Diverse polysaccharide production and biofilm formation abilities of clinical Klebsiella pneumoniae

Robert Beckman Iv, Elenora Cella, Taj Azarian, Olaya Rendueles, Renee Fleeman

npj | biofilms and microbiomes

Article

Published in partnership with Nanyang Technological University

https://doi.org/10.1038/s41522-024-00629-y

Diverse polysaccharide production and biofilm formation abilities of clinical *Klebsiella pneumoniae*

Robert L. Beckman IV[1], Elenora Cella[1], Taj Azarian[1], Olaya Rendueles[2] & Renee M. Fleeman[1]

Klebsiella pneumoniae infections have become a growing threat for human health. The lack of understanding of the relationship between antibiotic resistance, mucoviscosity, and biofilm formation impedes our abilities to effectively predict *K. pneumoniae* infection outcomes. The Multidrug-Resistant Organism Repository and Surveillance Network offers a unique opportunity into the genetic and phenotypic variabilities in the *K. pneumoniae* isolates. To this end, we compared the genetic profiles of these isolates with the phenotypic biofilm formation, percent mucoviscosity, and growth rates. There was a significant phenotype-genotype correlation with decreased biofilm formation and an insertion sequence in the transcriptional activator of the type III fimbrial system. Interestingly, the most mucoid strains in the populations were lacking the genetic element regulating the mucoid phenotype and three of these isolates were able to form robust biofilms. The combination of phenotypic, genomic, and image analyses revealed an intricate relation between growth, mucoviscosity and specific virulence-associated genetic determinants.

Klebsiella pneumoniae is an emerging threat to human health due to its extreme rate of drug resistance acquisition[1–3]. Until recently, these strains have primarily infected immune-compromised patients due to their lack of virulence factors to bypass the host immune system[4,5]. However, horizontal transfer of large virulence plasmids within the *K. pneumoniae* species has now allowed for hypervirulent isolates to emerge within the community and infect healthy individuals[6–8]. Historically, the hypermucoviscous phenotype was primarily observed with *K. pneumoniae* hypervirulent K1 and K2 capsule serotypes[9] and their increased mucoviscosity has been shown to be due to the virulence plasmid pLVKP encoding genes for production of RmpA, the transcriptional regulator of the mucoid phenotype[10]. However, recent studies have shown the spread of this virulence plasmid to other capsule serotypes[11]. Virulence acquisition in combination with the pan-drug resistance status of many isolates of *K. pneumoniae* has elevated this species to the CDC urgent threats for concern to human health[12]. Interestingly, difficulties have been encountered accurately predicting virulence in clinical *K. pneumoniae*[13,14].

Surface attachment and biofilm formation are a crucial first step of the infection process, but this important virulence trait is difficult to link to a genotype or predict from genomic data[15–17]. Characterization of the matrix of *K. pneumoniae* biofilms found it is largely composed of lipopolysaccharides and capsular polysaccharides[18,19]. Previous studies have highlighted the importance of capsular polysaccharides in biofilm formation and implicated its deleterious effect through occluding the fimbriae that allow for surface attachment[20,21]. The capsule has been shown to have a positive or negative correlation to biofilm formation depending on the overall ability of the isolate to form biofilms[22,23]. Furthermore, the deletion of capsule in a hypermucoviscous K1 serotype isolate increases biofilm formation abilities but decreases biofilm formation with a K124 serotype isolate, indicating adhesion factors and capsule abundance together drive biofilm formation[23]. These findings have highlighted the need to better understand how highly capsulated *K. pneumoniae* can bypass the impediment caused by capsule to form robust biofilms. *K. pneumoniae* attachment is mediated by the chaperone-usher systems type I and type III fimbriae encoded by the *fimA-K* and *mrkA-I* operons, respectively[22,24]. Mutations within *mrkD* and *fimH*, which are the tip adhesins of each fimbrial system, have been shown to vary based on the characteristics of the isolate that acquired the mutation[22].

Here we determine biofilm formation potential of the Multidrug-Resistant Organism Repository and Surveillance Network (MRSN)[25]. The MRSN collection is a diverse set of *K. pneumoniae* covering a broad range of the potential multi-locus sequence type (MLST) with the genome sequencing data and curated metadata available for all isolates. Therefore, this collection is optimal to compare the biofilm formation characteristics

[1]Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, FL, USA. [2]Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), UMR5100, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Toulouse III - Paul Sabatier (UT3), Toulouse, France. e-mail: Renee.Fleeman@ucf.edu

observed to the genetic profiles for a greater understanding of the range of physical attributes important for strong biofilm formation. To do so, we leveraged classical phenotypic tests with comparative genomics and confocal microscopy. We assessed biofilm formation, biofilm structure/morphology, mucoviscosity, and growth rates of the MRSN collection in order to align our results to their phylogeny and fimbriae allele variations. For deeper insight, we enriched the dataset with the virulence and antibiotic resistance scores from the original publication of this set, along with the host isolation source[25]. Our analyses identified a surprising diversity in the ability to form biofilms, encompassing both mucoid isolates and non-mucoid isolates. Additionally, we show that only two isolates displaying increased mucoviscosity carry the *rmpA* genetic element encoding for the hypermucoviscous phenotype[9], suggesting that other genetic mechanisms can determine mucoviscosity in a broader set of isolates. In line with previous studies, we identified mutations within the tip-adhesin of type 3 fimbriae, encoded by *mrkD*, were prevalent within the mucoid isolates and insertion sequences in *mrkH* were strongly correlated with decreased biofilm formation. Finally, our confocal imaging analysis reveals matrix composition variability between mucoid and non-mucoid isolates that form robust biofilms. Our work reveals the presence of a mucoid phenotype without the genetic elements and the composition of the matrix varies greatly depending on the mucoviscosity of the isolates.

Methods

Crystal violet biofilm staining

To assess for biofilm formation, the strains were cultured in lysogeny broth (LB) and placed in shaking incubator (220 rpm) for 24 h at 37 °C. The overnight cultures were diluted to an OD_{600} of 0.5 (9.75×10^9) in biofilm media (tryptic soy broth and 0.5% glucose). The biofilms were grown in tissue culture treated 6-well plates in triplicate for 24 h at 37 °C static. Supernatant was removed from wells and biofilms were washed with 1 mL phosphate-buffered saline (PBS). Biofilms were stained with 1 mL 0.1% crystal violet (CV) for 15 min. and placed in a fume hood for 24 h. Destaining of biofilms was performed by adding 1 mL 30% acetic acid to wells for 15 min. and transferring the mixture to a new 96-well plate. Microplate reader was used to measure the optical density at OD_{550}. Control wells were seeded with sterile biofilm media, incubated for 24 h, and stained alongside the MRSN biofilms. The OD_{550} values obtained from the control wells were subtracted from OD_{550} readouts from biofilm samples values to normalize and account for dye binding to the plastic of the wells.

Congo Red and calcofluor white agar plating

Overnight cultures were grown in lysogeny broth (LB) with shaking (220 rpm) for 24 h at 37 °C. Overnight cultures were diluted 1:100 with LB media in a 96-well plate. Congo Red plates were prepared by adding 4 mL 0.1% Congo Red solution and 3.6 mL 50% sucrose solution to LB agar for a final volume of 50 mL. 10 mL of calcofluor white staining dye was added to 40 mL LB agar for a final volume of 50 mL and a final concentration of 200 μg mL^{-1}. 5 μL diluted cultures were spot plated onto dried Congo Red and calcofluor white plates and imaged after 24 h of growth at 37 °C.

Mucoviscosity assay

To assess for the hypermucoviscous phenotype characteristic of hypervirulent *K. pneumoniae*, the strains were cultured in LB with shaking (220 rpm) for 24 h at 37 °C. 1 mL of overnight cultures were placed in 1.5 mL microcentrifuge tubes and OD_{600} values were recorded before and after centrifugation at 1000 × *g* for 5 min. Mucoviscosity was calculated as:

$$\text{Percent Mucoviscosity} = \frac{(\text{OD 600 nm after centrifuge})}{(\text{OD 600 nm before centrifuge})} \times 100 \quad (1)$$

The resulting values were graphed as mean percent mucoviscosity with error reported as ± SEM.

Percoll Gradient Assay

Percoll™ (GE Healthcare) solutions (15, 35 and 50%) were prepared by combining Percoll solution with phosphate-buffered saline (PBS) in 50 mL conical tube to achieve 50%, 35%, and 15% concentrations. 2 mL of the 50% Percoll solution was placed at the bottom of 15 mL centrifuge tube followed by 2 mL of 35% then 2 mL 15% Percoll solution. Strains were cultured in LB with shaking (220 rpm) for 24 h at 37 °C. Overnight cultures were centrifuged at 3434 × *g* for 10 min., the supernatant was removed, and the pellets were resuspended in 600 μL PBS. 400 μL of resuspended cultures were pipetted on top of Percoll gradients and centrifuged at 3000 × *g* (room temperature) for 30 min. All strains were tested in triplicate and representative photos were shown for each.

Genome alignment and tree generation

Genome sequencing data was acquired from the authors of the original manuscript at Walter Reed Army Institute of Research (WRAIR)[25]. Raw sequencing reads were filtered using Trimmomatic v.0.39 and quality assessed using FastQC[26]. De novo genome assemblies were constructed using Unicycler v0.4.8[27] and annotated using Prokka v1.14.15[28]. Pangenome analysis was performed using Roary v.3.12[29] and a core genome SNP alignment was extracted using snp-sites v2.4.0. A maximum likelihood (ML) phylogeny was inferred with IQTREE v1.6.8 using the ASC + GTR + GAMMA substitution model with 100 bootstrap replicates and rooted using MRSN 25947 strain (ST5447)[30]. The Kleborate tool[31] was then used to perform K and O typing using the -- kaptive option[32].

To descriptively assess the distribution of phenotypic and genotypic determinants, phenotypic biofilm formation, phenotypic mucoviscosity, genotypic mucoviscosity (*rmpADC*), genotypic virulence determinants, genotypic antibiotic resistance determinants, and allele typing of *mrkD*, *fimH*, and *ecpD* genes, were mapped onto the core-genome phylogeny using ggTree in RStudio, running R version 4.3.1[33]. We aligned the phylogeny of the isolates with their biofilm formation and mucoviscosity by grouping into bins to aid in visualization of the alignments (biofilm formation bin 1 – bin 4; mucoviscosity bin 1 – bin 5). For biofilm formation grouping (BIO), we assigned a cutoff of OD_{550} greater than 5 for the strongest biofilm formers in bin 1; OD_{550} greater than 2 and 1 for moderate biofilm formers in bins 2 and 3, respectively; and OD_{550} lower than 1 for very weak biofilm formers in bin 4. For mucoviscosity grouping (MV), we assigned a cutoff of >10% mucoviscosity for the most mucoid isolates in bin 1, between 5 and 10% mucoviscosity in bin 2, between 5 and 3% mucoviscosity in bin 3, between 3 and 1% mucoviscosity in bin 4, and 19 <1% mucoviscosity in bin 5.

Identification of tip adhesins, curli and cellulose operon

Identification of the cellulose operon and curli biogenesis apparatus was performed as described previously[34]. Briefly, the experimentally validated protein sequences involved in curli and cellulose synthesis and identified previously[34] were used as a query (Supplemental Table 1). For the identification of tip adhesins, the following proteins were used as query WP_004149659.1, BAH65076.1, WP_002890060.1 for MrkD, FimH and EcpD, respectively. BlastP (v2.7.1 +) with default parameters was used to search for each protein in the proteome of the MRSN collection. (i) *Curli*. Hits for the regulatory protein CsgD were identified, but no hits were obtained (*E*-value $<10^{-5}$ & identity >60%) for any other protein, indicative of the absence of curli synthesis. (ii) *Cellulose*. For each proteome, we search for hits in each protein of the *bcs* operon (*E*-value $<10^{-5}$ & identity >60%). All hits in each proteome colocalized, and the operon structure was inferred. Ninety-seven out of the 100 genomes had a genomic architecture corresponding to *bcsABCEFRGQZ*. (iii) *Tip adhesins*. Sequences with an identity percentage of less than 80% were discarded. Reducing the threshold to 70% or increasing it to 90% did not alter the number of sequences discarded. All genomes had either one or zero hits per adhesin. To identify the different alleles, the most common protein sequence was identified for *mrkH*, *mrkD*, and *fimH*. All other proteins were aligned to the most frequent protein sequence using the *pairwiseAlignment* function from Biostrings and differences were identified with *mismatchTable* function. For *mrkD* and *fimH*,

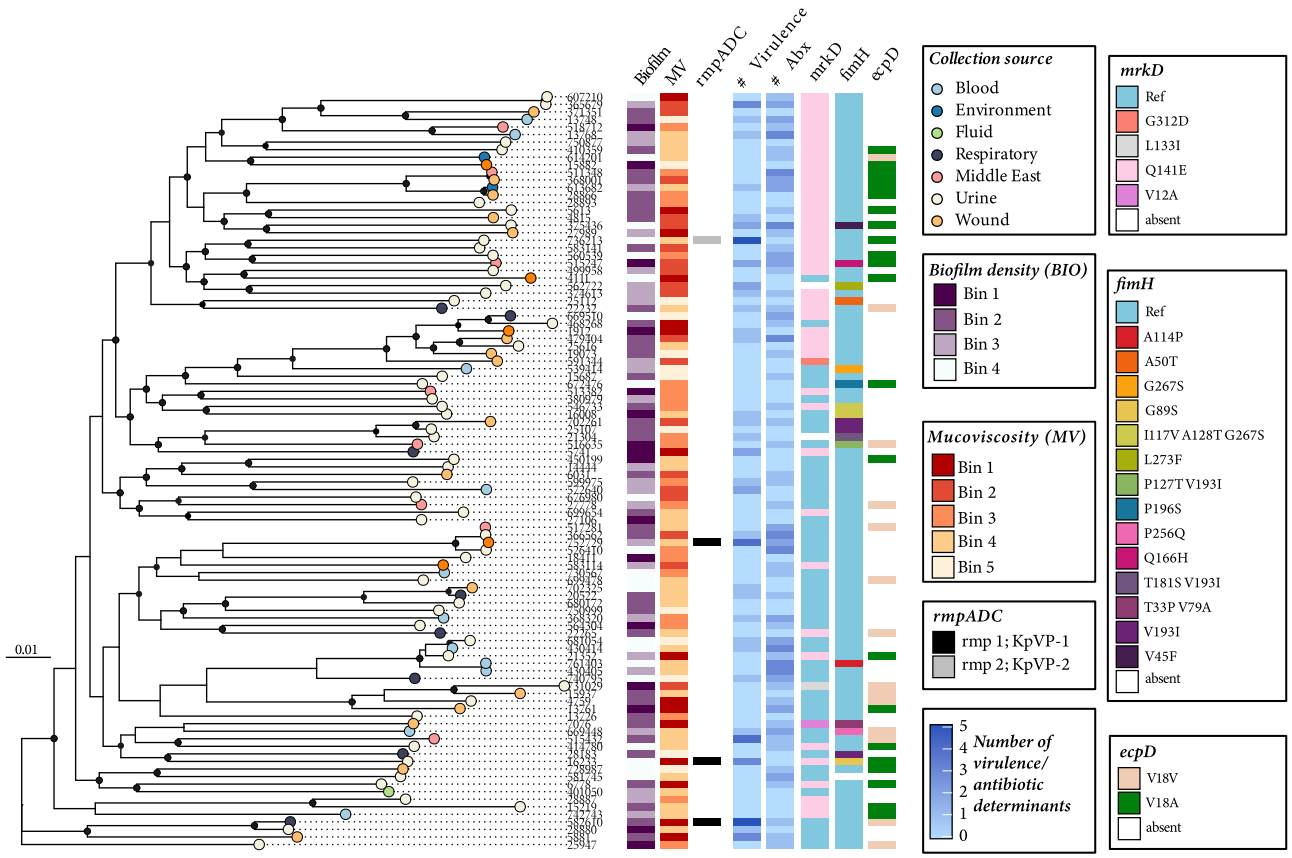

Fig. 1 | Genotype and phenotype aspects of the MRSN collection related to biofilm and mucoviscosity. The figure shows the maximum likelihood phylogeny illustrating population structure, phenotypic biofilm density using a normalized OD_{550} readout, phenotypic mucoviscosity, genotypic mucoid regulator presence (*rmpADC*), genotypic virulence and antibiotic resistance determinants, and typing of *mrkD*, *fimH* and *ecpD* genes among *K. pneumoniae* isolates ($N = 100$) previously published[25]. Tip shades indicate the host isolation source from the original publication[25]. A black circle along the branch indicated a higher bootstrap value (≥0.7). The presence of determinants is shown with a colored rectangle on the heatmap according to the legend on the right.

the most frequent protein sequence was referred to as the reference allele in Figs. 1 and 5. When performing the *ecpD* allele identification, the most common allele was the absence of an allele, therefore no reference gene is shown in Fig. 1. The AlphaFold Protein Structure Database was used to predict and generate the fimbriae structures, last accessed on March 2024[35,36].

Growth curve parameters

Overnight cultures were diluted at 1:100 in fresh LB media. Two hundred microliters of each subculture was transferred into 96-well microplate. Absorbance (OD_{600}) of the wells were measured with a TECAN Genios™ plate reader. Absorbance values from within-block technical replicates were averaged and these averages were used as statistically independent data points. (*i*) *Growth rate*. Minimum generation times were estimated across replicates for the 1 h interval (ΔT) spanning the fastest growth during the exponential growth phase. This was calculated as follows:

$$\frac{1}{Td} = \frac{\log_{10}(\text{maximum}_{OD600}/\text{minimum}_{OD600})}{\log_{10}(2) * \Delta T} \quad (2)$$

(ii) *Maximum yield*. This measure corresponds to the maximal OD_{600} reached by each culture. (*iii*) *Area under the curve* (AUC). AUC takes into account the lag phase, growth rate and population yield of the culture. It was calculated using the R function *trapz* from the *pracma* package.

Biofilm formation static growth for confocal imaging

Overnight cultures were grown in LB in shaking incubator (220 rpm) at 37 °C for 24 h. Overnight cultures were standardized to OD_{600} 0.5 (9.75×10^9) in biofilm media (TSB media with 0.5% glucose). 1 mL of OD_{600} 0.5 (9.75×10^9) standardized cultures were pipetted into Matsunami Glass bottom dishes (Glass Diameter: 14 mm, Glass Thickness: #1.5 (0.16–0.19 mm)). Matsunami dishes were parafilmed and placed in 37 °C static incubator for 24 h. After 24-h growth, supernatant of biofilms was removed, and samples were washed with 1 mL 1X PBS. 1 mL of 5 µM SYTO 9 green-fluorescent stain diluted in 1X PBS was pipetted onto biofilm samples and rocked at medium speed for 1 h. Dye was removed, and samples were washed with 1 mL 1X PBS. 1 mL of 50 µg mL^{-1} calcofluor white dye diluted in molecular grade water was pipetted onto all samples and rocked at medium speed for 5 min. Dye was removed, and samples were washed with 1 mL 1X PBS. Images were recorded using z-stack confocal microscopy imaging. For imaging of MRSN 1912 with constitutive RFP expression we electroporated pBTK1007 and post-stained with calcofluor white stain. pBTK1007 (pSL6) was a gift from Jeffrey Barrick (Addgene plasmid #191002; http://n2t.net/addgene:191002; RRID:Addgene_191002)[37].

Biofilm formation in microfluidic device

To introduce environmental shear flow, we utilized the BioFlux One system (Cell Microsystems, Durham, NC; https://cellmicrosystems.com/bioflux/). This is an automated electropneumatic pumping system with associated operating software and a proprietary Well Plate Microfluidic™ device to increase the throughput of analysis of biofilm formation under physiological shear flow. The system is designed for flow control and accuracy, with automated pressure controllers that can control shear force to ±0.05 dyne/cm^2. The plates utilize a standard well plate format with an imaging surface having a #1.5 170 µm glass bottom and imaging channel dimensions of 350 µm wide, 70 µm deep, and 4 mm long. Using the BioFlux One device

and a 48-well low shear plate (0–20 dyne/cm²), we tested the top biofilm formers from our crystal violet analysis. Briefly, overnight cultures were grown in LB in shaking incubator (220 rpm) at 37 °C for 24 h. Overnight cultures were standardized to OD_{600} 0.5 (9.75×10^9) in minimal media (with 0.4% glucose)[38] pre-warmed to 37 °C. After warming the media, outlet wells received 40 µL sterile minimal media (with glucose), and flow was directed from outlet to inlet wells at 20 dyne/cm² (2356 µL h^{-1}) for 40 s. 100 µL of sterile media was immediately added to inlet wells to avoid drying. Residual media was removed from outlet wells, and 20 µL standardized cultures in pre-warmed media were added to outlet wells. Flow was directed from outlet to inlet wells at 4 dyne/cm² (471 µL h^{-1}) for 5 s. 48-well low shear plate (0–20 dyne/cm²) was incubated for 1 h at 37 °C. Outlet wells were cleared of all liquid and rinsed with 100 µL fresh sterile minimal media (37 °C). 1 mL 37 °C fresh sterile minimal media was added to inlet wells, and flow was directed from inlet to outlet wells at 1 dyne/cm² (118 µL h^{-1}) for 5 min. Outlet wells were rinsed with 100 µL 37 °C fresh sterile minimal media, and inlet wells were assessed to ensure 1 mL media was still present. 48-well low shear plate (0–20 dyne/cm²) received flow from inlet to outlet wells at 0.55 dyne/cm² (65 µL h^{-1}) for 24 h. Images of 48-well low shear plate (0–20 dyne/cm²) were recorded using z-stack confocal microscopy imaging.

Confocal microscopy z-stack imaging

Samples were transported in light protected blackout bin due to light sensitive staining dyes. Zeiss LSM 710 confocal microscope and ZEN microscope software were powered on. 63X oil immersion lens was used to image samples. Small droplet of lens oil was applied to microscope lens and the Matsunami Glass dish or the Bioflux 48-well low shear plate containing sample was placed onto microscope lens holder. The microscope was lowered and GFP fluorescent detection laser was turned on and used to bring cells into focus. Once in focus, z-stack images were taken with using 488 and 543 nm laser channels for imaging of calcofluor white and SYTO 9 staining, respectively. The figures shown are 3D renderings of the z-stack images obtained to show biofilm thickness on the z-axis using a collective of images from multiple focal planes to produce a single 3D image.

Statistics

Many statistics were performed with R v 4.3.1. (*i*) *Correlations between continuous variables*. To correlate the different continuous variables (growth parameters, percent mucoviscosity, and biofilm) Spearman's rank correlation was used. (*ii*) *Correlations between discrete and continuous variables*. We used general linear models (GLM) to test the association between the different virulence, drug-resistance scores with mucoviscosity, and biofilm and other covariates using the *glm* function. We fitted the model with either resistance or virulence as dependent variables (Y) and percent mucoviscosity, biofilm and growth were independent variables (X), following the formula $Y \sim X$. We assessed the relevance of the focal independent variable by testing if the parameter estimate for the variable was significantly different from zero (when the overall model had an R^2 significantly higher than zero). (*iii*) *Estimate of phylogenetic inertia*. The presence of phylogenetic signal in the evolution of traits was estimated with Pagel's lambda and Bloomberg's K using the *phylosig* function of the *phytools* package v.2.1-1 for R[39]. These are two quantitative measures of phylogenetical signal, or the tendency of related species to have similar phenotypes to one another as a consequence of shared evolutionary history. Pagel's λ is a scaling factor that measures correlation between isolates relative to that expected by Brownian motion, whereas Bloomberg's K measures the variance across isolates[40,41].

Results

Biofilm density is greatest with a limited set of isolates, containing a disproportionately high ratio from the Middle East, mostly isolated from urine samples, and occurs in both classical and hypervirulent pathotypes

To rigorously test biofilm formation across the 100 diverse *K. pneumoniae* isolates of the MRSN collection, we grew all strains statically in TSB supplemented with 0.5% glucose, a robust medium providing the most optimal biofilm formation conditions for our analysis. We then used a crystal violet staining readout and measured optical density at 550 nm (OD_{550}) to determine biofilm density without bias for cells or polysaccharide, as this dye stains both of these elements[42]. The phylogeny was inferred from a core-genome alignment of 3729 Clusters of Orthologous Groups (COGs) of proteins and 149,624 single nucleotide polymorphism (SNP) sites (Fig. 1). This resulted in several well-supported clades, representing a broad range of MLSTs, which is in line with what was observed in the original work with these isolates[25]. When comparing the biofilm formation of the 100 isolates, we saw a great diversity of biofilm formation abilities, ranging from no visible biofilm to formation of a very robust biofilm. Such heterogeneity indicates that of the tested phenotypes in this work, biofilm is one of the most diverse and versatile traits in *K. pneumoniae*, as revealed by the high diversity index (Table 1). This is also observable in the distribution of biofilm formation, which showcases a long tail composed of *ca*. 15% of strains with an exceptionally strong capacity to form biofilm (Supplemental Fig. 1A). Furthermore, the diversity of biofilm formation was spread between the clades, with each clade composed of both strong and weak biofilm formers, suggesting skim influence of phylogeny (Table 2). Our analysis was deposited to be made publicly available and consultable (https://microreact.org/).

We found overall many isolates in the diversity set had limited biofilm formation, with 38 isolates displaying moderate biofilm formation (Supplemental Fig. 1B), and 46 isolates formed very limited biofilm (Supplemental Fig. 1C). However, a select group of 15 isolates formed extremely dense biofilms (upper 15% of the distribution tail; bin 1; crystal violet $OD_{550} > 5$) (Fig. 2a). To begin our comparison of the covariates which may influence biofilm formation, we assessed the geography, patient isolation, and pathotype and found there were no significant correlations with these covariates. Of note, we found that a sample isolated from Asia (MRSN 731029) had the greatest biofilm formation of the diversity set, and samples from the Middle East, on average, formed the most robust biofilms (Fig. 2b).

Table 1 | The diversity index weighing the variance of a trait, allowing direct comparisons across traits

Trait	*I_x
Biofilm	1.18
Mucoviscosity	1.52
AUC	0.1
Generation time	0.1
Maximum yield	0.11

*Diversity indexes (Ix) were calculated as the standard deviation of a sample, divided by the mean.

Table 2 | Phylogenetic inertia calculated using Bloomberg's K and Pagel's λ

Trait	Bloomberg's K	*P* value K	Pagel's λ	*P* value Pagel
FimH	0.07	0.540	0.73	0.008
MrkD	0.58	0.008	1.00	0.012
EcpD	0.05	0.329	0.84	<0.001
AMR	0.13	0.127	0.73	<0.001
Virulence	0.05	0.665	0.00	1.000
AUC	0.13	0.210	0.71	0.003
Biofilm	0.30	0.063	0.00	1.000
Mucoviscosity	0.06	0.650	0.00	1.000

We estimated the phylogenetic inertia of all tested variables using Pagel's λ and Bloomberg's K with the function included in the *phytools* package and a phylogenetic tree based on the core genome. The null hypothesis is $\lambda = 0$ (no phylogenetic effect).

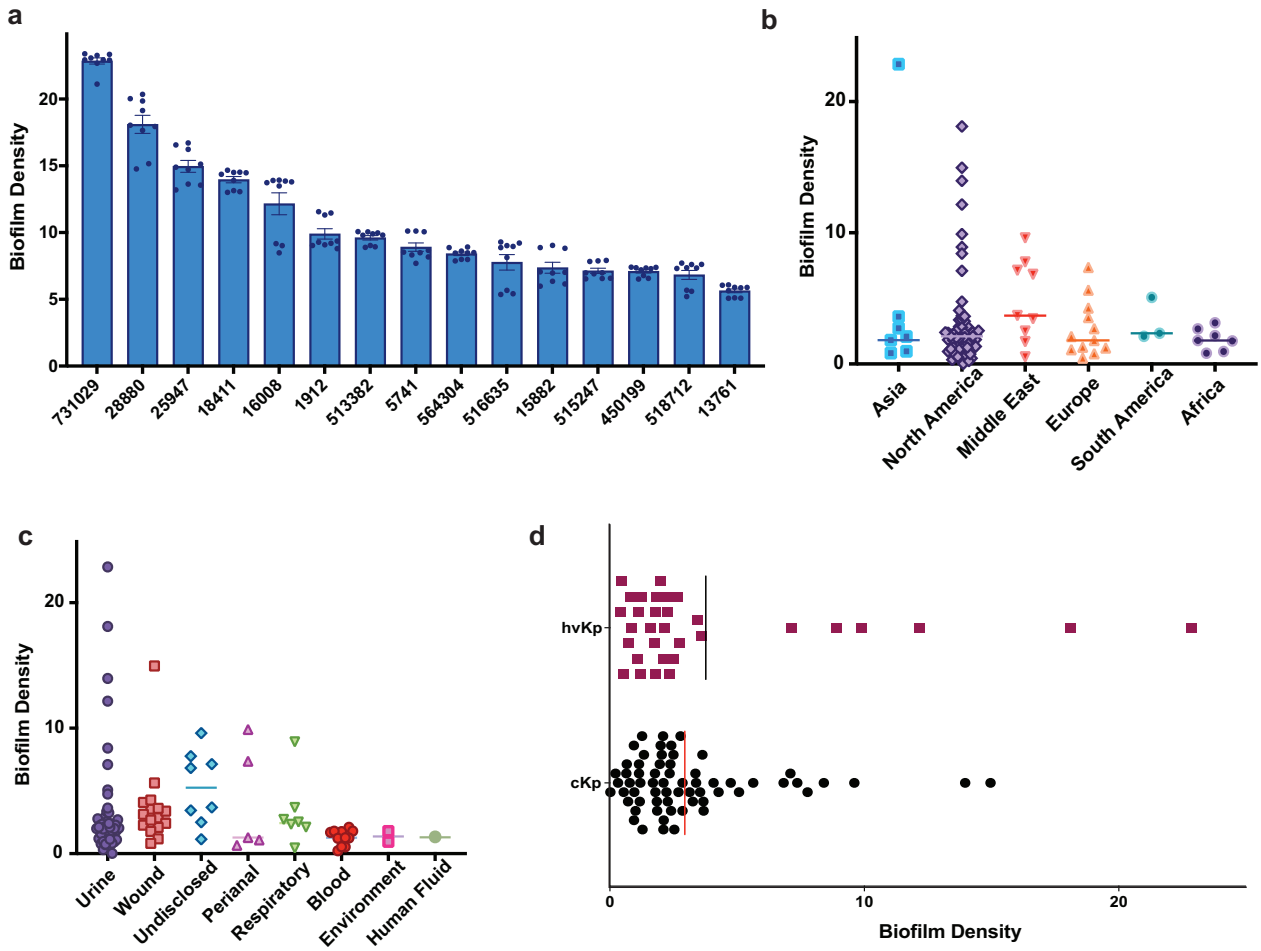

Fig. 2 | Biofilm density is greatest with a limited set of isolates, containing a disproportionately high ratio from the Middle East, mostly isolated from urine samples, and occurs in both classical and hypervirulent pathotypes. The figures show the biofilm density of the isolates using a normalized OD_{550} readout. Figure 2a shows the 15% of strains displaying the strongest biofilm formation of the MRSN collection. Figures 2b and 2c show the respective geographical location and host isolation site of the diversity panel isolates. Figure 2d displays biofilm density of both cKp and hvKp isolates. Figure 2a shows the mean of $n = 9$ biofilms for each isolate with error as ± SEM. Figures 2b, 2c, and 2d show the mean biofilm density of each isolate with a line for the mean of the population.

When considering the host isolation site, samples isolated from patient urine had the most isolates with strong biofilm formation, while isolates from the blood, human fluid, and the environment had the least ability to form biofilms (Fig. 2c). To observe the variability of biofilm formation with the presence of virulence genes, we graphed the *K. pneumoniae* isolates with virulence scores 1–5 (hvKp) next to those with a virulence score of 0 (cKp)[25]. We found biofilm formation abilities were similar between hypervirulent and classical pathotypes (Fig. 2d).

Since biofilm formation abilities have been previously linked to curli fimbriae and cellulose production[43,44], we assessed the abilities of the MRSN collection for production of these factors as previously described using Congo Red agar and calcofluor white agar plates, respectively[45]. From our Congo Red analysis, we found a great diversity between the 100 isolates within the collection (Supplemental Fig. 2). Specifically, MRSN 564304 produced dark red phenotype indicative of curli production, while the strongest biofilm former MRSN 731029 displayed very little color. Furthermore, MRSN isolates 1912 and 5741 displayed a pink color rather than red possibly due to the abundance of capsule production like what is observed with the control hypervirulent *K. pneumoniae* NTUH K2044 (Supplemental Fig. 2). However, similar to what was recently shown with *Klebsiella variicola*[34], within the entire collection, the operon *csgABCDEF* was not present suggesting the phenotypes observed on Congo Red are due to another factor. In complementation of the Congo Red analysis, when grown on calcofluor white agar and visualized under UV fluorescence we saw diverse fluorescence degrees, indicating variability in cellulose production within the entire collection (Supplemental Fig. 2). As opposed to the curli operon, we did find the *bcsABCEFGRQZ* operon encoding for production of cellulose within the genomes of the collection. Most isolates had a complete operon, with exception of MRSN 736213, 16008, and 450199 that had *bcsABCZ*, *bcsBCZ*, and *bcsABCEFQZ*, respectively. Although MRSN 736213 had limited biofilm formation, MRSN 16008 and 450199 were strong biofilm formers. Collectively, we conclude that the strongest biofilm formers were isolated from the urine, both cKp and hvKp are capable of forming robust biofilms, and the presence of curli and cellulose are not sufficient to describe the biofilm formation abilities of this collection suggesting other factors are contributing to biofilm formation.

Mucoviscosity limits biofilm formation and is not exclusive to the K1 and K2 capsule serotype

Literature to date has shown that capsule is important for biofilm formation[46,47]. However, the increased encapsulation has been suggested to interfere with the attachment of bacterial cells resulting in decreased biofilm formation[21]. Hypervirulent strains with K1 or K2 capsule serotypes are typically portrayed as hypermucoviscous, partly due to increased capsule production[9,48]. However, recent research has shown that lowly capsulated strains can be hypermucoviscous, and highly capsulated strains may have low mucoviscosity[14,49,50]. As mucoviscosity is a well-established proxy of hypervirulence, we hypothesize it could correlate with biofilm formation. To

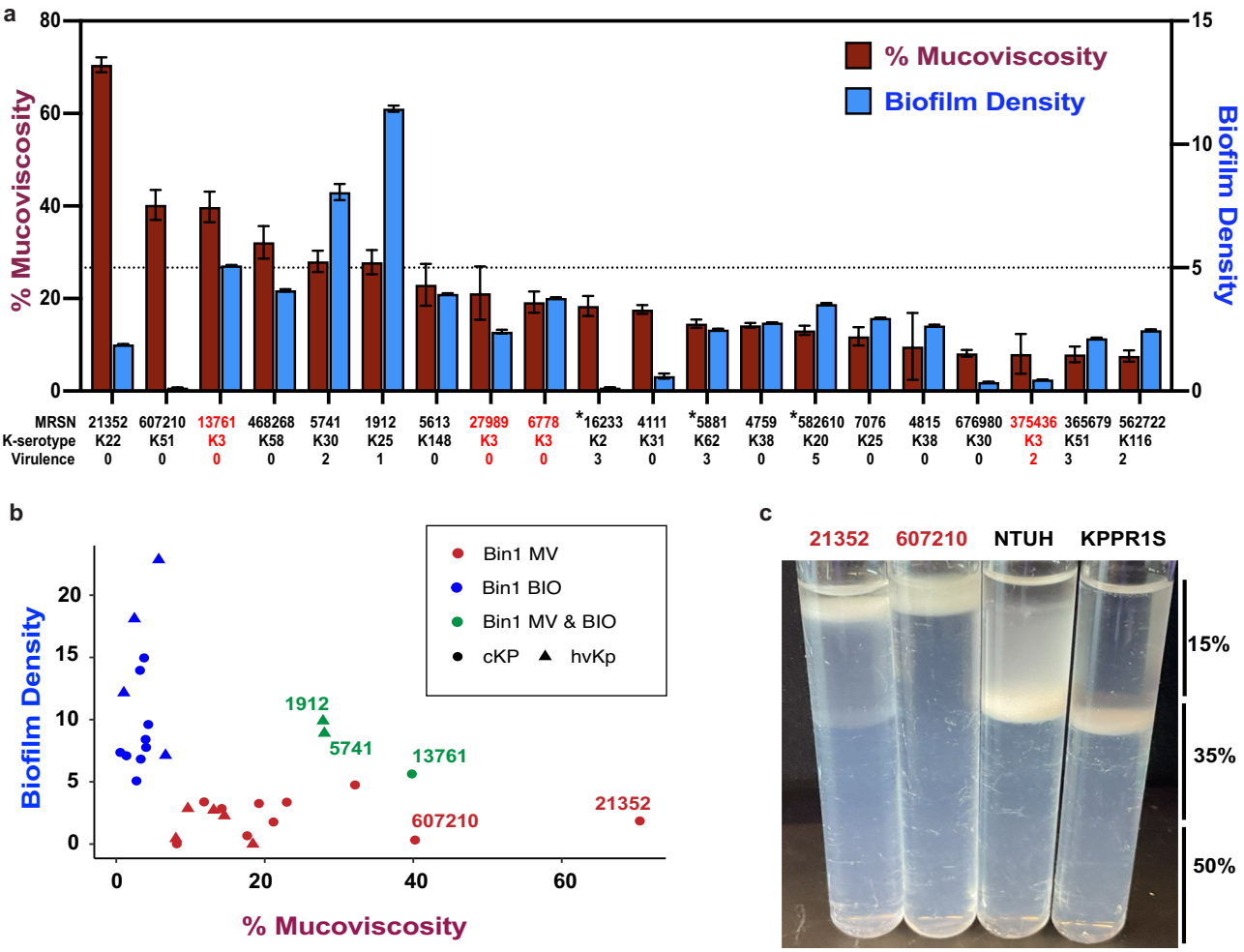

Fig. 3 | Mucoviscosity limits biofilm density and is not exclusive to the K1 and K2 capsule serotype. The figures show percent mucoviscosity compared to biofilm density normalized to OD_{550} and the Percoll migration of the most mucoid isolates in the collection compared to hypermucoviscous isolates. Figure 3a shows the 20 isolates (20% distribution tail) with % mucoviscosity >10% on the left *y*-axis and the biofilm density on the right *y*-axis. The K serotype of each isolate is shown along with the virulence score. The isolates that have the *rmpA* genetic element are marked with an asterisk and K3 capsule serotype are highlighted with red text. The dashed line represents the 15% distribution tail cutoff for strong biofilm formers. Figure 3b shows the percent mucoviscosity (*x*-axis) and biofilm density (*y*-axis) of the isolates found within upper 15% of the distribution tails of the biofilm (Bin1 BIO) and % mucoviscosity (Bin1 MV) analyses. The isolates that are in both distribution tails are colored separately for comparison. The pathotype (hypervirulent or classical) of each isolate is shown with a triangle or circle, respectively. Figure 3c shows the Percoll density gradient migration of the two most mucoid isolates in the collection compared to the control hypermucoviscous isolates NTUH K2044 (K1) and KPPR1S (K2). Figure 3a shows the mean of $n = 9$ biofilms for each isolate with error as ± SEM.

assess this, we used slow centrifugation and calculated the percent mucoviscosity as the change in supernatant optical density at 600 nm as previously described[14,19,51].

We found that most strains did not display high percent mucoviscosity, yet *ca.* 15% of strains exhibited high mucoviscosity (Supplemental Fig. 3A). The long tail in the distribution was similar to that of biofilm formation (Supplemental Fig. 1A). Therefore, we assessed correlations between percent mucoviscosity and biofilm formation. First, we verified whether phylogeny strongly bias the analyses using both Pagel's λ and Bloomberg's K calculations[40,41]. We did not detect a strong phylogenetic signal in mucoviscosity and biofilm formation (Table 2). This could be explained by extensive horizontal gene transfer of virulence factors in these species, including the above mentioned *rmp* locus encoded in the hypervirulence plasmid[10]. Overall, there was not a significant correlation with mucoviscosity and biofilm formation when assessing all isolates (Supplemental Fig. 4B), or when considering virulence as a covariate independently (Supplemental Figs. 4C and 4D). We found that although K1 and K2 capsule serotypes are often associated with increased mucoviscosity[9], among the 20% most mucoid isolates, only one had a K2 capsule serotype (Fig. 3a). Surprisingly, all isolates with a K3 capsule ($N = 4$) were among the most viscous, and significantly so ($X^2 = 11.8$, $P < 0.001$). Furthermore, this set of isolates included three of the four isolates from the collection that have the *rmpACD* genetic element shown to regulate the hypermucoviscous phenotype (indicated with asterisk)[48]. The analysis of the isolates with highest percent mucoviscosity and most robust biofilm formers suggest a trade-off between the two traits, with only three isolates displaying increased percent mucoviscosity being able to form high biofilm (Fig. 3b).

For a robust analysis, we included another readout of encapsulation and analyzed the isolates migration on a Percoll density gradient[52]. This approach was selected over the more traditional uronic acid quantification due to the large diversity of uronic acid content of each serotype, making difficult comparison across serotypes. Our results were in line with the percent mucoviscosity results as we observed none of the most mucoid isolates migrated to the bottom of the Percoll gradient, indicating high capsule production in this set of isolates; while the least mucoid isolates had the most isolates that migrated to the bottom of the gradient, indicating low levels of encapsulation (Supplemental Fig. 5). Surprisingly, the two most mucoid isolates MRSN 21352 and 607210 migrated less than the respective K1 and K2 capsule serotype hypermucoviscous controls *K. pneumoniae* NTUH K2044 and KPPR1S (Fig. 3c). Together, our results suggest

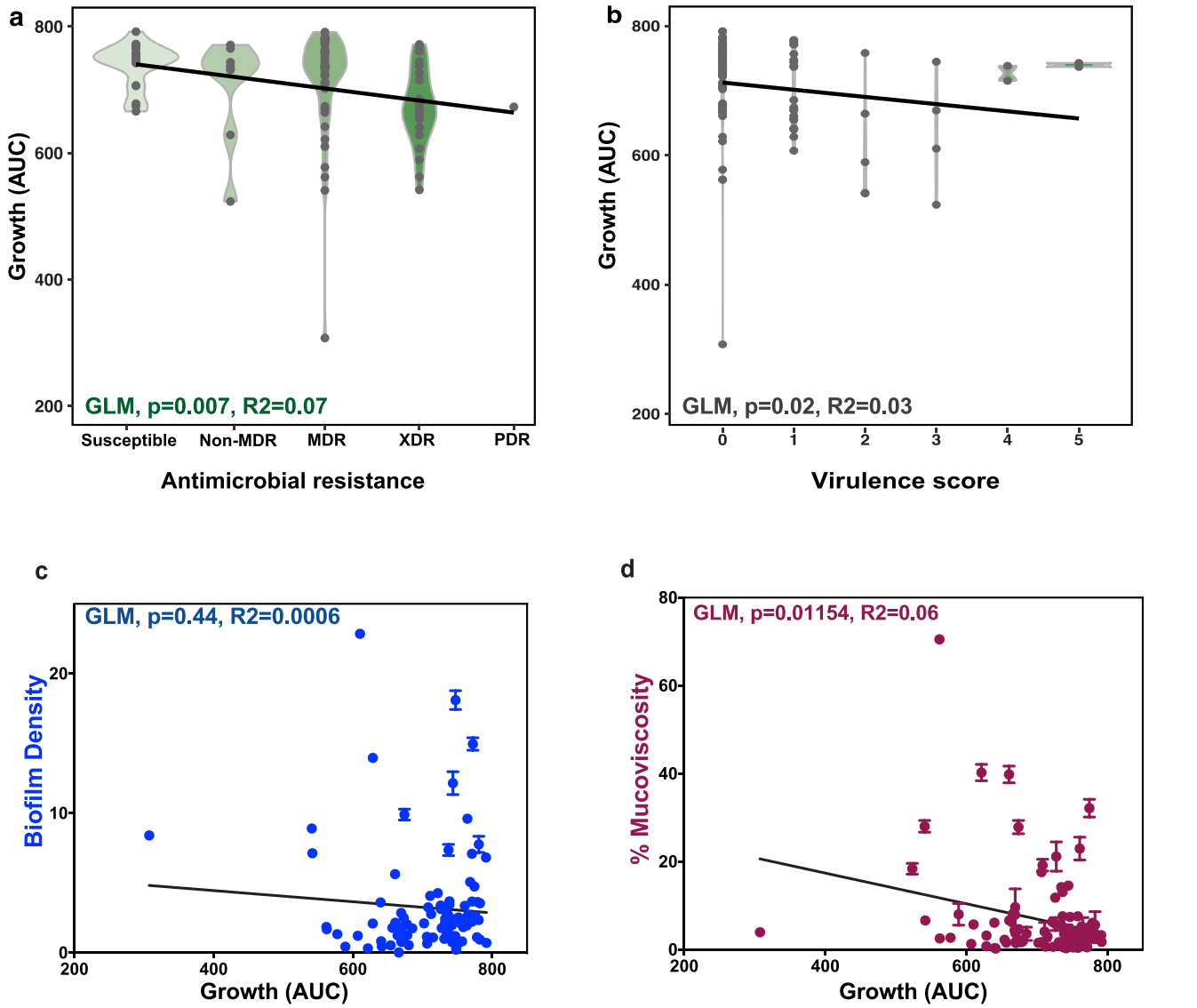

Fig. 4 | Mucoviscosity and antimicrobial resistance impact growth rate. The figures show area under the curve (AUC) of the growth rate analyses of the collection compared to potential covariates (antimicrobial resistance, virulence, biofilm and mucoviscosity). Figure 4a and 4b show the AUC for isolates with different antimicrobial resistance levels (MDR, multi-drug resistance; XDR, extensive-drug resistance; and PDR, pan-drug resistance) and different virulence scores (0–5), respectively. Figure 4c and 4d show the AUC compared to biofilm density normalized to OD_{550} and percent mucoviscosity, respectively. General linear model was used to assess *p*-value and R^2 values shown.

mucoviscosity limits biofilm formation, but highly mucoid strains can bypass this limitation to form robust biofilms.

Mucoviscosity and antimicrobial resistance impact growth rate

Previous work has highlighted the impact of population yield on biofilm formation[22]. Therefore, we wanted to determine if these findings applied more broadly than previously reported, and if like what was observed with our percent mucoviscosity and biofilm comparison, subsets of this collection had variable correlation with growth yield and mucoviscosity or biofilm. To test the growth rate of the MRSN diversity panel we performed growth curve analyses with the isolates to determine the generation time, maximum yield, and AUC. These parameters were then compared to the antibiotic resistance profile, virulence score, biofilm density, and percent mucoviscosity. Interestingly, we found that the growth rate parameters were much less diverse across the isolates than biofilm density (Table 1).

As expected, we observe a correlation between mucoviscosity and virulence score (GLM, *p*-value = <0.0001). Overall, we saw a negative correlation with all parameters tested when compared to AUC (Fig. 4). Specifically, when comparing the drug resistance status with the growth curve analysis, we found a significant (GLM, *p*-value = 0.007; $R^2 = 0.07$) negative correlation with drug resistance acquisition and AUC (Fig. 4a). This trend is further supported by a negative correlation (GLM, *p*-value = 0.009, $R^2 = 0.06$) between the maximum yield and drug resistance, and a positive correlation with generation time and drug resistance (GLM, *p*-value = 0.002; $R^2 = 0.09$) (Supplemental Fig. 6). Similar to antibiotic resistance, virulence factors such as mucoviscosity have been reported to be costly to the growth rate[34]. We found that an increase in virulence score also correlated with diminished growth (Fig. 4b), as estimated by the AUC (GLM, *p*-value = 0.02; $R^2 = 0.03$), and further validated by the positive correlation between generation time and virulence (GLM, *p*-value = 0.061; $R^2 = 0.18$) (Supplemental Fig. 6).

Next, we compared our growth analyses to the biofilm formation capabilities and percent mucoviscosity, as previous work has shown that biofilm formation can be impacted by the growth rate of the population[22]. We found a negative correlation between AUC and both biofilm formation (GLM, *p*-value = 0.44; $R^2 = 0.0006$) (Fig. 4c) and percent mucoviscosity

(GLM, p-value = 0.01154, R^2 = 0.06) (Fig. 4d). Interestingly, MRSN 564304 had a substantial growth defect when compared to representative strains, fast growers, the most mucoid strain, and the greatest biofilm former (Supplemental Fig. 7). With the concern that this outlier with an extremely slow growth rate may be skewing the significance of the correlation considering the sensitivity of Pearson's correlations, we removed it from the analyses and reassessed correlation with AUC and mucoviscosity. This resulted in no qualitative difference (GLM, p-value = 0.001, R^2 = 0.1). Interestingly, this strain was within our set of strong biofilm formers and was the one isolate that appeared dark red on Congo Red agar and dark blue on calcofluor white agar (Supplemental Fig. 2). These data suggest that mucoviscosity may affect the growth rate of the isolates, but growth rate does not affect biofilm formation.

Fimbriae mutations differential impact on biofilm density and mucoviscosity

It has recently been discovered that mutations within the gene encoding for the tip-adhesin of type III fimbriae (*mrkD*) or within the switch of type I fimbriae (*fimH*) impact biofilm formation in a capsule-dependent manner[22]. Furthermore, a recent study revealed interruption by an insertion sequence (IS) of *mrkH*, encoding for a c-di-GMP transcriptional activator, resulted in decreased biofilm formation[34,53]. In addition, another chaperon-usher system tip adhesin EcpD has been shown to be important for adherence to epithelial cells[27]. Therefore, to capitalize on the diversity of sequenced isolates within this collection, we next aimed to assess the fimbriae allele variations to determine impacts on biofilm.

We compared the *mrkH*, *mrkD, and fimH* genes to identify variations from the most common alleles (hereafter named as 'reference') and found a variety of alleles within the collection as well as isolates with the genetic elements absent from the genome (Fig. 1). For *ecpD*, we found that it was most commonly absent than present in the genomes, thus we refrained from using the term reference, and used the position found to be most commonly altered as nomenclature (V18A and V18V). First, we verified that, as expected, biofilm and mucoviscosity does not correlate with genome size (Spearman, p-value = 0.89 and 0.09, rho = 0.01 and −0.16 for biofilm and mucoviscosity, respectively). Overall, there was a clear genotype-phenotype correlation with the three isolates (MRSN 560539, 375436, and 730567) that have insertion sequences in the *mrkH* gene and are all deficient in biofilm formation (Kruskal–Wallis, p-value = 0.01) (Fig. 5a). In addition, there were many isolates with allelic variation in *mrkD* (10 different alleles with the dominant reference allele being present in 50 different isolates and Q141E present in 44 different isolates). Indeed, this mutation emerged early in the life history of *K. pneumoniae* and is present in most isolates from the first clade (Fig. 1). Interestingly, *mrkD* allelic variations did not show a correlation with biofilm changes but did have a minor impact on mucoviscosity (Kruskal–Wallis, p-value = 0.07). However, the importance of this fimbriae system for biofilm formation is revealed by the two strains lacking either a portion or all of the *mrk* operon (MRSN 562722 and 21304, respectively) being among the lowest biofilm formers (Fig. 1 and Supplemental Fig. 1). We observed a larger diversity of FimH alleles, a total of 17 different sequences, yet the reference allele was by far the most common, found in 81 isolates. The second most common allele was V193I found in five different isolates followed by G267S that was only present in three different isolates. No significant change in biofilm or mucoviscosity was seen with these mutations. Similar to the isolates with larger mutations of the *mrk* operon, we saw limited biofilm formation with MRSN 581745 that has an 829 base pair deletion between *fimG* and *fimH* (Supplemental Fig. 1). Finally, we did not identify a homolog of *ecpD* tip adhesin gene in almost half of the isolates. When present, the position with most polymorphism was amino acid 18 located near the signal peptide (Figs. 1 and 5a). Interestingly, even though there was no overall impact (Kruskal–Wallis, p-value = 0.46) of allele variations in the *ecpD* gene on mucoviscosity, those isolates with the V18A mutation had a wide range of mucoviscosity.

We found it extremely compelling that the Q141E allelic variation was found in 38 isolates as a single mutation and in six isolates with additional non-synonymous mutations in *mrkD* (Fig. 5b), many of which clustered together on the phylogenetic tree (Fig. 1). The Q141E mutation is located within the lectin binding domain of MrkD, shown to be important for binding affinity (Fig. 5c)[22]. Interestingly, the strongest biofilm former of the collection (MRSN 731029 collected from a urine sample in Asia) had a L133I mutation on the opposite side of the lectin binding domain (Fig. 5c). Comparatively, the mutations in *ecpD* occurred not only in the lectin domain but also very near to the signal peptide (Supplemental Fig. 8). These mutations could influence the export of the tip adhesin and thus binding abilities. Our observations suggest mutations within the tip adhesion can have differential effects on the isolate's biofilm formation and mucoviscosity but there is a strong correlation with a loss of biofilm formation associated with the *mrkH* insertion sequences disabling the c-di-GMP activator of the type III fimbriae.

Environmental shear flow influences the biofilm formation potential and spatial distribution in strong biofilm formers

With the diversity of attributes of the top biofilm formers we wanted to visualize their biofilm compositions to learn more about biofilm structure and potential heterogeneity in morphology. The biofilms of the highest biofilm formers (Fig. 2a) were grown in static conditions with the same growth media to mirror our crystal violet staining and in microfluidic conditions to understand the impact of environmental shear flow. To introduce environmental shear flow, we grew them under flow rate of 65 μL h^{-1} for 24-h in a microfluidic 48-well plate that has a confocal microscopy-compatible glass bottom under the channels. Due to the increased serpentine clogging encountered when growing our *K. pneumoniae* isolates in the microfluidic plate using the TSB with 0.5% glucose we used M9 minimal media with 0.4% glucose as previously described for *Escherichia coli*[38,54]. With both biofilm growth conditions, the bacterial population and polysaccharide matrix was stained with SYTO 9 (green) and calcofluor white (blue), respectively. Confocal z-stack imaging facilitated the 3D rendering images to visualize the height and composition of the biofilms formed by each isolate.

We chose crystal violet staining as our measure for the MRSN diversity set biofilm formation because it is a robust, well studied method for high-throughput biofilm assessments; yet, there is no distinction between the cells and the matrix materials because crystal violet stains both without bias[27]. Therefore, considering the diversity of the phenotypes of our 15 top biofilm-forming isolates we wanted to visualize the bacterial cells independently from the matrix polysaccharides and determine how these components change when the biofilms are grown under environmental shear flow. We found that in both static and microfluidic conditions, all isolates formed biofilms with average height between 30 and 50 μm and were found to have some level of polysaccharides within their matrix (Fig. 6 and Supplemental Fig. 9). Specifically, when grown in static conditions, MRSN 731029 and 564304 displayed the most biofilm height (~50 μm), although 564304 had more polysaccharide matrix (Fig. 6a, b). These isolates when viewed from the top of the matrix revealed the cellular population as aggregates in microcolonies, compared to MRSN 16008, 5741, and 513382 that had uniformly dispersed cellular population when viewed through the z-stacks (Fig. 6c–e). Interestingly, MRSN 16008 had a thin layer of polysaccharide matrix with an abundance of cells above and below but the mucoid isolate MRSN 5741 had two distinct layers of polysaccharide matrix with less cells outside of the matrix. MRSN 513382 had a robust layer of polysaccharide matrix but decreased matrix height compared to the other isolates within the top biofilm formers (Fig. 6e). The cellular staining of MRSN 1912 with SYTO 9 was limited and we hypothesized this was due to the thickness of the matrix layer. To test this, we visualized biofilms grown with MRSN 1912 harboring pSL6_RFP to allow for constitutive RFP expression. We saw a slight increase in the cellular population under the matrix cap, although this was still minimal compared to the other isolates (Supplemental Fig. 10).

When testing the effect of shear force on biofilm formation of the isolates we found both the microcolony formation (Fig. 6a, b) and the polysaccharide dense layers were no longer present (Fig. 6c–f). Strikingly,

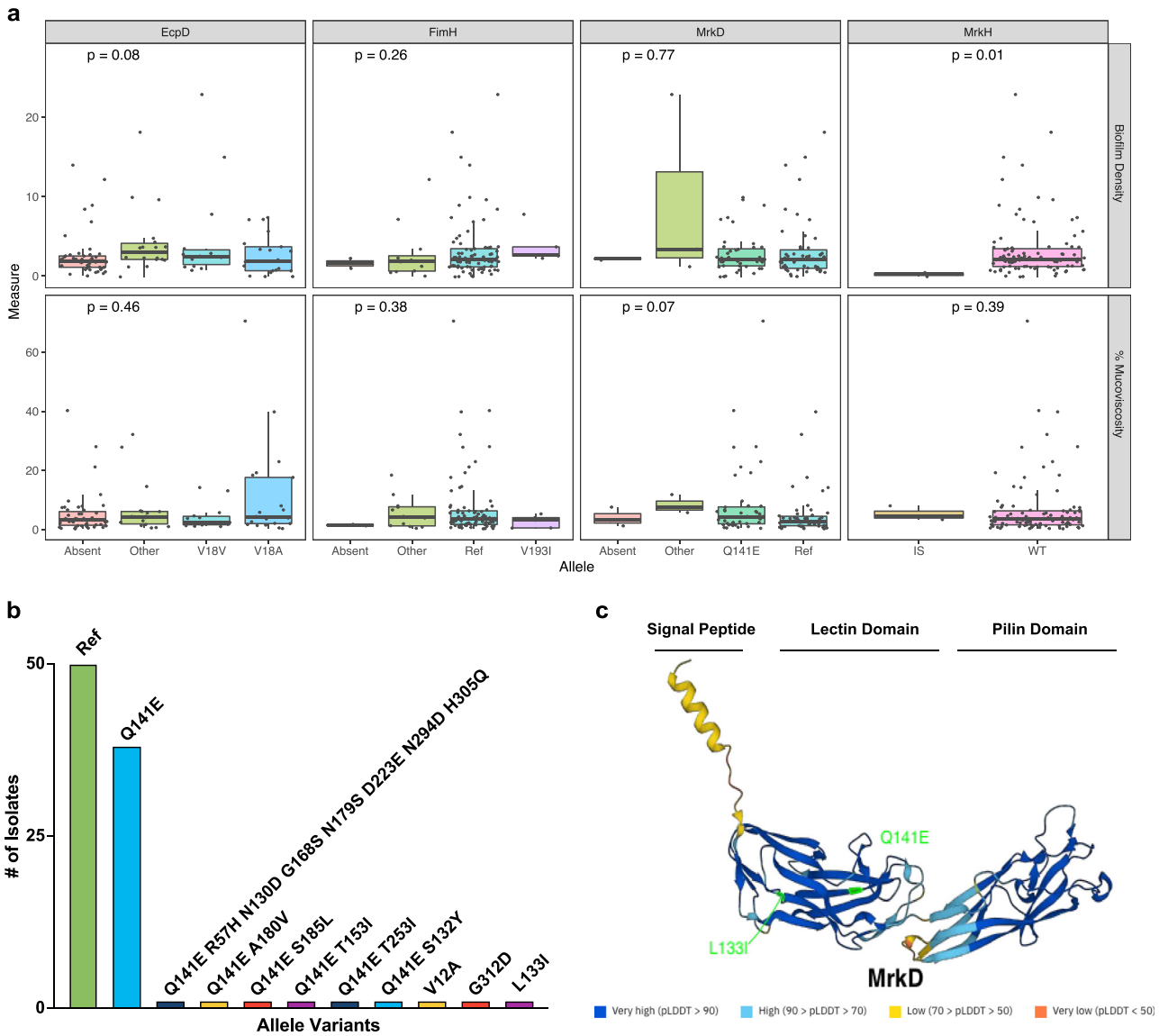

Fig. 5 | Fimbriae mutations differential impact on biofilm density and mucoviscosity. The figures show the allele variations within the fimbriae genes and their phenotypic biofilm density normalized to OD_{550} and percent mucoviscosity. Figure 5a shows the fimbriae allele variations and their impact on biofilm density and percent mucoviscosity. Kruskal–Wallis statistical analysis was used to determine the *p* values shown. Figure 5b shows the number of isolates within the collection that have each of the labeled mutation. Figure 5c is the AlphaFold generated structure of the MrkD parental protein with the signal peptide, lectin, and pilin domains labeled. The amino acids at position 141 and 133 are shown in bright green to display the localization of the mutation from glutamate amino acid to glutamic acid and leucine amino acid to isoleucine, respectively. The shades of blue in the predicted MrkD protein structure correspond to the pLDDT, which is the AlphaFold score for confidence per residue.

MRSN 731029 and 564304 still had the most cellular density, although the MRSN 564304 displayed large gaps in the matrix when viewed from the top. MRSN 16008 had an abundant amount of polysaccharide matrix but less cellular density compared to MRSN 731029 and 564304. In addition, the two mucoid isolates that were able to form robust biofilms from our crystal violet readout responded to shear force differently. MRSN 5741 had much less biofilm height and cellular staining when grown under shear flow (Fig. 6d), but MRSN 1912 had increased biofilm height and cellular staining (Fig. 6f). These results reveal the diversity of spatial distribution between strong biofilm formers and the impact of shear flow has differential effects on mucoid isolates that form robust biofilms.

Discussion

The MRSN diversity panel is a set of 100 *K. pneumoniae* isolates that include a broad spectrum of the MLST identified[25]. With such a large set of diverse isolates, investigating the covariates important for biofilm formation has potential to expand our current knowledge of *K. pneumoniae* biofilm formation abilities. The diversity and versatility of the *K. pneumoniae* species could preclude predictive analyses. Our findings suggest an important phenotypic diversity in this set of isolates and revealed an intricate relationship between growth, mucoviscosity, and specific genetic determinants. Furthermore, the large diversity revealed here raises questions concerning: (1) the relative roles of natural selection outside a host or during gut commensalism, (2) the potential alteration of mutation rates during infection, and (3) how these factors impact the capacity of *K. pneumoniae* to persist.

We found that the life history or origin of the isolates allowed for the selection of biofilm formation based on the structural organization of the environment. For example, there was a diverse range of biofilm formation abilities within the isolates collected from the urine, including the strongest biofilm former of the entire collection (Fig. 2b). This reflects the importance of the selection pressure for biofilm formation for bacteria in this environment, because establishing a urinary tract infection is largely impacted by

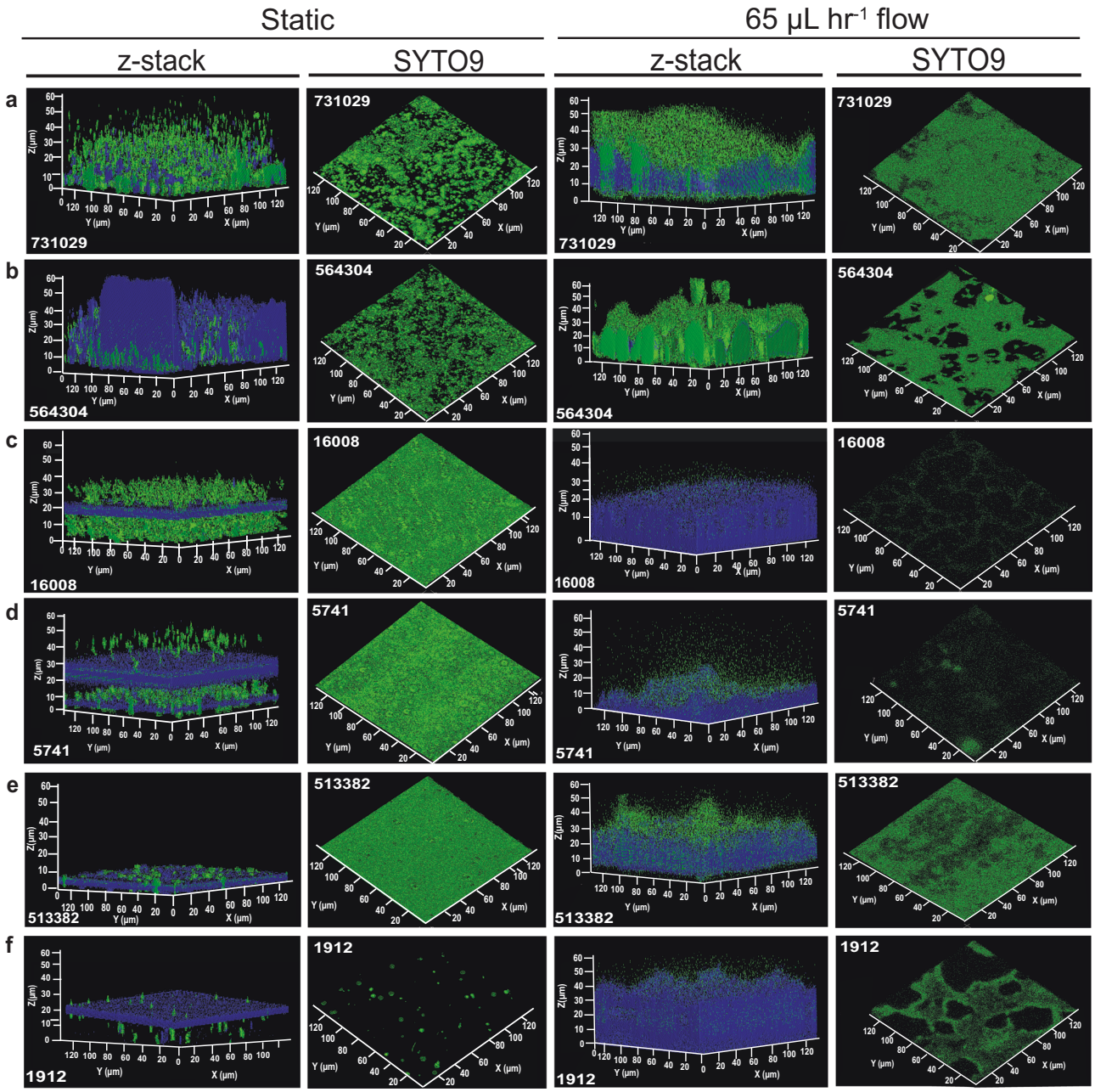

Fig. 6 | Environmental shear flow influences the biofilm formation potential and spatial distribution in strong biofilm formers. The figures show the 3D rendering of confocal z-stack imaging of six of the top biofilm forming isolates stained with SYTO 9 (green) to visualize the cells and calcofluor white (blue) to visualize the polysaccharide matrix, respectively. Figure 6a–f show MRSN 731029, 564304, 16008, 5741, 513382, and 1912 biofilms, respectively. Each isolate is shown grown under static conditions (Left side) and with 65 µl h^{-1} shear flow (Right side). For each isolate and with both conditions the z-stack images are shown to visualize the height of the biofilm and images from the top of the biofilm are shown to visualize the overall cellular density. $n = 3$ biofilms were grown and imaged with a representative image shown.

the ability to form biofilms both on indwelling devices and within the urinary tract[55]. Furthermore, isolates that came from very structured environments (i.e., wound) displayed strong biofilm formation and those that came from a less structured sample site (i.e., blood) had less biofilm formation. Therefore, our work suggests that the isolates were adapted during or prior to the infection to the environment where they were isolated. This finding builds upon the previous work that revealed the structure of the environment largely impacts population evolution[34].

Intriguingly, we found that percent mucoviscosity is not restricted to the K1/K2 serotype isolates nor exclusively associated to the presence of the *rmp* locus, the regulator of the mucoid phenotype in *K. pneumoniae* as 17 isolates were mucoid in the absence of such genetic element (Fig. 3a)[9]. Two isolates, MRSN 5881 (K62) and 736213 (K2) encoded for *rmpA1* and *rmpA2*, respectively, whereas three other isolates (MRSN 16233 (K2); 582610 (K20); and 752729 (K64)) carried both alleles. None of these isolates were strong biofilm formers and only three of these isolates (MRSN 16233, 5881, and 582610) displayed high percent mucoviscosity (Fig. 3a, marked with asterisk). The observation of increased mucoviscosity in the absence of *rmpA1* and *rmpA2* and in diverse capsular serotypes suggests there are other genetic determinants of mucoviscosity in *K. pneumoniae*. However, with such a diverse collection, there is a possibility of sampling bias due to the limited number of each capsular serotype included in the set. Therefore, continued investigation with more isolates with the same K serotype (K1, K2, and

K3) are needed to identify a trend for the serotype with the highest mucoviscosity. We show that within the top 15 biofilm formers only three are mucoid and most of the isolates with high percent mucoviscosity do not form robust biofilms (Fig. 3b). The identification of these outliers suggests these isolates can bypass the limitations of increased mucoviscosity on biofilm formation.

When looking at growth as a covariant, we found growth did not impact biofilm, as there was no significant correlation identified and one of the strongest biofilm formers was shown to have a noteworthy growth defect (Fig. 4c and Supplemental Fig. 7C). When considering the other parameters of our study, we found that antibiotic resistance and mucoviscosity had a significant negative impact on the growth rate, but the virulence score was not associated with the growth rate (Fig. 4a, b, d). These findings reveal the energetic burden antimicrobial resistance and increased mucoviscosity have on growth rate and suggests that the strict regulation of virulence traits can mitigate the effects on growth rate. This finding is important when considering the population dynamics suggesting that when removed from an environment where mucoviscosity or antibiotic resistance is not necessary, they may be outcompeted and less prone to spread a worrisome trait. When considering the fimbriae allele variants we found that, in line with recent work[34,53], insertion sequences in *mrkH* led to decreased biofilm formation, as all three isolates with this insertion displayed limited biofilm formation (Fig. 5). Interestingly, the mutations within *mrkD* had more of an effect on mucoviscosity than biofilm formation. Furthermore, the independent emergence of the Q141E allele throughout the life history of *K. pneumoniae* suggests it could provide a strong fitness advantage in the clinic. Alternatively, it could imply that it could be hitchhiking along with several other mutational events. Conversely, *ecpD* allele variation impacted biofilm formation (p-value = 0.08) but not mucoviscosity (p-value = 0.46). Specifically, mutations within the tip adhesion showed increased biofilm formation, while a lack of an *ecpD* homolog within the genome led to a decrease in biofilm formation.

When visualizing the strongest biofilm formers using confocal microscopy we continued to see suggestions of adaptation to the environment. In particular, isolated from a respiratory sample and having high percent mucoviscosity MRSN 5741 displayed limited biofilm formation ability under 65 $\mu L\ h^{-1}$ shear flow compared to static growth. Conversely, MRSN 1912 was collected from a perianal sample and although it had high percent mucoviscosity was able to form a more robust biofilm in the presence of environmental shear flow, although the cellular population was not as dense as MRSN 731029 and 564304 isolated from urine samples. Although MRSN 564304 retained biofilm height with shear flow, this isolate had a striking decrease in polysaccharide abundance, consistent with its dark phenotype when grown on Congo Red and calcofluor white agar (Supplemental Fig. 2). This could be suggestive that the polysaccharide abundance observed in our static growth conditions are being washed away under shear flow. The confocal analysis of the strong biofilm formers under environmental shear flow not only suggests environmental adaptation but reveals the limitations of biofilm analyses using a single condition.

In conclusion, our investigation of the MRSN diversity panel has revealed a striking quantitative and morphological variation in the biofilm formation abilities within this set, reflecting the diversity of genetic components impacting biofilm growth. We found that mucoviscosity and antibiotic resistance largely impact the growth parameters of the isolate, while the tight regulation of virulence may offset effects on growth potential.. Our work has significantly improved our knowledge of the covariates of biofilm formation and our understanding of the genetic variabilities of the attributes important for *K. pneumoniae*.

Data availability

The phylogenetic analysis done for this manuscript was deposited to be made publicly available and consultable at microreact (https://microreact.org/).

Received: 24 June 2024; Accepted: 5 December 2024;
Published online: 19 December 2024

References

1. Effah, C. Y., Sun, T., Liu, S. & Wu, Y. Klebsiella pneumoniae: an increasing threat to public health. *Ann. Clin. Microbiol. Antimicrob.* **19**, 1, https://doi.org/10.1186/s12941-019-0343-8 (2020).
2. Kot, B. et al. Virulence analysis and antibiotic resistance of Klebsiella pneumoniae isolates from hospitalised patients in Poland. *Sci. Rep.* **13**, 4448, https://doi.org/10.1038/s41598-023-31086-w (2023).
3. Wyres, K. L. & Holt, K. E. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria.*Curr. Opin. Microbiol.* **45**, 131–139, https://doi.org/10.1016/j.mib.2018.04.004 (2018).
4. Gonzalez-Ferrer, S. et al. Finding order in the chaos: outstanding questions in Klebsiella pneumoniae pathogenesis. *Infect. Immun.* **89**, e00693–20, https://doi.org/10.1128/IAI.00693-20 (2021).
5. Paczosa, M. K. & Mecsas, J. Klebsiella pneumoniae: going on the offense with a strong defense. *Microbiol. Mol. Biol. Rev.* **80**, 629–661, https://doi.org/10.1128/MMBR.00078-15 (2016).
6. Zhang, J. et al. Mobilizable plasmids drive the spread of antimicrobial resistance genes and virulence genes in Klebsiella pneumoniae. *Genome Med.* **15**, 106, https://doi.org/10.1186/s13073-023-01260-w (2023).
7. Ramirez M. S., Traglia G. M., Lin D. L., Tran T., Tolmasky M. E. Plasmid-mediated antibiotic resistance and virulence in gram-negatives: the klebsiella pneumoniae paradigm. *Microbiol. Spectr.* **2**. https://doi.org/10.1128/microbiolspec.PLAS-0016-2013 (2014).
8. Wyres, K. L. et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. *PLoS Genet.* **15**, e1008114, https://doi.org/10.1371/journal.pgen.1008114 (2019).
9. Walker, K. A. & Miller, V. L. The intersection of capsule gene expression, hypermucoviscosity and hypervirulence in Klebsiella pneumoniae. *Curr. Opin. Microbiol.* **54**, 95–102 (2020).
10. Cheng, H. Y. et al. RmpA regulation of capsular polysaccharide biosynthesis in Klebsiella pneumoniae CG43. *J. Bacteriol.* **192**, 3144–3158, https://doi.org/10.1128/JB.00031-10 (2010).
11. Yeh, K. M. et al. Capsular serotype K1 or K2, rather than magA and rmpA, is a major virulence determinant for Klebsiella pneumoniae liver abscess in Singapore and Taiwan. *J. Clin. Microbiol.* **45**, 466–471, https://doi.org/10.1128/JCM.01150-06 (2007).
12. Centers for Disease Control and Prevention (U.S.), Unit NCfEZaIDUSDoHQPARCaS. Antibiotic resistance threats in the United States, 20192019. https://doi.org/10.15620/cdc:82532.
13. Kochan, T. J. et al. Klebsiella pneumoniae clinical isolates with features of both multidrug-resistance and hypervirulence have unexpectedly low virulence. *Nat. Commun.* **14**, 7962, https://doi.org/10.1038/s41467-023-43802-1 (2023).
14. Russo, T. A. et al. Differentiation of hypervirulent and classical Klebsiella pneumoniae with acquired drug resistance. *mBio* **15**, e0286723, https://doi.org/10.1128/mbio.02867-23 (2024).
15. Guerra, M. E. S. et al. Klebsiella pneumoniae biofilms and their role in disease pathogenesis. *Front. Cell Infect. Microbiol.* **12**, 877995, https://doi.org/10.3389/fcimb.2022.877995 (2022).
16. Guilhen, C. et al. Colonization and immune modulation properties of Klebsiella pneumoniae biofilm-dispersed cells. *NPJ Biofilms Microbiomes* **5**, 25, https://doi.org/10.1038/s41522-019-0098-1 (2019).
17. Heindl, J. E. et al. Mechanisms and regulation of surface interactions and biofilm formation in Agrobacterium. *Front. Plant Sci.* **5**, 176, https://doi.org/10.3389/fpls.2014.00176 (2014).
18. Benincasa, M. et al. Biofilms from Klebsiella pneumoniae: matrix polysaccharide structure and interactions with antimicrobial peptides. *Microorganisms* **4**, 26, https://doi.org/10.3390/microorganisms4030026 (2016).

19. De Los Santos L. et al. Polyproline peptide targets Klebsiella pneumoniae polysaccharides to collapse biofilms. *Cell Rep. Phys. Sci.* 2024;5. https://doi.org/10.1016/j.xcrp.2024.101869.
20. Zheng, J. X. et al. Biofilm formation in Klebsiella pneumoniae bacteremia strains was found to be associated with CC23 and the presence of wcaG. *Front. Cell Infect. Microbiol.* **8**, 21, https://doi.org/10.3389/fcimb.2018.00021 (2018).
21. Schembri, M. A., Blom, J., Krogfelt, K. A. & Klemm, P. Capsule and fimbria interaction in Klebsiella pneumoniae. *Infect. Immun.* **73**, 4626–4633, https://doi.org/10.1128/IAI.73.8.4626-4633.2005 (2005).
22. Nucci, A., Rocha, E. P. C. & Rendueles, O. Latent evolution of biofilm formation depends on life-history and genetic background. *NPJ Biofilms Microbiomes* **9**, 53, https://doi.org/10.1038/s41522-023-00422-3 (2023).
23. Buffet, A., Rocha, E. P. C. & Rendueles, O. Nutrient conditions are primary drivers of bacterial capsule maintenance in Klebsiella. *Proc. Biol. Sci.* **288**, 20202876, https://doi.org/10.1098/rspb.2020.2876 (2021).
24. Nirwati, H. et al. Biofilm formation and antibiotic resistance of Klebsiella pneumoniae isolated from clinical samples in a tertiary care hospital, Klaten, Indonesia. *BMC Proc.* **13**, 20, https://doi.org/10.1186/s12919-019-0176-7 (2019).
25. Martin, M. J. et al. A panel of diverse Klebsiella pneumoniae clinical isolates for research and development. *Micro. Genom.* **9**, mgen000967, https://doi.org/10.1099/mgen.0.000967 (2023).
26. Chen, C., Khaleel, S. S., Huang, H. & Wu, C. H. Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol. Med.* **9**, 8, https://doi.org/10.1186/1751-0473-9-8 (2014).
27. Alcantar-Curiel, M. D. et al. Multi-functional analysis of Klebsiella pneumoniae fimbrial types in adherence and biofilm formation. *Virulence* **4**, 129–138, https://doi.org/10.4161/viru.22974 (2013).
28. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, https://doi.org/10.1093/bioinformatics/btu153 (2014).
29. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693, https://doi.org/10.1093/bioinformatics/btv421 (2015).
30. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274, https://doi.org/10.1093/molbev/msu300 (2015).
31. Lam, M. M. C. et al. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. *Nat. Commun.* **12**, 4188, https://doi.org/10.1038/s41467-021-24448-3 (2021).
32. Wyres, K. L. et al. Identification of Klebsiella capsule synthesis loci from whole genome data. *Micro. Genom.* **2**, e000102, https://doi.org/10.1099/mgen.0.000102 (2016).
33. R Core Team R. R: A language and environment for statistical computing 2013.
34. Nucci, A., Janaszkiewicz, J., Rocha, E. P. C. & Rendueles, O. Emergence of novel non-aggregative variants under negative frequency-dependent selection in Klebsiella variicola. *Microlife* **4**, uqad038, https://doi.org/10.1093/femsml/uqad038 (2023).
35. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589, https://doi.org/10.1038/s41586-021-03819-2 (2021).
36. Mihaly Varadi ea. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkad1011 (2024).
37. Elston, K. M. et al. The Pathfinder plasmid toolkit for genetically engineering newly isolated bacteria enables the study of Drosophila-colonizing Orbaceae. *ISME Commun.* **3**, 49, https://doi.org/10.1038/s43705-023-00255-3 (2023).
38. M9 minimal medium standard. Cold Spring Harb Protocols. 2010 (2010:pdb.rec12295). https://doi.org/10.1101/pdb.rec12295.
39. Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (adn other things). *PeerJ* **12**, e16505, https://doi.org/10.7717/peerj.16505 (2024).
40. Blomberg, S. P., Garland, T. Jr. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745, https://doi.org/10.1111/j.0014-3820.2003.tb00285.x (2003).
41. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884, https://doi.org/10.1038/44766 (1999).
42. Petrachi, T. et al. An alternative approach to investigate biofilm in medical devices: a feasibility study. *Int. J. Environ. Res. Public Health* **14**, 1587, https://doi.org/10.3390/ijerph14121587 (2017).
43. Barnhart, M. M. & Chapman, M. R. Curli biogenesis and function. *Annu. Rev. Microbiol.* **60**, 131–147, https://doi.org/10.1146/annurev.micro.60.080805.142106 (2006).
44. York, A. Biofilms: Naturally modified cellulose in bacterial biofilms. *Nat. Rev. Microbiol.* **16**, 123, https://doi.org/10.1038/nrmicro.2018.22 (2018).
45. Reichhardt, C. et al. Congo Red interactions with Curli-producing E. coli and native curli amyloid fibers. *PLoS ONE* **10**, e0140388, https://doi.org/10.1371/journal.pone.0140388 (2015).
46. Davis, R. T. & Brown, P. D. spoT-mediated stringent response influences environmental and nutritional stress tolerance, biofilm formation and antimicrobial resistance in Klebsiella pneumoniae. *APMIS* **128**, 48–60, https://doi.org/10.1111/apm.13006 (2020).
47. Balestrino, D., Ghigo, J. M., Charbonnel, N., Haagensen, J. A. & Forestier, C. The characterization of functions involved in the establishment and maturation of Klebsiella pneumoniae in vitro biofilm reveals dual roles for surface exopolysaccharides. *Environ. Microbiol.* **10**, 685–701, https://doi.org/10.1111/j.1462-2920.2007.01491.x (2008).
48. Walker, K. A. et al. A Klebsiella pneumoniae regulatory mutant has reduced capsule expression but retains hypermucoviscosity. *mBio* **10**, e00089–19, https://doi.org/10.1128/mBio.00089-19 (2019).
49. Russo, T. A. & Marr, C. M. Hypervirulent Klebsiella pneumoniae. *Clin. Microbiol. Rev.* **32**, e00001–e00019, https://doi.org/10.1128/CMR.00001-19 (2019).
50. Xu, Q., Yang, X., Chan, E. W. C. & Chen, S. The hypermucoviscosity of hypervirulent K. pneumoniae confers the ability to evade neutrophil-mediated phagocytosis. *Virulence* **12**, 2050–2059, https://doi.org/10.1080/21505594.2021.1960101 (2021).
51. Fleeman, R. M. & Davies, B. W. Polyproline peptide aggregation with Klebsiella pneumoniae extracellular polysaccharides exposes biofilm associated bacteria. *Microbiol. Spectr.* **10**, e0202721, https://doi.org/10.1128/spectrum.02027-21 (2022).
52. Feltwell T., Dorman M. J., Goulding D. A., Parkhill J., Short F. L. Separating bacteria by capsule amount using a discontinuous density gradient. *J. Vis. Exp.* 2019. https://doi.org/10.3791/58679.
53. Wilksch, J. J. et al. MrkH, a novel c-di-GMP-dependent transcriptional activator, controls Klebsiella pneumoniae biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathog.* **7**, e1002204, https://doi.org/10.1371/journal.ppat.1002204 (2011).
54. Tremblay, Y. D. N. et al. High-throughput microfluidic method to study biofilm formation and host-pathogen interactions in pathogenic Escherichia coli. *Appl. Environ. Microbiol.* **81**, 2827–2840, https://doi.org/10.1128/AEM.04208-14 (2015).
55. Hatt, J. K. & Rather, P. N. Role of bacterial biofilms in urinary tract infections. In: Romeo, T. (ed.) Bacterial Biofilms. Berlin, Heidelberg: p. 163-92 (Springer Berlin Heidelberg, 2008).

Acknowledgements

The authors wish to acknowledge Dr. Patrick T McGann, Francois Lebreton, Melissa Martin, and Yoon Im Kwak from Walter Reed Army Institute of Research

for sharing the raw genetic information from their phylogenetic analysis performed in their 2023 publication[25]. This work is supported by the National Institutes of Health R00AI163295 to RMF. OR is supported by the grant from ANR (Agence national de recherche) grant [ANR 22 CE20 00181 BETinCAP].

Author contributions

R.M.F. conceived the idea and designed the experimentation. The biofilm and mucoviscosity experiments were performed by R.L.B. The phylogenetics analysis was done by E.C. and T.A. The growth curve assessment, fimbriae allele variant identification, and AMR, virulence and growth correlations were done by O.R. R.M.F. wrote the manuscript with input from other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41522-024-00629-y.

Correspondence and requests for materials should be addressed to Renee M. Fleeman.