRPK values.

Figure 6: *Galleria mellonella* infection model.

The graph shows the rate of dead larvae 72 hours post infection. Data points represent the mean of duplicate experiments with 10 larvae each. They are colored according to increasing rates of dead larvae (green inverted pyramid ≤ 0.1 ; yellow circles < 0.5 ; orange diamonds < 0.9 and red triangles ≥ 0.9). Horizontal and vertical bars indicate mean and standard deviation of each phylogenetic group with at least two isolates, respectively.

Figure 7: Expression of virulence associated genes.

The circular plot shows the normalized read counts (in nRPK) of 86 virulence-associated genes (extracted from the virulence factor data base VFDB (Chen et al., 2012) and the *K. pneumoniae* BIGSdb database (Bialek-Davenet et al., 2014)) of each of the 37 clinical isolates. Expression varied from 12 nRPK (dark gray) to less than the sensitivity limit of 2.20 nRPK (white). Circular plots of the clinical strains are arranged according to their virulence phenotype as determined in the *Galleria* infection model from highly virulent (outside, dark red) to lowly virulent (inside, yellow). The full dataset is given in supplementary table S6.

Figure 8: Quantitative analysis of biofilm formation by *K. pneumoniae* clinical isolates.

A: Raw absorbance at 590 nm of 8 replicate measurements with standard error of the mean (SEM). Isolates are arranged according to their MLST sequence type. The dashed line shows the separation into high and low biofilm producing isolates (which corresponds to three times the medium control (absorbance at 590nm = 0.187)).

B: Differences in biofilm formation as determined by crystal violet staining between groups of isolates belonging to the same MLST type were analyzed using two-tailed Student's t-test. Statistical significance is indicated by asterisk according to the legend on the right.

Figure 9: Differences in gene expression between high and low biofilm forming isolates.

The circular plot shows the normalized read counts (in nRPK) of all genes which were identified to be highly expressed in biofilm forming *K. pneumoniae* isolates. Expression varied from 13 nRPK (dark gray) to less than the sensitivity limit of 2.20 nRPK (white). Circular plots of the clinical strains are arranged according to their biofilm phenotype. The 12 high biofilm forming isolates are shown in the outer circles, followed by the 14 low biofilm forming isolates in the inner circles. The MLST type is indicated as colored segments: ST512 (magenta), ST258 (purple), ST101 (green), ST15 (orange) and ST38 (MGH 78578) (blue). The full dataset is given in supplementary table S8.

Tables

Table 1: List of completely sequenced *K. pneumoniae* genomes used in this study

GenBank Accession ID	<i>K. pneumoniae</i> strain	Locus ID / format used	Number of genes	MLST sequence type
AP006725	NTUH-K2044 (Wu et al., 2009)	KP1_0001	5123	23
CP000647	MGH 78578	KPN_00001	4887	38
CP002910	KCTC 2242 (Shin et al., 2012)	KPN2242_r25078	5035	375
CP003200	HS11286 (Liu et al., 2012)	KPHS_00010	5404	11
CP003785	1084 (Lin et al., 2012)	A79E_0001	5067	23
CP003999	Kp13 (Ramos et al., 2014)	KP13_00049	5299	442
CP006648	CG43	D364_00005	4897	86
CP006656	JM45	N559_0001	4980	11
CP006918	30684/NJST258_2 (Deleo et al., 2014)	KPNJ2_00001	5545	258
CP006923	30660/NJST258_1 (Deleo et al., 2014)	KPNJ1_00001	5577	258
CP006659	ATCC BAA-2146 (Hudson et al., 2014)	KPN2146_0001	5498	11
Total genes (including 30 predicted small RNAs):			57342	
pan-genome:			7859	
unique genes (singletons):			1598	
core-genome:			3336	
accessory genes:			2925	

Table 2: List of genes being higher expressed in biofilm forming isolates

Gene name / locus ID	dnRPK	Gene name / Locus ID	dnRPK
<i>bcsB</i>	2.66	A79E_2827	3.62
<i>bcsC</i>	2.67	A79E_2841	3.03
A79E_0470	5.91	<i>ompW (yciD)</i>	3.26
A79E_0471	6.76	A79E_3106	3.57
A79E_0472	2.58	<i>dmsA</i>	2.07
<i>ttdA</i>	3.40	A79E_4441	6.15
<i>ttdB</i>	2.27	<i>ulaC (ptxA)</i>	2.30
A79E_0491	2.09	<i>ulaB (sgaB)</i>	2.36
<i>mrkA</i>	6.00	<i>frdD</i>	2.21
<i>dalT</i>	2.31	D364_02760	6.30
A79E_2240	3.19	D364_02765	6.86
A79E_2397	4.75	D364_17515	5.08
A79E_2413	6.18	D364_17530	2.71
A79E_2451	5.51	D364_17540	3.26
A79E_2489	2.12	D364_17550	6.62
A79E_2492	2.07	<i>acsA</i>	2.89
A79E_2600	3.31	KP13_01340	4.29
A79E_2740	2.63	KP13_04177	6.84
A79E_2741	3.16	KP13_32375	2.45
<i>astB</i>	3.38	KPN2242_20090	5.99
<i>astD</i>	3.10	KPN_01389	2.37
A79E_2744	3.20	KPN_01390	3.36
<i>astC</i>	2.88	KPN_01391	3.52

dnRPK: Difference in nRPK between the groups of high and low biofilm forming isolates

Supplemental Material

Supplementary Figure S1: Analysis of the *Klebsiella pneumoniae* genomic content.

Supplementary Figure S2: Functional annotation and analysis of different sets of genes

Supplementary Table S1: Detailed description of the clinical isolates used in this study

Supplementary Table S2: Complete list of genes in the *K. pneumoniae* pan-genome with information about gene expression and variation

Supplementary Table S3: Complete list of genes of the accessory transcriptome

Supplementary Table S4: Expression of sRNAs in *K. pneumoniae* clinical isolates

Supplementary Table S5: Growth parameters in LB, BM2+mannitol and BM2+glucose

Supplementary Table S6: Expression of known virulence associated genes in the pan-genome

Supplementary Table S7: Virulence associate genes in the accessory transcriptome

Supplementary Table S8: Genes identified to be differentially expressed in high biofilm producing isolates vs. low biofilm producing isolates