www.sciencetranslationalmedicine.org/cgi/content/full/4/148/148ra116/DC1

Supplementary Materials for

Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* with Whole-Genome Sequencing

Evan S. Snitkin, Adrian M. Zelazny, Pamela J. Thomas, Frida Stock, NISC Comparative Sequencing Program, David K. Henderson, Tara N. Palmore,* Julia A. Segre*

*To whom correspondence should be addressed. E-mail: tpalmore@mail.nih.gov (T.N.P.); jsegre@nhgri.nih.gov (J.A.S.)

Published 22 August 2012, *Sci. Transl. Med.* **4**, 148ra116 (2012)
DOI: 10.1126/scitranslmed.3004129

The PDF file includes:

Supplementary Materials

Methods

Variant filtering

Single nucleotide variants (SNVs) were filtered to remove those SNVs that were likely to be a result of alignment or sequencing errors. SNVs were filtered out if: 1) they resided in genes annotated as phage, transposase or integrase, 2) they resided in genomic regions annotated as phage by the Phage Finder program*(44)* 3) they resided within 20 bp of the start or end of a contig, 4) they resided in tandem repeats of total length greater than 20 bp, as determined by the exact-tandem program associated with MUMmer*(45)*, 5) they resided in large inexact repeats as determined by nucmer, 6) they were within two positions of a second putative SNV, 6) the SNV position was ambiguous or low quality in any of the aligned genomes, 7) the 10 bp window surrounding the putative SNV contained more than two ambiguous or low quality base calls, or 8) the 10 bp window surrounding the putative SNV contained a A/T homopolymer run of length five or longer.

Constructing the putative transmission map

Our approach for construction of a putative outbreak transmission map built upon the method described by Jombart et al *(42)(40)*[3]. In their approach the most parsimonious transmission map was generated by first computing all pairwise genetic distances among isolates, and then finding the set of links that spans all isolates and has the minimal total genetic distance. Edmonds algorithm*(46),* which identifies the minimal spanning tree for a directed graph, was applied to identify most parsimonious transmission graph. Here, we use the same algorithm, but compute distances between patients with not only genetic

data, but also quantitative epidemiological data in a manner that accounts for the current understanding of nosocomial outbreaks of *K. pneumoniae*.

Two defining features of *K. pneumoniae* outbreaks are patient-to-patient transmission via hospital personnel, and the potential for silently colonized patients to act as hidden reservoirs. We aimed to capture both of these features in quantifying the relative likelihoods of different patient transmission routes. To capture patient-to-patient spread as the most likely mode of transmission, we considered transmission opportunities to occur when patients overlapped in the same hospital ward (Fig. S6A). The rational for this is that patients in the same ward typically share the same hospital staff, which can in turn act as vectors of transmission between patients. To capture silent colonization, we considered two possibilities. First, silent colonization of a potential donor can result in the donor culturing positive only after the recipient (Fig. S6B). Second, silent colonization of a recipient can result in transmission events facilitated by a patient overlap that occurred well before the recipient cultures positive (Fig. S6C).

These aspects of *K. pneumoniae* epidemiology were quantified for each putative transmission event between two patients. First, the requirement for patient overlap was implemented by assigning a maximal distance for a transmission from patient A to B, if B cultured positive before ever overlapping with A. For all other transmission events the total number of days of silent colonization in the donor (Fig. S6B) and recipient (Fig. S6C) were summed. Note that the likelihood of transmission between two patients does not have to be symmetrical, resulting in the transmission from patient A to B potentially

having a different weight than the transmission from B to A. Thus, the epidemiological distance matrix was calculated as:

$$E_{AB} = \begin{cases} \max(E) & No\ transmission\ opportunity \\ D(A) + R(B) & Otherwise \end{cases}$$

where *D* is the minimum number of days of silent colonization in the donor and *R* is the minimum number of days of silent colonization in the recipient required for the transmission event to have occurred.

Finally, we combined epidemiological and genetic weights into a single distance matrix. In integrating these two types of data, we desired that epidemiological weights should only be used to distinguish between scenarios that are equally probable based upon the genetic data. We therefore calculated the distance from patient A to B as the sum of the number of nucleotide differences between their respective genomes, and the number of days of silent colonization normalized to be between 10^{-5} and 10^{-2}. Thus the integrated distance matrix D was calculated as:

$$D_{AB} = G_{AB} + \left(\frac{999 * \left(\frac{E_{AB}}{\max(E)} \right)}{10^5} \right)$$

Edmonds algorithm was then applied to this integrated distance matrix to identify the most likely transmission map. In some cases there were alternative transmission maps that were equally likely. To identify variable links Edmonds algorithm was applied an additional 100 times to distance matrices with random noise added, with variable links being designated as those appearing less than 100 times.

Supplementary figures

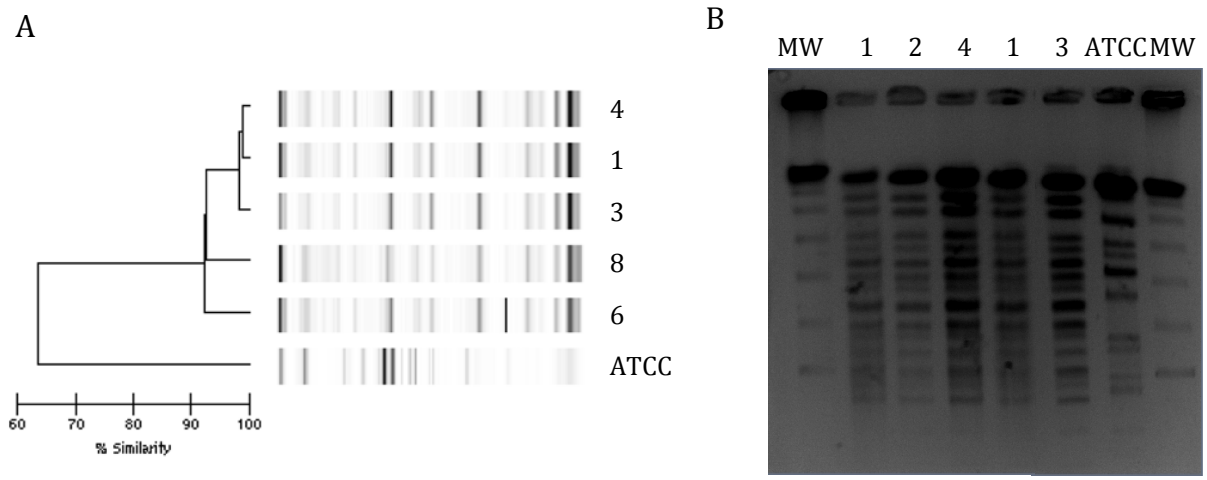

Fig. S1. Repetitive element PCR and pulsed-field gels of representative outbreak isolates. (A) Repetitive element PCR (rep-PCR) was performed on each isolate cultured during the outbreak, to provide a rapid indication as to whether the strain was part of the outbreak. Shown are rep-PCR banding patterns for representative isolates taken from outbreak patients, as well as for a non-KPC carrying ATCC strain, which acted as a reference. **(B)** Pulsed-field gel electrophoresis (PFGE) was performed on outbreak isolates to determine whether more resolution could be gained than with rep-PCR. Select isolates from outbreak are shown, in addition to a KPC carrying ATCC strain. The closeness between the outbreak strain and the ATCC strain demonstrates that the resolution provided by PFGE is not sufficient to distinguish transmission of the outbreak strain between patients and an independent introduction of a new isolate.

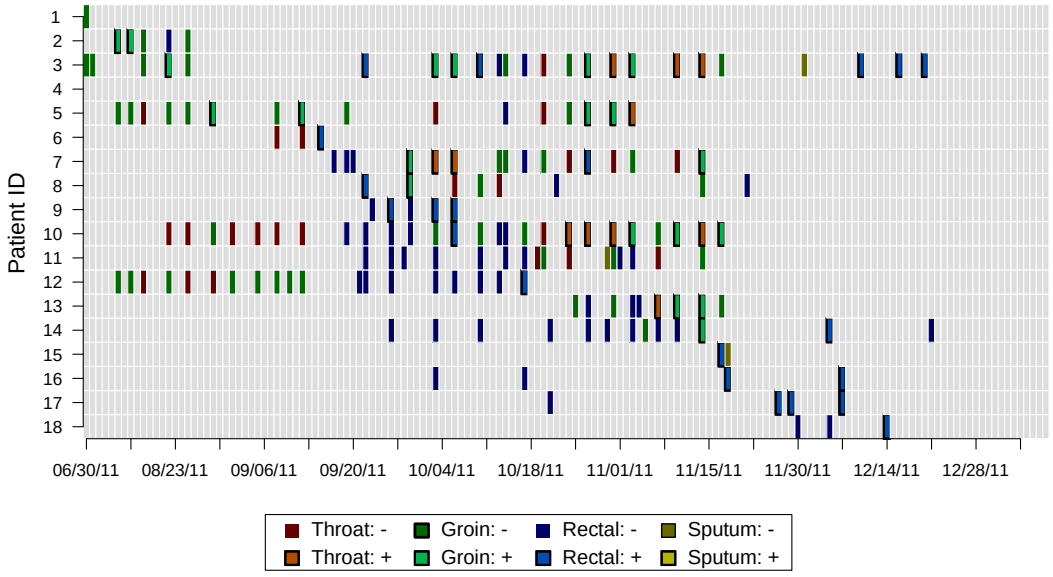

Fig. S2. Surveillance cultures for outbreak patients. Patients culturing positive for the outbreak strain of *K. pneumoniae* are listed on the y-axis and the dates during which the outbreak occurred are represented on the x-axis. Red, green, blue and yellow bars are used to indicate when throat, groin, rectal and sputum surveillance cultures, respectively, were performed. Darker shades of each color represent a negative culture and lighter shades a positive culture.

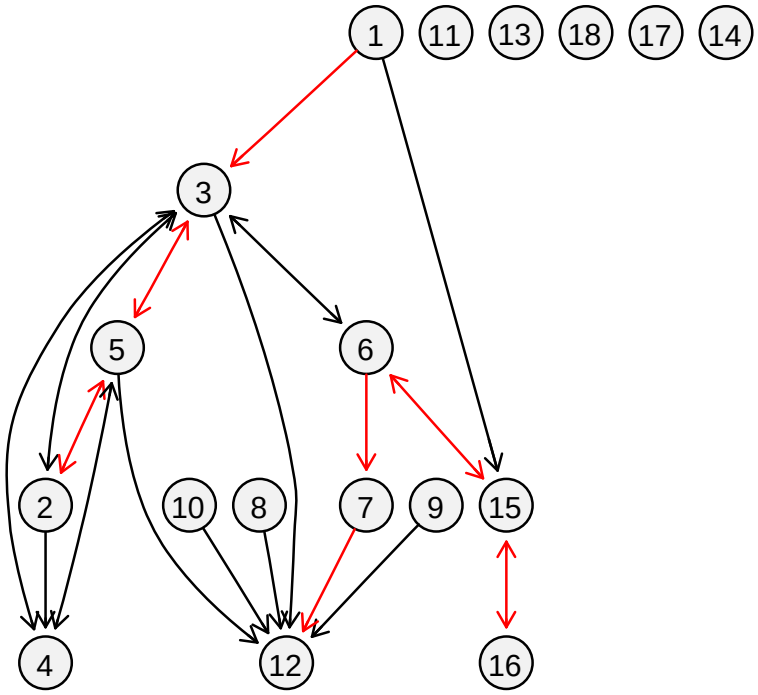

Fig. S3. Transmission opportunities between patients when using negative rectal surveillance to exclude patient colonization. Nodes in the graph represent patients, and edges between patients indicate possible transmission links. An arrow is present from one patient to another if the two patients overlapped in the same unit prior to the potential recipient culturing positive. Note that this figure is distinguished from Fig. 1C in the main text in that rectal surveillance cultures were used to limit possible transmission links between patients. Red links, the transmission event is predicted by our analysis.

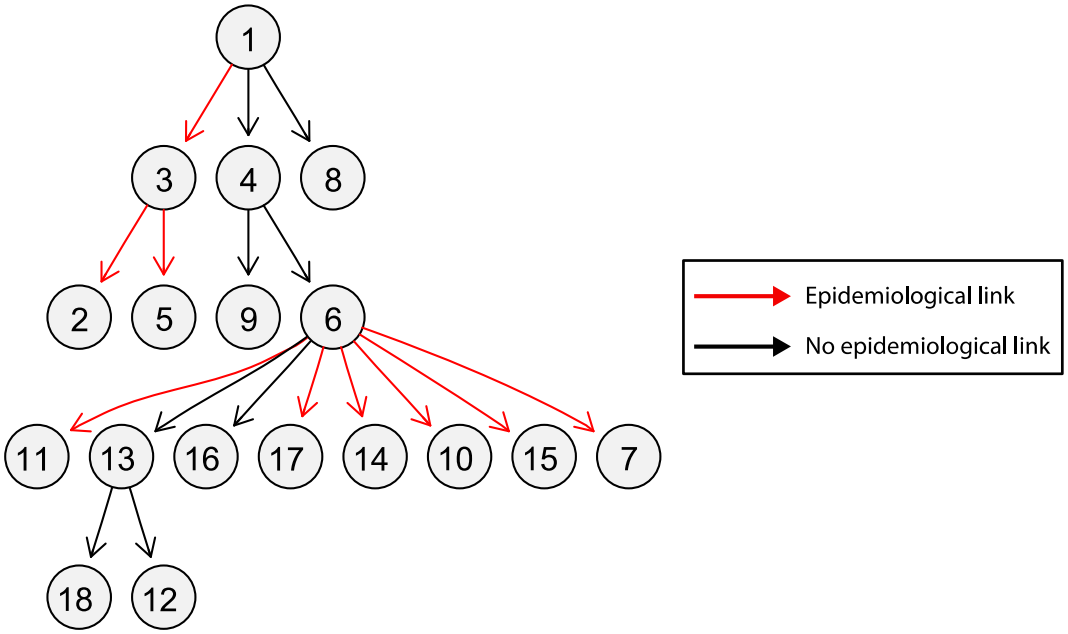

Fig. S4. Predicted transmission chart based only on genetic data. The transmission map was constructed by using the same approach as that presented in the main text, but only genetic variation among patients was considered. Circles, patients with ID; black arrows, a predicted transmission event leading either directly or indirectly from one patient to another; red arrows, an opportunity for a direct transmission event, as defined in Fig. 1C.

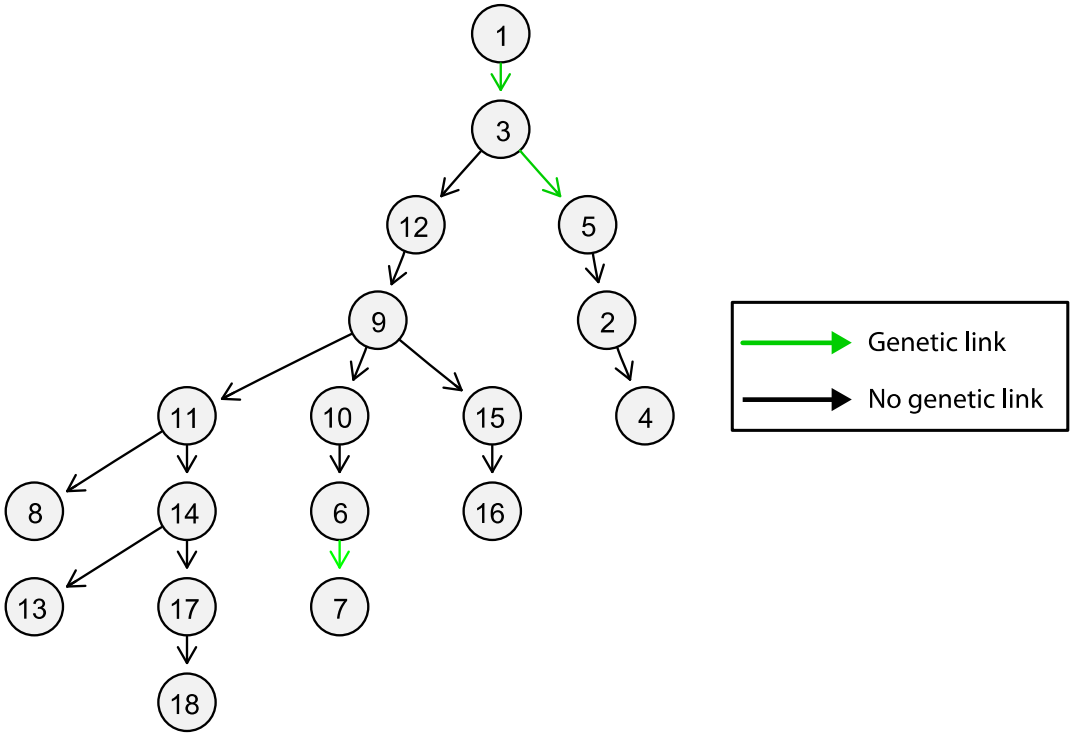

Fig. S5. Predicted transmission chart based only on epidemiological data. The transmission map was constructed with the same approach as in the main text, but only epidemiological links among patients were considered. Circles, patients with ID; black arrows, a predicted transmission event leading either directly or indirectly from one patient to another; green arrows, a predicted link between patients when considering only the genetic data, as shown in Fig. S4.

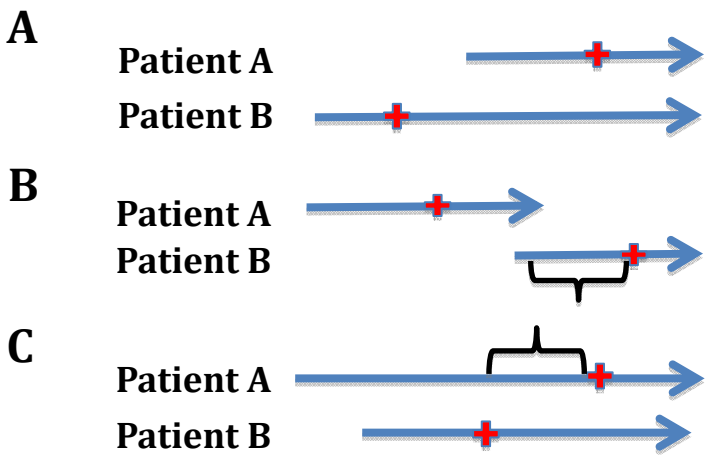

Fig. S6. Computing epidemiological distances between patients. The quantification of epidemiological distance is demonstrated with model examples of transmission from hypothetical patient A to patient B. Blue arrows represent when patients were present in a given ward over time. The red "+" indicates when the first positive culture for A or B occurred. **(A)** A transmission event from patient A to B was deemed to have a maximal distance (be least likely) if patient B cultured positive before ever overlapping with patient A. If there was an overlap between A and B before B cultured positive, then the weight of this link was calculated as the minimum number of days of silent colonization required to explain the event. Silent colonization can manifest as silent colonization of both the donor (patient A) and the recipient (patient B). **(B)** Silent colonization of the recipient is quantified as the number of days after overlapping with the donor, that the recipient cultures positive. **(C)** Silent colonization of the donor is quantified as the number of days after the recipient cultures positive, that the donor first cultures positive.

Locus Tag	Strain	Mean/median depth	Number of contigs	Contig N50	Number of bases	Number of protein coding genes
KPNIH1	1	33/30	116	158289	5725345	5612
KPNIH5	2	21/20	147	126254	5718888	5626
KPNIH6	2-R	27/25	132	130041	5716025	5620
KPNIH2	3	28/27	123	154879	5725591	5630
KPNIH4	4	23/21	144	120800	5719493	5621
KPNIH10	5	42/40	155	158000	5717574	5621
KPNIH9	6	38/37	105	178365	5721171	5613
KPNIH8	7	20/19	151	158021	5715293	5634
KPNIH11	8	26/25	157	150495	5714746	5562
KPNIH12	9	24/23	143	149661	5721646	5567
KPNIH14	10	23/22	140	147593	5725273	5517
KPNIH20	11	36/35	136	167325	5726024	5525
KPNIH17	12	37/35	120	158097	5716707	5571
KPNIH16	13	33/31	153	158009	5716543	5568
KPNIH21	14	43.3/41	148	117020	5760928	5641
KPNIH19	15	20/19	174	98646	5718161	5474
KPNIH18	16	30/27	141	153222	5771043	5631
KPNIH22	17	38.8/38	134	131068	5751965	5626
KPNIH23	18	23.5/23	282	47638	5748041	5631
KPNIH7	VENT	30/29	139	158114	5725101	5619

Table S1. Genome sequencing statistics.

Demographic characteristics	
Female	**5**
Median age (yrs)	**44**
Underlying malignancy	**9**
Solid tumor	**5**
Hematologic malignancy	**4**
Primary Immunodeficiency	**2**
Aplastic anemia	**2**
Lung disease	**2**
Other	**2**
HSCT recipients	**6**
Acquisition of KPC	
Acquired KPC in ICU	**12**
Acquired KPC on medical or surgical ward	**5**
First detected in clinical culture	**2**
First detected in surveillance culture	**15**
KPC grown from clinical cultures at any time	**10**
Bloodstream infections	**8**
Outcome	
Died	**10**
Died from KPC	**6**
Died from underlying condition	**4**

Table S2. Characteristics of patients who acquired outbreak strain. MUD, matched unrelated donor; MRD, matched related donor; NA, not applicable.

Patient	Amikacin	Amox/Clav'ate	Ampicillin	Aztreonam	Cefazolin	Cefepime	Cefotaxime	Cefoxitin	Ceftazidime	Ceftriaxone	Ciprofloxacin	Colistin	Ertapenem	Gentamicin	Imipenem	Levofloxacin	Meropenem	Pip/Tazo	Rifampin	Tetracycline	Tigecycline	Tobramycin	Trimeth/Sulfa
1	32	16/8	>16	>16	>16	>16	32	>16	>256	>32	>2	ND	>4	<=2	8	>4	>8	>64/4	>32	>8	ND	>8	>2/38
2	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	1	>4	4	8	>4	>8	>64/4	>32	>8	2	>8	>2/38
3	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	>8	4	>8	>2/38
4	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	>8	2	>8	>2/38
5	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	>8	2	>8	>2/38
6	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	ND	>4	4	8	>4	>8	>64/4	>32	ND	16	>8	>2/38
7	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
8	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	128	>4	4	8	>4	>8	>64/4	>32	>8	2	>8	>2/38
9	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	>32	>4	>32	>64/4	>32	8	2	>8	>2/38
10	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	4	>4	4	8	>4	>8	>64/4	>32	>8	1	>8	>2/38
11	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	0.25	>4	4	8	>4	>8	>64/4	>32	>8	16	>8	>2/38
12	32	16/8	>16	>16	>16	>16	>32	>16	>2	>32	>2	2	>4	4	>8	>4	>8	>64/4	>32	4	1	>8	>2/38
13	32	16/8	>16	>16	>16	>16	>32	>16	>2	>32	>2	4	>4	4	>8	>4	>8	>64/4	>32	8	2	>8	>2/38
14	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
15	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	4	1	>8	>2/38
16	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	4	1	>8	>2/38
17	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	4	1	>8	>2/38
18	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	2	>4	4	8	>4	>8	>64/4	>32	4	1	>8	>2/38
VENT	32	16/8	>16	>16	>16	>16	32	>16	>2	>32	>2	8	>4	4	8	>4	>8	>64/4	>32	4	2	>8	>2/38

Table S3. MICs for antibiotic susceptibility of outbreak isolates.

Mutation	NTUH-K2044	1 (6-15: Urine)	1 (6-16: BAL)	1 (6-17: Urine)	1 (6-19: Urine)	1 (6-27: Urine)	1 (6-30: Groin)	1 (6-30: Throat)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	VENT
Noncoding mutation (G -> T) between NTUH-K2044:KP1_0263(aspartate kinase III) and NTUH-K2044:KP1_0264(glucose-6-phosphate isomerase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ACG -> TCG (T -> S) at 1504 of 1563 in NTUH-K2044:KP1_0734(4-hydroxyphenylacetate 3-hydroxylase)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TCA -> TCT (S -> S) at 882 of 1221 in NTUH-K2044:KP1_0750(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CAT -> TAT (H -> Y) at 118 of 717 in NTUH-K2044:KP1_0818(negative response regulator of genes in aerobic pathways)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1
GCG -> CCG (A -> P) at 328 of 582 in NTUH-K2044:KP1_1490(putative regulatory protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
GGC -> GAC (G -> D) at 179 of 474 in NTUH-K2044:KP1_1641(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
TAT -> TTT (Y -> F) at 713 of 1479 in NTUH-K2044:KP1_1673(PTR2-family transport protein)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1
ATG -> TTG (M -> L) at 886 of 1233 in NTUH-K2044:KP1_1836(multidrug/chloramphenicol efflux transport protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
ATC -> ATT (I -> I) at 1173 of 2424 in NTUH-K2044:KP1_1913(putative recombination protein with metallo-hydrolase domain)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
CTG -> ATG (L -> M) at 130 of 309 in NTUH-K2044:KP1_2318(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GCC -> GTC (A -> V) at 527 of 789 in NTUH-K2044:KP1_2327(enoyl-(acyl carrier protein) reductase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
TTG -> CTG (L -> L) at 22 of 912 in NTUH-K2044:KP1_2432(NmrA family protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
TTT -> TTC (F -> F) at 393 of 1647 in NTUH-K2044:KP1_2571(malate:quinone oxidoreductase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ACC -> ACT (T -> T) at 855 of 2436 in NTUH-K2044:KP1_2603(putative dimethyl sulfoxide reductase major subunit)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
ACC -> ATC (T -> I) at 1175 of 1188 in NTUH-K2044:KP1_2941(L-Ala-D/L-Glu epimerase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CTT -> CCT (L -> P) at 965 of 1500 in NTUH-K2044:KP1_2943(energy-dependent efflux protein for methyl viologen resistance)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GAC -> GAA (D -> E) at 411 of 510 in NTUH-K2044:KP1_2961(putative acyltransferase)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
GGT -> GAT (G -> D) at 20 of 1089 in NTUH-K2044:KP1_3175(putative ABC transport system periplasmic binding component)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
TGG -> AGG (W -> R) at 472 of 963 in NTUH-K2044:KP1_3203(putative ABC-type transport system periplasmic component)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
GTA -> GTG (V -> V) at 135 of 1185 in NTUH-K2044:KP1_3241(L-lactate dehydrogenase)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Noncoding mutation (G -> T) between NTUH-K2044:KP1_3338(putative phosphatase) and NTUH-K2044:KP1_3339(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CTG -> CAG (L -> Q) at 2597 of 3417 in NTUH-K2044:KP1_3370(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Noncoding mutation (G -> A) between NTUH-K2044:KP1_3721(outer membrane protein) and NTUH-K2044:KP1_3725(putative acid phosphatase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

GGC -> TGC (G -> C) at 811 of 1110 in NTUH-K2044:KP1_3857(putative transport protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
AGC -> GGC (S -> G) at 598 of 888 in NTUH-K2044:KP1_4082(putative LysR-family transcriptional regulator)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
GAG -> GGG (E -> G) at 1511 of 2556 in NTUH-K2044:KP1_4249(putative export and assembly usher protein of type 1 fimbriae)	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GGC -> TGC (G -> C) at 199 of 912 in NTUH-K2044:KP1_4467(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Noncoding mutation (A -> G) between NTUH-K2044:KP1_4544(hypothetical protein) and NTUH-K2044:KP1_4545(acetyl-CoA acetyltransferase)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
GAG -> GAA (E -> E) at 315 of 636 in NTUH-K2044:KP1_4555(fimbriae stability-associated protein)	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
TCG -> ACG (S -> T) at 202 of 777 in NTUH-K2044:KP1_4820(negative regulator of exu regulon)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
CAG -> CCG (Q -> P) at 377 of 393 in NTUH-K2044:KP1_4942(30S ribosomal protein S9)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GTC -> CTC (V -> L) at 169 of 312 in NTUH-K2044:KP1_5045(30S ribosomal protein S10)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GCC -> GCT (A -> A) at 546 of 714 in NTUH-K2044:KP1_5417(hypothetical protein)	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
ACC -> ACT (T -> T) at 774 of 1590 in NTUH-K2044:KP1_0789(peptide chain release factor 3)	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACT -> TCT (T -> S) at 4 of 1674 in NTUH-K2044:KP1_1910(30S ribosomal protein S1)	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TGG -> TTG (W -> L) at 677 of 696 in NTUH-K2044:KP1_2873(putative Mg(2+) transport ATPase)	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Noncoding mutation (G -> A) between NTUH-K2044:KP1_2988(hypothetical protein) and NTUH-K2044:KP1_2989(putative aldehyde dehydrogenase)	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GTA -> GTG (V -> V) at 285 of 1044 in NTUH-K2044:KP1_0928(guanosine 5'-monophosphate oxidoreductase)	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GGA -> GTA (G -> V) at 125 of 582 in NTUH-K2044:KP1_1490(putative regulatory protein)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CTG -> CTA (L -> L) at 441 of 804 in NTUH-K2044:KP1_2425(probable dehydrogenase/reductase oxidoreductase protein)	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CAG -> TAG (Q -> *) at 310 of 711 in NTUH-K2044:KP1_4551(hypothetical protein)	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table S4. Mutations identified among outbreak genomes