Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health

Kathryn E. Holt[a,b,1], Heiman Wertheim[c,d], Ruth N. Zadoks[e,f], Stephen Baker[g], Chris A. Whitehouse[h], David Dance[d,i], Adam Jenney[b,j], Thomas R. Connor[k,l], Li Yang Hsu[m], Juliëtte Severin[n], Sylvain Brisse[o], Hanwei Cao[b,p], Jonathan Wilksch[b,p], Claire Gorrie[a,b,p], Mark B. Schultz[a], David J. Edwards[a], Kinh Van Nguyen[q], Trung Vu Nguyen[q], Trinh Tuyet Dao[q], Martijn Mensink[e], Vien Le Minh[g,r], Nguyen Thi Khanh Nhu[g,s], Constance Schultsz[g,t], Kuntaman Kuntaman[u], Paul N. Newton[d,i], Catrin E. Moore[d,i], Richard A. Strugnell[b,p], and Nicholas R. Thomson[k,v,1]

[a]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia; [b]Department of Microbiology and Immunology, The University of Melbourne, Parkville, VIC 3010, Australia; [c]Oxford University Clinical Research Unit, Wellcome Trust Major Overseas Programme, National Hospital for Tropical Diseases, Hanoi, Vietnam; [d]Nuffield Department of Clinical Medicine, Centre for Tropical Medicine, University of Oxford, OX3 7BN Oxford, United Kingdom; [e]Quality Milk Production Services, Cornell University, Ithaca, NY 14853; [f]Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, G12 8QQ Glasgow, United Kingdom; [g]The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam; [h]United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD 21702; [i]Lao-Oxford-Mahosot Hospital Wellcome Trust Research Unit, Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People's Democratic Republic; [j]Departmemt Infectious Diseases and Microbiology Unit, The Alfred Hospital, Melbourne, VIC 3004, Australia; [k]Pathogen Genomics, Wellcome Trust Sanger Centre, CB10 1SA Cambridge, United Kingdom; [l]Cardiff University School of Biosciences, Cardiff University, Cardiff, Wales, CF10 3AX, United Kingdom; [m]Department of Medicine, National University Health System, Singapore 119228; [n]Department of Medical Microbiology and Infectious Diseases, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands; [o]Microbial Evolutionary Genomics, Institut Pasteur, CNRS, UMR3525, Paris, France; [p]Peter Doherty Institute, The University of Melbourne, Parkville, VIC 3010, Australia; [q]National Hospital for Tropical Diseases, Hanoi, Vietnam; [r]Division of Infectious Diseases, Department of Medicine, University of California, San Francisco, CA 94118-6215; [s]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia; [t]Academic Medical Center, University of Amsterdam, 1012 WX Amsterdam, The Netherlands; [u]Department of Clinical Microbiology, Dr. Soetomo Academic Hospital – School of Medicine, Airlangga University, Surabaya, Jawa Timur, Indonesia; and [v]Department of Pathogen Molecular Biology, The London School of Hygiene and Tropical Medicine, WC1E 7HT London, United Kingdom

Edited by Simon A. Levin, Princeton University, Princeton, NJ, and approved May 11, 2015 (received for review January 16, 2015)

***Klebsiella pneumoniae* is now recognized as an urgent threat to human health because of the emergence of multidrug-resistant strains associated with hospital outbreaks and hypervirulent strains associated with severe community-acquired infections. *K. pneumoniae* is ubiquitous in the environment and can colonize and infect both plants and animals. However, little is known about the population structure of *K. pneumoniae*, so it is difficult to recognize or understand the emergence of clinically important clones within this highly genetically diverse species. Here we present a detailed genomic framework for *K. pneumoniae* based on whole-genome sequencing of more than 300 human and animal isolates spanning four continents. Our data provide genome-wide support for the splitting of *K. pneumoniae* into three distinct species, KpI (*K. pneumoniae*), KpII (*K. quasipneumoniae*), and KpIII (*K. variicola*). Further, for *K. pneumoniae* (KpI), the entity most frequently associated with human infection, we show the existence of >150 deeply branching lineages including numerous multidrug-resistant or hypervirulent clones. We show *K. pneumoniae* has a large accessory genome approaching 30,000 protein-coding genes, including a number of virulence functions that are significantly associated with invasive community-acquired disease in humans. In our dataset, antimicrobial resistance genes were common among human carriage isolates and hospital-acquired infections, which generally lacked the genes associated with invasive disease. The convergence of virulence and resistance genes potentially could lead to the emergence of untreatable invasive *K. pneumoniae* infections; our data provide the whole-genome framework against which to track the emergence of such threats.**

Klebsiella pneumoniae | genomics | virulence | antimicrobial resistance | population structure

The Gram-negative bacterium *Klebsiella pneumoniae* is a leading cause of hospital-acquired (HA) infections and neonatal sepsis globally (1–3). Widely considered an opportunistic pathogen, *K. pneumoniae* can be carried asymptomatically in the intestinal tract, skin, nose, and throat of healthy individuals (4, 5) but can also cause a range of infections in hospitalized patients, most commonly pneumonia, wound, soft tissue, or urinary tract infections. *K. pneumoniae* infections are particularly a problem among neonates, the elderly, and the immunocompromised (4) but also cause significant numbers of serious community-acquired (CA) infections, including pyogenic liver abscess, pneumonia, and meningitis (6). Virulence factors thought to be associated with invasive CA infections include various siderophores, specific polysaccharide capsule serotypes, and *rmpA* genes that are associated with hypermucoidy (7).

Significance

***Klebsiella pneumoniae* is rapidly becoming untreatable using last-line antibiotics. It is especially problematic in hospitals, where it causes a range of acute infections. To approach controlling such a bacterium, we first must define what it is and how it varies genetically. Here we have determined the DNA sequence of *K. pneumoniae* isolates from around the world and present a detailed analysis of these data. We show that there is a wide spectrum of diversity, including variation within shared sequences and gain and loss of whole genes. Using this detailed blueprint, we show that there is an unrecognized association between the possession of specific gene profiles associated with virulence and antibiotic resistance and the differing disease outcomes seen for *K. pneumoniae*.**

Author contributions: K.E.H., H.W., R.A.S., and N.R.T. designed research; K.E.H. performed research; H.W., R.N.Z., S. Baker, C.A.W., D.D., A.J., L.Y.H., J.S., S. Brisse, H.C., J.W., K.V.N., T.V.N., T.T.D., M.M., V.L.M., N.T.K.N., C.S., K.K., P.N.N., and C.E.M. contributed new reagents/analytic tools; K.E.H., T.R.C., C.G., M.B.S., D.J.E., and N.R.T. analyzed data; and K.E.H. and N.R.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the European Nucleotide Archive database (accession no. ERP000165).

[1]To whom correspondence may be addressed. Email: nrt@sanger.ac.uk or kholt@unimelb.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1501049112/-/DCSupplemental.

K. pneumoniae, particularly when hypermucoid, can cause invasive disease in several animal species (8, 9) and is a common cause of mastitis in dairy herds (10). Moreover it can thrive in a range of plant hosts and environmental niches, including water, soil, and plant matter (4, 5, 11). Although it is clear that *K. pneumoniae* is genetically and phenotypically diverse (12, 13), previous efforts to identify specific features that can distinguish human clinical isolates from plant, animal, or environmental isolates have yielded no markers of human-specific lineages (14). Three distinct phylogroups of *K. pneumoniae*—KpI, KpII, and KpIII—have been defined based on sequencing of a small number of genes (15, 16), and it has been proposed that these phylogroups be redesignated as distinct species, namely, *K. pneumoniae* (KpI), *K. quasipneumoniae* (KpII) (17), and *K. variicola* (KpIII) (18); however, all three cause infections in humans (15, 19).

Critically, the emergence of multiple drug-resistant (MDR) *K. pneumoniae* has been identified as an urgent threat to human health, featuring, for example, in recent reports on antimicrobial resistance (AMR) from the US Centers for Disease Control and Prevention (CDC) (20) and the UK Department of Health (21), because of a high prevalence of resistance to carbapenems and broad-spectrum β-lactams (22–25). The most notorious example of AMR *K. pneumoniae* is a lineage identified as clonal complex (CC) 258 by multilocus sequence typing (MLST) (13); CC258 frequently carries the *K. pneumoniae* carbapenemase (KPC) gene as well as numerous other acquired AMR genes and has been responsible for hospital outbreaks on several continents (13, 26, 27).

The tracking of AMR organisms is one of the four core actions proposed in the CDC AMR action plan to limit the emergence and spread of AMR bacteria. Several recent genomic analyses indicate that sequence type (ST) 258 is a recombinant strain that has undergone capsular exchange since its emergence as a cause of KPC outbreaks (28–30). However, little attention has been paid to other MDR clones, which also are common and can spread carbapenem resistance (31). Relatively little is known about this broader population of *K. pneumoniae*, and there remains a lack of data regarding transmission, pathogenicity, and the evolution and spread of MDR clones globally. Moreover, *K. pneumoniae* is considered a source and a reservoir of AMR genes, with many of the major families being described first in *K. pneumoniae* (22–25) before being identified in a range of other Gram-negative bacteria; hence it is crucial to improve our understanding of the broader population of *K. pneumoniae* beyond a handful of well-known clones. Many consider this knowledge to be fundamental to support efforts to control the threat to human health posed by this bacterium.

With this aim, we sequenced the genomes of nearly 300 diverse *K. pneumoniae* isolates spanning four continents and collected from a range of human and animal sources, including infection, colonization, and the environment (Dataset S1). We also performed a pangenome-wide association study (PGWAS) to look for associations between gene repertoire and disease potential/outcome and to identify distinct sets of accessory genes associated with virulence traits in humans, world-wide.

Results and Discussion

A total of 288 *K. pneumoniae* isolates were sequenced and compared with publicly available whole-genome sequences for an additional 40 isolates (Dataset S1). A total of 1,743 core genes, encoded in 1.48 Mbp of sequence, were conserved in all 328 genomes, and we identified 175,120 SNPs within these genes. Split network analysis and maximum likelihood (ML) phylogenetic analysis of these SNPs (Fig. 1*A*) identified four phylogroups, with 100% bootstrap support and corresponding to the groups previously defined as KpI, KpII-A, KpII-B, and KpIII.

We identified a pangenome of 29,886 unique protein-coding sequences among the 328 *K. pneumoniae* genomes. The gene accumulation curve (Fig. 1*B*) revealed an open pangenome, indicating that further genes will continue to be detected as additional *K. pneumoniae* genomes are sequenced. KpI, KpII, and KpIII shared 1,888 "common" genes that were present in ≥95% of genomes from each phylogroup. However, each individual *K. pneumoniae* carried thousands of additional accessory genes (median 3,817, yielding a median of 5,705 genes per genome). Some of these are likely to be on plasmids. It is not feasible to reconstruct whole novel plasmid sequences, at scale, from short-read data; however many genes associated with virulence and AMR were correlated with the presence of known plasmids (*SI Appendix*, Table S1).

The majority of accessory genes were rare, with 66% of genes found in ≤5% of *K. pneumoniae* and one third found in only one genome. Analysis of G+C content diversity and taxonomy indicated the *K. pneumoniae* accessory genes likely were acquired from a wide range of bacterial taxa including Enterobacteriaceae, *Vibrio*, and *Acinetobacter* (*SI Appendix*, Fig. S1). Accessory genes of intermediate frequency tended to be associated with one of the major phylogroups (*SI Appendix*, Fig. S2) or were correlated with phylogenetic lineages of KpI (*SI Appendix*, Fig. S3). These data highlight how broadly *K. pneumoniae* samples genetic diversity from other genera and, importantly, the considerable genomic plasticity that is contained within this species.

Whole-Genome Analysis Supports KpI, KpII, and KpIII as Distinct Species *K. pneumoniae*, *K. quasipneumoniae*, and *K. variicola*. Within each phylogroup the mean pairwise nucleotide divergence between genomes was ∼0.5%, whereas nucleotide divergence between phylogroups was 3–4% (calculated across the core genes). The two KpII-A isolates were 1.8–1.9% divergent from KpII-B and were 3.2–3.7% divergent from KpI and KpIII. As the split network indicates (Fig. 1*A*), there was very little evidence of homologous recombination between phylogroups, with the exception of a single human gut carriage isolate from Vietnam (Fig. 1*A* and *SI Appendix*, Fig. S4). Further, principal components analysis (PCA) on accessory gene content clearly distinguished the four phylogroups (Fig. 1*C*). These data provide whole-genome support for the proposal that KpI, KpII, and KpIII are distinct species by demonstrating these phylogroups constitute discrete bacterial populations that are evolving independently, with limited homologous recombination between groups (Fig. 1). Between-phylogroup nucleotide conservation,

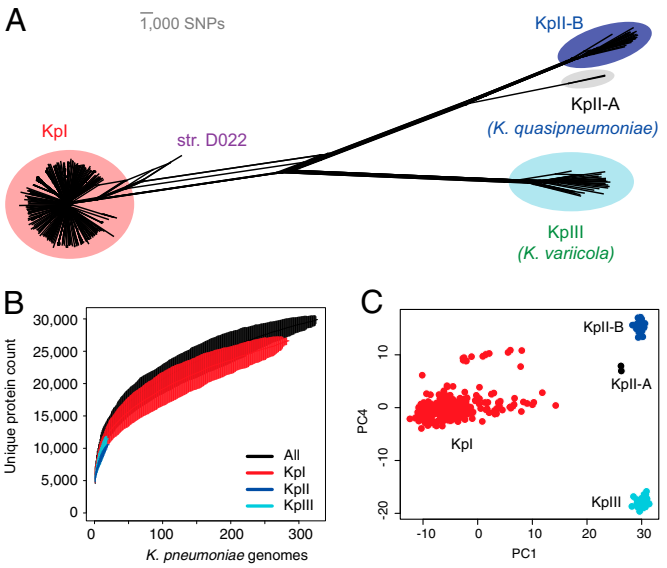

Fig. 1. The phylogroups and pangenome of *K. pneumoniae*. (*A*) Split network of 328 *K. pneumoniae* genomes with phylogroups highlighted. (*B*) Pangenome accumulation curves. (*C*) PCA analysis based on the presence of common (5–95% prevalence) accessory genes.

at 96%, is at the level commonly used as a cutoff for species differentiation in taxonomic analysis (32), and differences in gene content (Fig. 1*C*) further support the proposition that KpI, KpII, and KpIII are separately evolving populations that can be considered as separate species.

The observed speciation into genetically distinct phylogroups indicates that there are barriers to gene flow among these closely related populations. These barriers could arise through ecological separation into distinct niches, mechanistic barriers to homologous recombination, or adaptive selection against hybrid genotypes (33). There are no obvious mechanistic barriers to homologous recombination between KpI, KpII, and KpIII; indeed the observation of a large recombination between KpI and KpII (*SI Appendix*, Fig. S4) shows that homologous recombination is possible, although the rarity of this event (1 out of >300 genomes) suggests there could be selection against such hybrids. Although our sampling of *K. pneumoniae* isolates was blind to the distinction of KpI from KpII and KpIII, the characteristics of isolates falling into each phylogroup were quite distinct (*SI Appendix*, Fig. S5), suggesting that their speciation is likely driven by long-term separation in distinct ecological niches.

Our isolate collection, which focused on human and bovine-associated bacteria but also contained isolates from nonhuman primates and marine mammals, was comprised mainly (87%) of *K. pneumoniae* KpI. Because a number of different criteria, mainly unrelated to core phylogeny, were used to select the isolates included in this study, there is unlikely to have been a sampling bias with respect to phylogroup. Therefore, we hypothesize that this preponderance of KpI is associated with the bias of our collection toward mammalian-associated infection isolates. Other studies of human clinical *K. pneumoniae* isolates report similarly high rates of KpI and low rates of KpII or KpIII (34, 35), and all the sequence types reported in the literature as being linked to hospital outbreaks or pyogenic liver abscess belong to KpI (including CC258 and CC23). Notably, all the publicly available genomes of *K. pneumoniae* clinical isolates that we analyzed, including the *K. pneumoniae* subsp. *rhinoscleromatis* reference genome, clustered within KpI (15).

Although both KpII (*K. quasipneumoniae*) and KpIII (*K. variicola*) are capable of causing infections in humans, they appear to be less pathogenic than KpI, being associated more frequently with carriage (*SI Appendix*, Fig. S5). KpII was found almost exclusively in humans but was generally associated with colonization (50%) or HA infection (25%), consistent with low virulence and opportunistic infection (*SI Appendix*, Fig. S5*A*). No KpII or KpIII isolates in our collection were linked to either liver abscess or the death of a patient. We detected no KpII among the bovine isolates. In contrast, almost half of our KpIII isolates were of bovine origin, compared with 20% of KpI isolates [odds ratio (OR) 5.2; $P = 0.001$; Fisher's exact test) (*SI Appendix*, Fig. S5). The KpIII phylogroup was proposed in 2004 to be a distinct plant-associated nitrogen-fixing species, *K. variicola*, based on DNA–DNA hybridization and gene-sequence analysis (18). It has been isolated frequently from a wide range of plants (18, 36) and also has been shown to be an important nitrogen-fixing symbiont of leaf-cutter ants (37). Consistent with these reports, in our analysis the two public reference genomes of plant-associated *K. pneumoniae* belonged to KpIII (*SI Appendix*, Fig. S5). It is likely that the high number of bovine-derived KpIII isolates compared with human-derived KpIII isolates reflects bovine consumption of raw plant matter rather than any particular adaptation of KpIII to colonize or infect bovine hosts. Consistent with this notion, seven of the nine bovine KpIII were fecal carriage isolates, and only two were associated with infection. Importantly, our data show that the *nif* nitrogen-fixing operon (36) was present in all KpIII (*K. variicola*) genomes, supporting its identification as a nitrogen-fixing species. In contrast, *nif* was detected in only one KpI genome (a bovine mastitis isolate) and in half of the KpII-B genomes. This finding strongly supports the ecological separation of KpIII from KpI, with KpIII occupying a niche in which the ability to fix nitrogen is essential and KpI occupying a niche in which such ability is unnecessary and possibly disadvantageous and selected against. The intermediate frequency of *nif* in KpII is intriguing; because all our KpII isolates originated from humans, we hypothesize that nitrogen fixing is important in environmental-source populations of KpII but the *nif* operon is lost rapidly upon colonization of humans, possibly through negative selection.

Population Structure and Dynamics of *K. pneumoniae* KpI. We identified a total of 91,898 core genome SNPs among 283 KpI genomes (247 newly sequenced and 36 publicly available genome sequences) and inferred from these SNPs an ML phylogeny (Fig. 2*A*) and neighbor-joining split network (*SI Appendix*, Fig. S6*A*). These revealed a deep branching, star-like population structure, suggesting an early radiation of *K. pneumoniae* KpI into hundreds of distinct equally distant lineages (Fig. 2*A*). The deep branching structure, which was supported by genome-specific and lineage-specific SNPs (*SI Appendix*, Fig. S6*B*), is polytomous at the root with low bootstrap support for sequential branching patterns (Fig. 2*A*). Differences in gene content provided further support for the inferred population structure (*SI Appendix*, Fig. S3).

We divided the KpI genomes into 157 distinct phylogenetic lineages based on analysis of the core gene ML tree using RAMI (Fig. 2*A*) (38). Median divergence between lineages was 0.46% (range 0.04–0.61%), whereas genomes within the same lineage differed by a median of 0.02% (range 0–0.08%) and generally shared the same MLST sequence type. We used fineSTRUCTURE

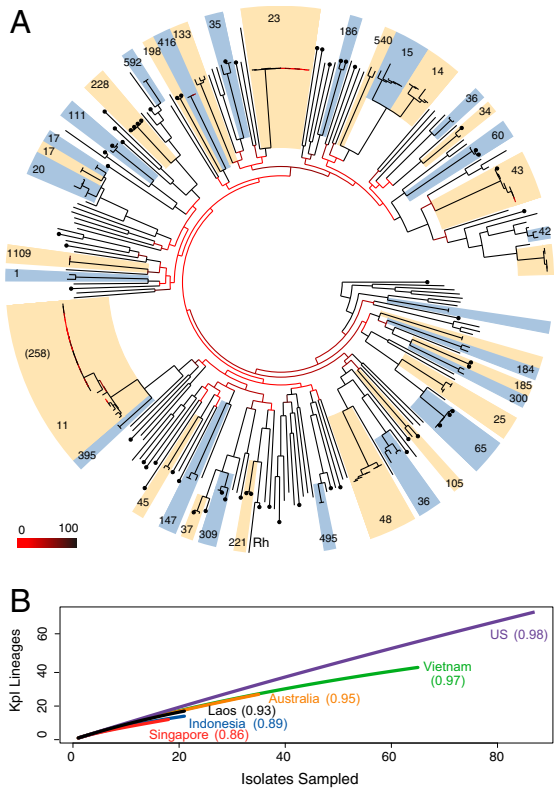

Fig. 2. Population structure of the *K. pneumoniae* KpI phylogroup. (*A*) Phylogeny of core gene SNPs. Branch colors indicate bootstrap support according to the legend provided in the figure. Black leaves indicate bovine isolates. Lineages with more than one genome are highlighted in alternating colors and labeled by sequence type. Rh, rhinoscleromatis. (*B*) Rarefaction curves show the accumulation of KpI lineages in each country, labeled with Simpson's diversity index (1-D) on a scale of 0–1 (0 = no diversity, i.e., all isolates are in same lineage; 1 = total diversity, i.e., every isolate is in a different lineage).

(39) to investigate the relationships between KpI genomes and to identify clusters in a phylogeny-independent manner. These analyses identified some recombination within KpI, but the observed relationships were consistent with the ML tree and the phylogenetically defined lineages (Fig. 2*A* and *SI Appendix*, Fig. S6*C*). The lineage accumulation curve for each geographical location (Fig. 2*B*) indicated that further sequencing of many hundreds of additional isolates would be needed to capture the full diversity of the broader *K. pneumoniae* KpI population. The maintenance of so many distinct lineages of KpI in each geographical area studied is intriguing and may be driven by adaptive selection or may be simply a product of genetic drift in a large population. However, these factors likely play out outside of any association with humans; hence our dataset, which is heavily biased toward animal-associated isolates, is not well suited to distinguish these possibilities. The inclusion of isolates from wider environmental sources will be important to extend the population framework further, against which we can begin to separate clones associated with disease from those causing sporadic infections that, although acutely important to individuals, can be seen at the wider population level as background epidemiological noise.

Genetic Determinants of CA Invasive Infection in Humans. *K. pneumoniae* KpI is best known as an opportunistic cause of HA infections, likely depending more on host factors such as compromised immunity than on specific pathogenicity factors in the bacterium. However, our human isolate collection included both HA (isolated >48 h after admission to hospital) and CA (isolated within 48 h of admission to hospital) KpI (Fig. 3). Further, 38 isolates came from CA invasive infections (defined as isolation of *K. pneumoniae* from a normally sterile site such as blood, CSF, intraocular, pleural, pericardial, or joint fluids, or deep-seated tissue abscesses), in which bacterial factors are likely to play a role in infection. Most (74%) of the 157 KpI phylogenetic lineages were observed only once within our collection; however, those that were represented by multiple isolates came from a diversity of specimen types (e.g., respiratory tract, urinary tract, digestive tract, blood) and from a mixture of carriage, invasive infection, and noninvasive infection (*SI Appendix*, Fig. S7). Thus, the ability of *K. pneumoniae* to cause invasive CA infections is not determined by lineage per se but may be associated with a specific virulence gene profile acquired horizontally and accessible through the pangenome.

Several genetic loci have been identified as virulence factors in *K. pneumoniae* on the basis of murine models of infection. These include gene clusters associated with the synthesis of siderophore systems yersiniabactin, aerobactin, colibactin, salmochelin (40–43), or microcin (44); the "regulators of mucoid phenotype" *rmpA* and *rmpA2*, which can up-regulate capsule production (45, 46); an allantoinase gene cluster (47); the ferric uptake operon *kfuABC* (48); and the two-component regulator *kvgAS* (49). Most of these genes were detected only in KpI, except for *kfuABC* (found in all KpII-B, 75% of KpIII, and 20% of KpI) and allantoinase (found in 50% of KpII-B and 5% of KpI). To investigate whether these genes also are associated with the ability cause disease in humans, we examined their distribution among human KpI isolates associated with invasive infection, noninvasive infection, and asymptomatic gut or throat carriage (Fig. 3*A*). Genes *rmpA* and *rmpA2* and the siderophore clusters were significantly associated with invasive human infection, compared with noninvasive or carriage isolates (ORs 3–15) (Fig. 3*A*).

Yersiniabactin, whose synthesis is encoded by the *ybt*, *irp1*, *irp2*, and *fyuA* genes which form the *Yersinia* high-pathogenicity island, HPI (40), was the most prevalent virulence-associated locus, present in one third of the KpI human isolates (Fig. 3*A*). Despite this high prevalence it was a strong predictor of infection vs. carriage in humans, with an OR of 7.4 [95% confidence interval (CI), 2.2–40; $P = 0.0001$; Fisher's exact test] and a positive predictive value of 95%. This effect was not dependent on chromosomal background, because yersiniabactin was significantly associated with infection in

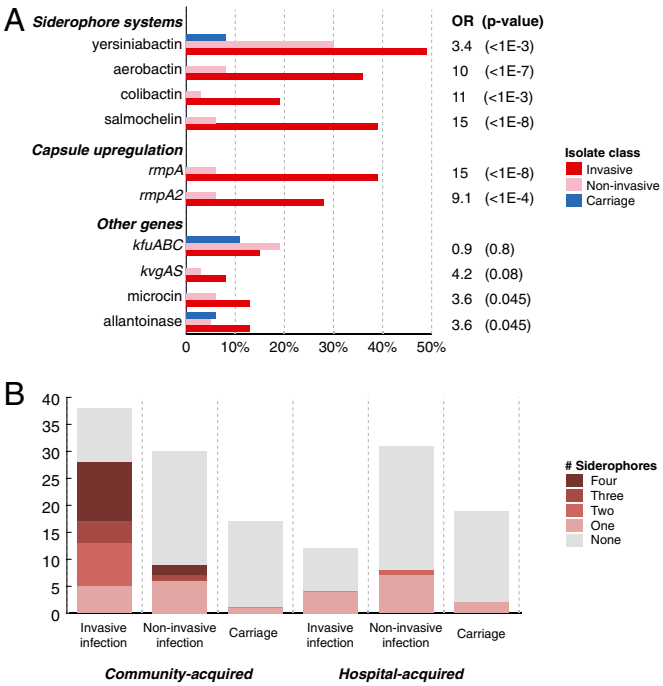

Fig. 3. Virulence genes in human *K. pneumoniae* KpI isolates. (*A*) Frequency of gene clusters among KpI isolated from different human sources. Invasive, isolated from a normally sterile site; noninvasive, associated with infections and isolated from respiratory, urinary tract, or wound infections in the absence of bacteremia; carriage, isolated in the absence of symptoms. OR, odds ratio for association between the presence of the gene cluster and invasive infection vs. others; *P* values were calculated using Fisher's exact test. (*B*) Number of KpI in each sample group, colored to indicate the total number of siderophore gene clusters in each isolate. Community-acquired, isolated ≤48 h after hospital admission; hospital-acquired, isolated >48 h after admission.

a logistic regression model that included phylogenetic lineage (OR 1.3; $P = 0.003$). Similarly, all *K. pneumoniae* encoding aerobactin (*iucABCD*, *iutA*), colibactin (*clbA-R*), salmochelin (*iroN*, *iroBCD*), and *rmpA* or *rmpA2* were isolated from human infections. These gene clusters were in strong linkage with each other as well as with yersiniabactin, and each was detected in 9–16 different KpI lineages (*SI Appendix*, Figs. S7 and S8). The combination of salmochelin, aerobactin, and *rmpA* was frequently, but not always, linked to the presence of genes from the known *K. pneumoniae* virulence plasmids pLVPK and pK2044 (43, 50) (*SI Appendix*, Table S1).

To extend these observations, we performed a PGWAS, screening each gene in the KpI pangenome for association with infection in humans. The strongest associations, reaching pangenome-wide significance after correcting for multiple testing, were the *rmpA*/*2* and siderophore genes and five additional predicted iron-metabolism genes (OR 8–10; 95% CI, 3–35) (*SI Appendix*, Fig. S9), also present on the virulence plasmid pK2044. Next we assessed whether the siderophores and predicted iron-metabolism genes could explain CA invasive infections (Fig. 3*B*). They were very common among CA invasive infections (75% carried one or more, and 60% carried three or more) but were rare among other classes of isolates (<10%) (Fig. 3*B*). These results are striking and suggest that access to iron may be central to the ability of *K. pneumoniae* to cause invasive disease in immunologically competent human hosts. Iron is essential for cell growth and replication in the human body, and nearly all free iron is bound by host proteins. Therefore bacteria must compete with host systems for iron, and siderophores are known to be critical to the ability of many bacteria to grow and replicate during colonization and infection of hosts (51). In

addition, siderophores also can play other roles in the interaction of *K. pneumoniae* and other Enterobacteriaceae with hosts, including modulating immune responses via the host protein lipocalin 2, binding noniron metal ions, and protecting against reactive oxygen species (51).

Because serotyping is technically challenging and unreliable, we assessed capsular variation based on allelic diversity of the *wzi* gene. We identified 136 alleles, including 118 in KpI. There was strong evidence of horizontal transfer: 27 *wzi* alleles were found in two or more different KpI lineages, and 20 of the 37 lineages with two or more genomes contained two or more *wzi* alleles (*SI Appendix*, Fig. S7). It is unclear what the drivers of capsular diversity are in *K. pneumoniae*. *K. pneumoniae* is an opportunistic pathogen that does not depend on colonization or infection of humans and other animals for survival; therefore it seems unlikely that capsular exchange within clones is associated with selection for host immune evasion, as is evident in human-adapted bacteria such as *Streptococcus pneumoniae* (52) or *Neisseria meningitidis* (53). *K. pneumoniae* capsule types K1, K2, and K5 have been proposed as virulence factors associated with pathogenicity in humans and in mouse models; however, in our dataset, all CA invasive infection isolates with the K1, K2, or K5 *wzi* alleles also carried siderophore genes that potentially could explain their virulence, and these *wzi* alleles also were found in isolates from asymptomatic carriage and HA infections.

The Emergence of Virulent *K. pneumoniae* Clones. We identified several *K. pneumoniae* clones that were significantly enriched for siderophores and/or *rmpA* genes (compared with the rest of KpI using Fisher's exact test) (*SI Appendix*, Fig. S7*A*). The best-known virulent *K. pneumoniae* clone is ST23, which is able to cause severe disease in apparently healthy individuals (13, 54) and typically carries all four acquired siderophore systems as well as *rmpA* and *rmpA2*. The 13 ST23 isolates in our collection were isolated from invasive infections more frequently than were other KpI (69% of ST23 were invasive vs. 35% of other KpI; OR 4.4, $P = 0.01$; Fisher's exact test) and were associated solely with CA infections. Our population genomic framework highlights how unusual these characteristics of ST23 are in the context of the broader population of KpI and also shows that each of the yersiniabactin, salmochelin, aerobactin, colibactin, *rmpA*, *rmpA2*, and microcin gene clusters of ST23 were associated with severe human disease in other KpI lineages with global distributions and high rates of invasive disease (*SI Appendix*, Fig. S7*B*). A feature shared by these clones that distinguishes them from the other noninvasive lineages is the presence of yersiniabactin, salmochelin, and *rmpA* in various combinations: (*i*) ST65, which also carried colibactin and *rmpA2*, was associated with lethal infections in humans and marine mammals [and has been associated with liver abscesses in Taiwan (55)]; (*ii*) ST592, which also carried aerobactin and *rmpA2*, was associated with a liver abscess and sepsis; (*iii*) ST25, which sometimes also carried aerobactin and *rmpA2*, was associated with liver abscesses in Vietnam and sepsis in Laos; (*iv*) ST60 was isolated from sepsis and also from an abdominal abscess in a monkey. ST231, which carried yersiniabactin and sometimes aerobactin, was isolated from cases of CA lethal pneumonia and abscesses and was multiply drug resistant.

Using the *K. pneumoniae* population framework, we also were able to make finely detailed comparisons between closely related isolates of the same lineage(s) that were differentiated by recorded clinical outcome. These data revealed the emergence of virulence within individual KpI clones, whereby invasive isolates were differentiated from noninvasive isolates of the same lineage by the presence of *rmpA* and siderophores. These clones include ST43 (yersiniabactin, salmochelin, aerobactin, and *rmpA*) and ST36 (yersiniabactin, colibactin, aerobactin, and *rmpA*), both linked to bacteremia. Within ST1, ST14, ST15, ST35, and ST48, the acquisition of yersiniabactin was linked to bacteremia and sepsis (*SI Apppendix*, Fig. S7*B*). These data build on previous observations made from a restricted number of isolates (47, 56, 57) by showing that the effect on virulence of acquiring iron-scavenging genes is not dependent on strain background or on specific combinations of siderophores and that the acquisition of iron-scavenging systems is central to the ability of *K. pneumoniae* to cause invasive disease in nonimmunocompromised patients. Importantly, this finding suggests that the acquisition of any one of the siderophore clusters increases the risk of severe infection in humans.

Genetic Determinants of Mastitis in Cows. *K. pneumoniae* is a frequent cause of bovine mastitis in dairy herds, but no known pathogenicity factors have been reported. Here we found that there was no association between observed mastitis and *K. pneumoniae* lineage(s). Moreover, genes that were associated with invasive infection in humans were rare among bovine isolates and were not associated with mastitis. PGWAS analysis for mastitis vs. carriage among bovine KpI isolates identified three genes significantly associated with KpI mastitis in cows (OR 30; 95% CI, 5.7–247; $P < 1 \times 10^{-6}$, adjusted for multiple testing $P = 0.03$). These genes formed a *lac* operon cluster encoding a transcriptional regulator (LacI), β-D-galactosidase (LacZ), and lactose permease (LacY) that is distinct from the lactose operon conserved in the chromosome of other KpI and KpIII isolates. The acquired *lac* operon was present in 18/20 KpI clinical mastitis isolates, 8/10 KpI subclinical mastitis isolates, and 3/19 KpI bovine fecal isolates, indicating association with both clinical and subclinical infection. This finding suggests that the utilization of lactose from the cow using this second lactose operon may confer an important selective growth advantage to *K. pneumoniae* isolates associated with mastitis. Similar associations have been made in *Streptococcus agalactiae* (58).

The acquired *lac* operon was common in *K. pneumoniae*, present in 50% of the KpI (*SI Appendix*, Fig. S10), 60% of the KpII, and 16% of the KpIII, but was not associated with infection or specimen type among human isolates. Notably, the acquired *lac* operon was located adjacent to a copy of the *fec* iron-enterobactin operon, which also showed a positive association with mastitis (OR 20; $P < 1 \times 10^{-4}$) and bovine host (OR 2; $P = 0.0005$). The combination of *fec* and *lac* operons likely provides *K. pneumoniae* the ability to invade via the udder and to thrive within mammary epithelial cells. The 30 bovine isolates carrying these operons were found in 23 different KpI lineages as well as in a KpII, confirming they are subject to extensive horizontal transfer (*SI Appendix*, Fig. S10). Moreover the *fec* and *lac* operons are known to be colocated together on numerous unrelated *K. pneumoniae* plasmids (pKPN3, pKN_LS6, pKPN_IT, pKP007, pKPN_CZ, pK29, and JM45_p1), indicating that the operons themselves are mobile and in strong genetic linkage within *K. pneumoniae*.

Antimicrobial Resistance Genes in *K. pneumoniae*. Given the clinical importance of AMR in *K. pneumoniae*, we performed a targeted analysis of all known AMR genes within our genomic dataset. We detected 84 AMR genes among the newly sequenced isolates (*SI Appendix*, Fig. S8). Our data confirm that the SHV, OKP, and LEN β-lactamases are core chromosomal genes of KpI, KpII, and KpIII, respectively (*SI Appendix*, Fig. S8). *FosA* and *oqxAB*, which confer low-level resistance to fosfomycin and quinolones, respectively, and have been detected as horizontally acquired AMR determinants in *Escherichia coli*, were core to all three phylogroups (*SI Appendix*, Fig. S8) and likely originated in *K. pneumoniae*. A further 78 AMR genes were detected in our collection, in 150/288 isolates (*SI Appendix*, Fig. S8). It was evident that the distribution of these AMR genes varied substantially among locales, with many resistance genes associated with particular countries or regions but not with lineage, presumably reflecting differing local antimicrobial use or availability (*SI Appendix*, Fig. S11) and most likely resulting from horizontal transfer of AMR

genes among locally circulating bacterial populations. We refer to these genes as "acquired AMR genes" to differentiate them from the core β-lactam resistance genes described above.

Acquired AMR genes were found commonly in the human-associated lineages KpI and KpII-B. In KpII-B, the human carriage-linked phylogroup, the median number of observed AMR genes was five per isolate. In KpI, acquired AMR genes were associated with humans (median 5 vs. 0; $P = 0.002$; Wilcoxon rank test) and were more common in carriage isolates than in infection isolates (median 9–10 vs. 4; $P < 1 \times 10^{-8}$) (Fig. 4*A*). Similarly, extended-spectrum β-lactamases (ESBLs), including ESBL alleles of SHV, were strongly associated with human isolates overall (detected in 40% of human isolates vs. 2% of bovine isolates; $P < 1 \times 10^{-6}$; Fisher's exact test), especially in the human-associated KpII-B group (50%). ESBL genes were not found in our KpIII isolates (*SI Appendix*, Fig. S5).

Nosocomial carriage isolates had more acquired AMR genes than community carriage isolates (median 13 vs. 7; $P = 0.08$) (Fig. 4*C*), implying that AMR genes have a role in opportunistic HA infection within the *K. pneumoniae* population. These patterns also were evident at the level of individual genes: Among human *K. pneumoniae*, most AMR genes were more common among HA isolates than among CA isolates ($P < 1 \times 10^{-16}$; paired Wilcoxon test) and also were more common among carriage isolates than among isolates associated with infection ($P < 1 \times 10^{-6}$) (Fig. 4*B*). Each of the individual AMR genes was detected at lower rates in bovine isolates than in human isolates, with the exception of *tet(B)*,which was found in 10% of bovine isolates and in only 1% of human isolates.

We were unable to link resistance loci reliably to specific plasmids because of the constraints inherent in using short-read Illumina data to generate whole-genome assemblies and the often repetitive nature of mobile elements that frequently disrupt those assemblies. However, screening against the PlasmidFinder database (59) identified 28 known plasmid replicons in 69/150 *K. pneumoniae* isolates with acquired AMR genes. These included six colicin plasmid replicons but also included a large range of replicons associated with large known conjugative AMR plasmids (*SI Appendix*, Table S1).

These data show that hospital isolates and carriage isolates accumulate AMR genes, likely through selection from antimicrobial exposure during long-term carriage or during hospital treatment. The more "successful" HA infection-associated clones, which lack virulence genes but appear multiple times in our collection, were MDR and had the greatest numbers of AMR genes (*SI Appendix*, Fig. S7). However, the precise complement of AMR genes differed among isolates of the same lineage (*SI Appendix*, Fig. S7), indicating distinct and potentially quite frequent gene-acquisition events. This finding suggests that the accumulation of resistance genes in non-virulent *K. pneumoniae* clones is a consequence, rather than a driver, of the successful spread of such clones within the human population, which is achieved mainly through effective asymptomatic colonization. This notion is consistent with the characterization of *K. pneumoniae* HA infections as resulting from opportunistic growth in nongut replicative niches that is selected for by the use of antimicrobials and can overwhelm compromised host immunity.

The Emerging Threat of Highly Pathogenic XDR *K. pneumoniae*. Given the existence of multiple virulent and MDR KpI clones that have access to a diverse mobile pool of virulence and AMR genes of

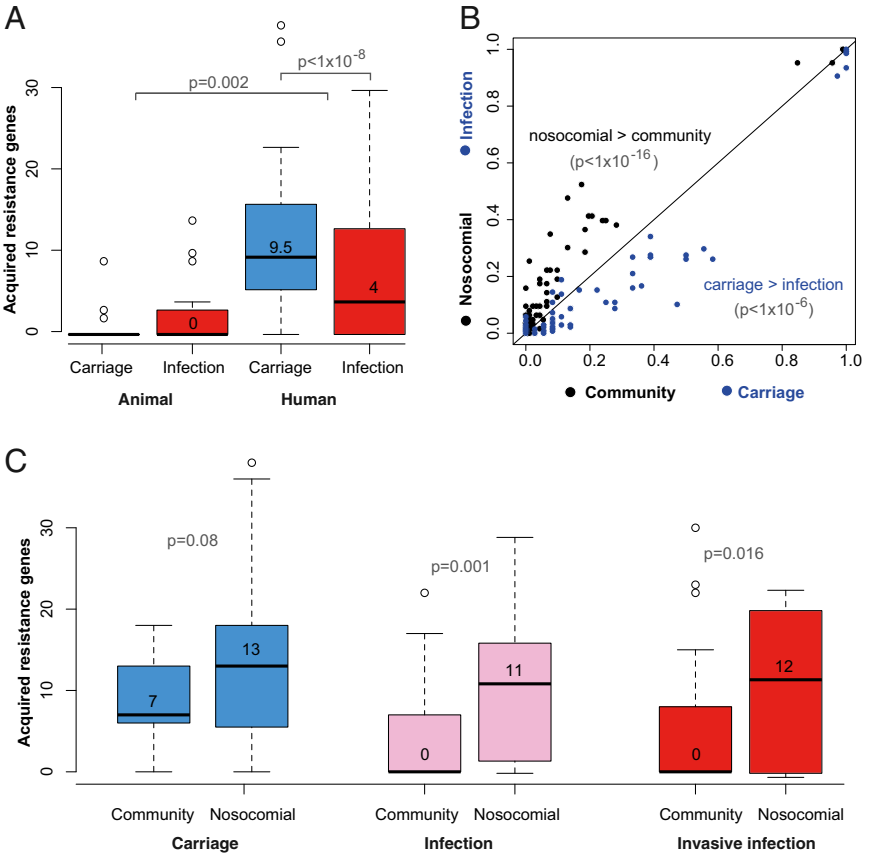

Fig. 4. AMR genes in KpI isolates. (*A*) Number of acquired AMR genes per isolate. (*B*) Frequency of individual AMR genes in human isolates by hospital status (black) or by infection status (blue). (*C*) Number of acquired AMR genes per isolate for different classes of human-associated isolates. *P* values were calculated using the Wilcoxon test. For boxplots, numbers indicate average numbers within each group.

high penetrance, and given the strong selective pressures imposed on bacteria in the hospital setting, there is potential for the emergence of an extremely drug-resistant (XDR), hypervirulent *K. pneumoniae* clone capable of causing severe, untreatable infections in healthy individuals. Indeed, convergence of specific virulence genes with AMR in *K. pneumoniae* already is beginning to occur. Many isolates of the epidemic KPC-producing ST258/ST11 clonal complex (CC258) already have acquired yersiniabactin (*SI Appendix*, Fig. S7). Experimental infection models show that yersiniabactin in CC258 enhances the bacterium's ability to colonize the respiratory tract and cause pneumonia (60). The presence of yersiniabactin in CC258, and also in the ESBL clones ST14 and ST15, is worrying, because our data show that it not only is strongly associated with infection in humans but also appears to be a frequent first step in the acquisition of additional siderophores that augment the ability of KpI clones to cause invasive non–hospital-related infections. There also is a high risk that current hypervirulent clones may acquire AMR, and MDR ST23 already has been reported in Korea, Vietnam, China, Madagascar, and Brazil (12, 61–63).

Now that the associations have been recognized, new efforts can be more focused on the task of reconstructing the mobile elements that carry the critical virulence genes, ideally using long-read sequencing approaches that have proven successful for unraveling MDR plasmids (64). However, given the plasticity of the *K. pneumoniae* pangenome and the broad distribution of key virulence and AMR genes, we need to place all *K. pneumoniae* within a wider population framework to target known epidemic clones and increase the likelihood of identifying new and emerging clones. Clinically, AMR phenotypes are monitored routinely in most hospital laboratories; however, our study indicates that it also will be crucial to perform active surveillance for key virulence genes and to determine clonal background. Genomic surveillance is being used increasingly to monitor KPC-producing *K. pneumoniae*, especially ST258, and the emergence of NDM-1 and colistin-resistant (XDR) isolates (27, 28, 65–67). Crucially, our study shows that we can augment these surveillance efforts by using key virulence genes as strong predictors of invasive disease in humans, and by determining clonal background, so that we can identify and track XDR hypervirulent clones as they emerge. To facilitate future genome-based surveillance, the genomes presented here have been deposited in the *K. pneumoniae* BIGSdb database (bigsdb.web.pasteur.fr) (12), which includes a core genome MLST scheme as well as key *K. pneumoniae* accessory genes encoding critical determinants of AMR, virulence, and capsule type. The data presented here provide a new genomic framework with which a new and deeper understanding of the *K. pneumoniae* population can be developed (*SI Appendix*, Fig. S12). Together with the *K. pneumoniae* BIGSdb database, this framework will provide a critical foundation and practical support for future studies investigating ecological niche adaptation, pathogenicity, and lineage diversification in *K. pneumoniae* and will facilitate more deeply informed genomic tracking and surveillance for the emergence and convergence of virulence and AMR in this increasingly important pathogen.

Materials and Methods

Bacterial Isolates and Sequencing. A total of 288 bacteria isolates, sampled to maximize diversity, were contributed from coauthors in six countries (*SI Appendix, Supplementary Text*). Genomic DNA was extracted and sequenced via Illumina HiSeq; data have been deposited in the European Nucleotide Archive under accession no. ERP000165. All *K. pneumoniae* genome data publicly available in the PATRIC database in March 2013 also were included in the analysis; all isolates and metadata are given in Dataset S1.

Clinical Definitions. Infection vs. carriage status was defined on the basis of available clinical data associated with each sample, that is, whether the isolate was considered, at the time of isolation, to be the cause of an infection. Invasive infections were defined as those associated with the isolation of *K. pneumoniae* from a normally sterile site (blood, CSF, intraocular, pleural, pericardial or joint fluids, or deep-seated tissue abscesses). The remaining infections, classed as noninvasive, were pneumonia, urinary tract infections, or wound infections with no recorded bacteremia. Bacteria isolated from samples taken >48 h after admission to hospital were classified as nosocomial; those isolated within 48 h of admission to hospital were classed as CA. This information was unavailable for 31 human KpI isolates, which were excluded from all analyses of mode of acquisition.

Variant Detection and Phylogenetic Analysis. SNPs were identified via mapping of Illumina reads to a reference genome (*K. pneumoniae* strain NTUH-K2044, NC_006625.1). Core genes were defined as the 1,743 NTUH-K2044 chromosomal genes with 100% coverage in all 328 genomes; sites outside these genes were excluded, leaving 175,120 SNPs for phylogenetic analysis. The alleles at these sites were concatenated to form a multiple alignment of SNPs for phylogenetic analyses. KpI lineages were defined based on patristic distances in the ML tree using RAMI (38), supported by a phylogeny-free approach using fineSTRUCTURE (39). STs were assigned to each genome according to the *K. pneumoniae* MLST database (13) by mapping to known alleles using SRST2 (68). Full details of all analyses are provided in *SI Appendix, Supplementary Text*.

Gene Content Analysis. De novo assemblies of Illumina reads were generated using Velvet and were combined with publicly available genomes to generate a nonredundant set of pangenome sequences. Taxonomic assignment was performed using MG-RAST v3.2 (69). A gene content matrix, indicating coverage of each gene in each genome, was generated by mapping to the annotated pangenome sequence. Read sets also were screened for known alleles of important genes using a read mapping approach with SRST2 (68). Gene databases analyzed were AMR alleles [ARG-Annot (70)], plasmid replicons [PlasmidFinder (59)], and virulence and *wzi* alleles [*K. pneumoniae* BIGSdb (12, 71)]. Full details of all analyses are provided in *SI Appendix, Supplementary Text*.

Statistical Analyses. Statistical analyses were performed in R as detailed in *SI Appendix, Supplementary Text*. Briefly, accumulation curves were generated using the *vegan* package. PCA was performed using the *prcomp* function. Pairwise gene content distances were calculated as Jaccard distances. PGWAS was done using Fisher's exact test and Benjamini–Hochberg correction of *P* values. Differences in the total number of acquired AMR genes among various isolate classes were assessed using the Wilcoxon test. Median nucleotide divergence within and between lineages or phylogroups was calculated from pairwise distances between genomes, obtained for each pair by dividing the total number of variant SNPs by the total length of the core genes analyzed (1,475,502 bp).

ACKNOWLEDGMENTS. We thank the sequencing teams at the Wellcome Trust Sanger Institute (WTSI) for genome sequencing; Dr. Rattanaphone Phetsouvanh and the directors of Mahosot Hospital (Vientiane, Lao People's Democratic Republic) and Joy Silisouk for performing DNA extractions on the Lao strains; Associate Professor Tse Hsien Koh for isolates from the Singapore General Hospital (Singapore); T. Williammee and D. Sentochnik (Basett Healthcare) for human isolates from the United States; K. W. Simpson and B. Dogan (Cornell University) for isolates from human intestinal biopsies; and Brenda Werner and other staff at Quality Milk Production Services (Cornell University) for help in collecting and processing bovine and human samples from the United States. This work was funded by National Health and Medical Research Council of Australia Fellowships 628930 and 1061409 (to K.E.H.) and Program Grant 606788 (to R.A.S.); Wellcome Trust Grants 098051 (to the WTSI) and 089275/H/09/Z (to the Lao-Oxford-Mahosot Hospital-Wellcome Trust Research Unit); Sir Henry Dale Fellowship (cofunded by the Royal Society) (to S. Baker); and by Victorian Life Sciences Computation Initiative Grant VR0082.

1. Jones RN (2010) Microbial etiologies of hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia. *Clin Infect Dis* 51(Suppl 1):S81–S87.
2. Falade AG, Ayede AI (2011) Epidemiology, aetiology and management of childhood acute community-acquired pneumonia in developing countries—a review. *Afr J Med Med Sci* 40(4):293–308.
3. Jarvis WR, Munn VP, Highsmith AK, Culver DH, Hughes JM (1985) The epidemiology of nosocomial infections caused by Klebsiella pneumoniae. *Infect Control* 6(2):68–74.
4. Podschun R, Ullmann U (1998) Klebsiella spp. as nosocomial pathogens: Epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* 11(4): 589–603.
5. Brisse S, Grimont F, Grimont P (2006) The genus Klebsiella. *The Prokaryotes A Handbook on the Biology of Bacteria, 3rd edition*, eds Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (Springer, New York), 3rd Ed. Vol 6: Proteobacteria: Gamma Subclass.

6. Shon AS, Bajwa RP, Russo TA (2013) Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: A new and dangerous breed. *Virulence* 4(2):107–118.
7. Broberg CA, Palacios M, Miller VL (2014) Klebsiella: A long way to go towards understanding this enigmatic jet-setter. *F1000Prime Rep* 6:64–76.
8. Jang S, et al. (2010) Pleuritis and suppurative pneumonia associated with a hypermucoviscosity phenotype of Klebsiella pneumoniae in California sea lions (Zalophus californianus). *Vet Microbiol* 141(1-2):174–177.
9. Twenhafel NA, et al. (2008) Multisystemic abscesses in African green monkeys (Chlorocebus aethiops) with invasive Klebsiella pneumoniae—identification of the hypermucoviscosity phenotype. *Vet Pathol* 45(2):226–231.
10. Schukken Y, et al. (2012) The "other" gram-negative bacteria in mastitis: Klebsiella, serratia, and more. *Vet Clin North Am Food Anim Pract* 28(2):239–256.
11. Bagley ST (1985) Habitat association of Klebsiella species. *Infect Control* 6(2):52–58.
12. Bialek-Davenet S, et al. (2014) Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. *Emerg Infect Dis* 20(11):1812–1820.
13. Brisse S, et al. (2009) Virulent clones of Klebsiella pneumoniae: Identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS ONE* 4(3):e4982.
14. Struve C, Krogfelt KA (2004) Pathogenic potential of environmental Klebsiella pneumoniae isolates. *Environ Microbiol* 6(6):584–590.
15. Brisse S, Verhoef J (2001) Phylogenetic diversity of Klebsiella pneumoniae and Klebsiella oxytoca clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping. *Int J Syst Evol Microbiol* 51(Pt 3):915–924.
16. Fevre C, Passet V, Weill FX, Grimont PA, Brisse S (2005) Variants of the Klebsiella pneumoniae OKP chromosomal beta-lactamase are divided into two main groups, OKP-A and OKP-B. *Antimicrob Agents Chemother* 49(12):5149–5152.
17. Brisse S, Passet V, Grimont PA (2014) Description of Klebsiella quasipneumoniae sp. nov., isolated from human infections, with two subspecies, Klebsiella quasipneumoniae subsp. quasipneumoniae subsp. nov. and Klebsiella quasipneumoniae subsp. similipneumoniae subsp. nov., and demonstration that Klebsiella singaporensis is a junior heterotypic synonym of Klebsiella variicola. *Int J Syst Evol Microbiol* 64(Pt 9):3146–3152.
18. Rosenblueth M, Martínez L, Silva J, Martínez-Romero E (2004) Klebsiella variicola, a novel species with clinical and plant-associated isolates. *Syst Appl Microbiol* 27(1):27–35.
19. Maatallah M, et al. (2014) Klebsiella variicola is a frequent cause of bloodstream infection in the stockholm area, and associated with higher mortality compared to K. pneumoniae. *PLoS ONE* 9(11):e113539.
20. Centers for Disease Control and Prevention (2013) Antibiotic Resistance Threats in the United States, 2013 (Centers Dis Control, Atlanta).
21. Department of Health and Department for Environment Food & Rural Affairs (2013) UK Five Year Antimicrobial Resistance Strategy 2013 to 2018. (Department of Health, London).
22. Chaves J, et al. (2001) SHV-1 beta-lactamase is mainly a chromosomally encoded species-specific enzyme in Klebsiella pneumoniae. *Antimicrob Agents Chemother* 45(10):2856–2861.
23. Sirot J, et al. (1988) Klebsiella pneumoniae and other Enterobacteriaceae producing novel plasmid-mediated beta-lactamases markedly active against third-generation cephalosporins: Epidemiologic studies. *Rev Infect Dis* 10(4):850–859.
24. Nordmann P, Cuzon G, Naas T (2009) The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. *Lancet Infect Dis* 9(4):228–236.
25. Nordmann P, Poirel L, Walsh TR, Livermore DM (2011) The emerging NDM carbapenemases. *Trends Microbiol* 19(12):588–595.
26. Chen L, et al. (2014) Carbapenemase-producing Klebsiella pneumoniae: Molecular and genetic decoding. *Trends Microbiol* 22(12):686–696.
27. Snitkin ES, et al. (2012) Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 4(148):148ra116.
28. Wright MS, et al. (2014) Population structure of KPC-producing Klebsiella pneumoniae isolates from midwestern U.S. hospitals. *Antimicrob Agents Chemother* 58(8):4961–4965.
29. Chen L, Mathema B, Pitout JD, DeLeo FR, Kreiswirth BN (2014) Epidemic Klebsiella pneumoniae ST258 is a hybrid strain. *MBio* 5(3):e01355-14.
30. Deleo FR, et al. (2014) Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. *Proc Natl Acad Sci USA* 111(13):4988–4993.
31. Mathers AJ, et al. (2015) Klebsiella pneumoniae carbapenemase (KPC)-producing K. pneumoniae at a single institution: Insights into endemicity from whole-genome sequencing. *Antimicrob Agents Chemother* 59(3):1656–1663.
32. Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 106(45):19126–19131.
33. Sheppard SK, McCarthy ND, Falush D, Maiden MC (2008) Convergence of Campylobacter species: Implications for bacterial evolution. *Science* 320(5873):237–239.
34. de Melo ME, Cabral AB, Maciel MA, da Silveira VM, de Souza Lopes AC (2011) Phylogenetic groups among Klebsiella pneumoniae isolates from Brazil: Relationship with antimicrobial resistance and origin. *Curr Microbiol* 62(5):1596–1601.
35. Brisse S, van Himbergen T, Kusters K, Verhoef J (2004) Development of a rapid identification method for Klebsiella pneumoniae phylogenetic groups and analysis of 420 clinical isolates. *Clin Microbiol Infect* 10(10):942–945.
36. Fouts DE, et al. (2008) Complete genome sequence of the N2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice. *PLoS Genet* 4(7):e1000141.
37. Pinto-Tomás AA, et al. (2009) Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* 326(5956):1120–1123.
38. Pommier T, Canbäck B, Lundberg P, Hagström A, Tunlid A (2009) RAMI: A tool for identification and characterization of phylogenetic clusters in microbial communities. *Bioinformatics* 25(6):736–742.
39. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.
40. Carniel E (2001) The Yersinia high-pathogenicity island: An iron-uptake island. *Microbes Infect* 3(7):561–569.
41. Nassif X, Sansonetti PJ (1986) Correlation of the virulence of Klebsiella pneumoniae K1 and K2 with the presence of a plasmid encoding aerobactin. *Infect Immun* 54(3):603–608.
42. Putze J, et al. (2009) Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect Immun* 77(11):4696–4703.
43. Chen YT, et al. (2004) Sequencing and analysis of the large virulence plasmid pLVPK of Klebsiella pneumoniae CG43. *Gene* 337:189–198.
44. Lagos R, et al. (2001) Structure, organization and characterization of the gene cluster involved in the production of microcin E492, a channel-forming bacteriocin. *Mol Microbiol* 42(1):229–243.
45. Cheng HY, et al. (2010) RmpA regulation of capsular polysaccharide biosynthesis in Klebsiella pneumoniae CG43. *J Bacteriol* 192(12):3144–3158.
46. Lai YC, Peng HL, Chang HY (2003) RmpA2, an activator of capsule biosynthesis in Klebsiella pneumoniae CG43, regulates K2 cps gene expression at the transcriptional level. *J Bacteriol* 185(3):788–800.
47. Chou HC, et al. (2004) Isolation of a chromosomal region of Klebsiella pneumoniae associated with allantoin metabolism and liver infection. *Infect Immun* 72(7):3783–3792.
48. Ma LC, Fang CT, Lee CZ, Shun CT, Wang JT (2005) Genomic heterogeneity in Klebsiella pneumoniae strains is associated with primary pyogenic liver abscess and metastatic infection. *J Infect Dis* 192(1):117–128.
49. Lai YC, Lin GT, Yang SL, Chang HY, Peng HL (2003) Identification and characterization of KvgAS, a two-component system in Klebsiella pneumoniae CG43. *FEMS Microbiol Lett* 218(1):121–126.
50. Wu KM, et al. (2009) Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. *J Bacteriol* 191(14):4492–4501.
51. Holden VI, Bachman MA (2015) Diverging roles of bacterial siderophores during infection. *Metallomics*, 10.1039/c4mt00333k.
52. Croucher NJ, et al. (2013) Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 45(6):656–663.
53. Ibarz-Pavón AB, et al. (2011) Changes in serogroup and genotype prevalence among carried meningococci in the United Kingdom during vaccine implementation. *J Infect Dis* 204(7):1046–1053.
54. Turton JF, et al. (2007) Genetically similar isolates of Klebsiella pneumoniae serotype K1 causing liver abscesses in three continents. *J Med Microbiol* 56(Pt 5):593–597.
55. Liao CH, Huang YT, Chang CY, Hsu HS, Hsueh PR (2014) Capsular serotypes and multilocus sequence types of bacteremic Klebsiella pneumoniae isolates associated with different types of infections. *Eur J Clin Microbiol Infect Dis* 33(3):365–369.
56. Siu LK, et al. (2011) Molecular typing and virulence analysis of serotype K1 Klebsiella pneumoniae strains isolated from liver abscess patients and stool samples from noninfectious subjects in Hong Kong, Singapore, and Taiwan. *J Clin Microbiol* 49(11):3761–3765.
57. Yu WL, et al. (2008) Comparison of prevalence of virulence factors for Klebsiella pneumoniae liver abscesses between isolates with capsular K1/K2 and non-K1/K2 serotypes. *Diagn Microbiol Infect Dis* 62(1):1–6.
58. Richards VP, et al. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. *Infect Genet Evol* 11(6):1263–1275.
59. Carattoli A, et al. (2014) In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58(7):3895–3903.
60. Bachman MA, Lenio S, Schmidt L, Oyler JE, Weiser JN (2012) Interaction of lipocalin 2, transferrin, and siderophores determines the replicative niche of Klebsiella pneumoniae during pneumonia. *MBio* 3(6):e00224-11.
61. Shin J, Soo Ko K (2014) Single origin of three plasmids bearing blaCTX-M-15 from different Klebsiella pneumoniae clones. *J Antimicrob Chemother* 69(4):969–972.
62. Liu YM, et al. (2014) Clinical and molecular characteristics of emerging hypervirulent Klebsiella pneumoniae bloodstream infections in mainland China. *Antimicrob Agents Chemother* 58(9):5379–5385.
63. Cejas D, et al. (2014) First isolate of KPC-2-producing Klebsiella pneumonaie sequence type 23 from the Americas. *J Clin Microbiol* 52(9):3483–3485.
64. Conlan S, et al. (2014) Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Sci Transl Med* 6(254):254ra126.
65. Gaiarsa S, et al. (2014) Genomic epidemiology of Klebsiella pneumoniae: The Italian scenario, and novel insights into the origin and global evolution of resistance to carbapenem antibiotics. *Antimicrob Agents Chemother* 59(1):389–396.
66. Köser CU, Ellington MJ, Peacock SJ (2014) Whole-genome sequencing to control antimicrobial resistance. *Trends Genet* 30(9):401–407.
67. Stoesser N, 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(12):7347–7357.
68. Inouye M, et al. (2014) SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 6(11):90–106.
69. Meyer F, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386–396.
70. Gupta SK, et al. (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother* 58(1):212–220.
71. Brisse S, et al. (2013) wzi Gene sequencing, a rapid method for determination of capsular type for Klebsiella strains. *J Clin Microbiol* 51(12):4073–4078.