Epidemic population structure of clinical Klebsiella aerogenes inferred by multilocus sequence typing

CURRENT STATUS: POSTED

Sheng Chen
Department of Pediatric, Southwest Hospital, the Third Military Medicinal University (Army Medical University)

Yu Wang
Department of Pharmacy, Southwest Hospital, the Third Medical University (Army Medical University)

Lijun Zeng
State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology

Wenbo Luo
Department of Pharmacy, Southwest Hospital, the Third Military University (Army Medical University); State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology

Wei Feng
Department of Pharmacy, Southwest Hospital, the Third Military Medical University (Army Medical University)

Ping Li
State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology

Qian Wang
Department of Pharmacy, Southwest Hospital, the Third Military Medical University (Army Medical University)

Dongsheng Zhou
State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology

Chenyi Guo
State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology

Fengjun Sun  fengj_sun@163.com
Army Medical University
Abstract

Background: Multilocus sequence typing (MLST) act as an accurate approach to characterize bacterial population genetics, phylogeny and epidemiology, and has not yet been applied to Klebsiella aerogenes. Results: A MLST scheme was established for a collection of 213 isolates of K. aerogenes. These strains exhibited considerable sequence diversity under purifying selection, and could be assigned into 135 sequence types, which were further divided into 8 clonal complexes and a lot of doubletons and singletons scattered in the population snapshot. Five separately clustering lineages were presented in the population, which displayed evident homologous recombination occurred within and across lineages, with a tendency of linkage disequilibrium. Conclusions: K. aerogenes shows an epidemic population structure displaying high levels of recombination occurring more frequently than point mutation. Key words: Klebsiella aerogenes; multilocus sequence typing; sequence types; purifying selection; linkage disequilibrium; population structure

Background

*Klebsiella aerogenes* (formerly *Enterobacter aerogenes*) (1), belonging to the family *Enterobacteriaceae*, is a rod-shaped, nonspore-forming bacterium capable of motility via peritrichous flagella. *K. aerogenes* is commonly nonpathogenic to healthy humans but, since 1990s, it has been recognized as an important opportunistic pathogen responsible for nosocomial infections, such as meningitis, sepsis, bacteremia, respiratory and urinary tracts infections, and surgical site infections, in humans with weakened immune systems (2). Clinical *K. aerogenes* strains often exhibit multi-drug resistance to commonly used broad-spectrum antibiotics, including third-generation cephalosporins and carbapenems frequently used for treatment of *Enterobacteriaceae*-induced severe or refractory
infections (2).

Arbitrarily primed-PCR (3), enterobacterial repetitive intergenic consensus PCR (4), random amplification of polymorphic DNA (5), and pulsed field gel electrophoresis (PFGE) (6) have been applied to epidemiologically characterize outbreaks of K. aerogenes-induced infections. Compared to these traditional genotyping methods, multilocus sequence typing (MLST), based on the sequences of multiple (commonly 7) housekeeping genes, is able to provide more discriminatory power to differentiate isolates, and thus act as a more accurate approach to characterize bacterial population genetics, phylogeny and epidemiology, enabling to steadily track source and spreading of bacterial infections. The present work established a MLST scheme for K. aerogenes, which provided an insight into the high-level genetic diversity and an epidemic population structure of K. aerogenes.

Methods

Bacterial strains

A total of 184 isolates of K. aerogenes (Additional file 1: Table S1) were isolated between 2013 and 2015 from patients with nosocomial infections from 8 hospitals in 6 Chinese cities, including Chongqing (Southwest China), Beijing, Tianjin and Zhengzhou (North China), and Hefei and Fuzhou (East China). Bacterial species identification was performed by PCR detection of a 981-bp K. aerogenes-specific ampC sequence (7). Genomic DNA of each isolate was isolated by classical phenol/chloroform method followed by methoxyethanol removal of polysaccharides that contaminate genomic DNA (8), and then arrayed in 96-well PCR plates for further analyses. Additional 29 available complete or draft genome sequences of K. aerogenes derived from NCBI (last accessed May 1st 2016) were included for the MLST analysis (Additional file 1: Table S1).

K. aerogenes MLST scheme

The seven housekeeping genes, namely dnaA, fusA, gyrB, leuS, pyrG, rplB and rpoB, and
their sequences involved in MLST analysis in this study (Table 1) were derived from the 
MLST scheme of *K. cloacae* (http://pubmlst.org/ecloacae/). PCR primers (Table 1) of each 
indicative target gene were designed from the chromosome sequence of *K. aerogenes*
KCTC 2190 (9) (accession number CP002824). A volume of 50 µl PCR mixture contained 50 
mM KCl, 10 mM Tris-HCl (pH8.0), 2.5 mM MgCl2, 0.001% gelatin, 0.1% BSA, 100 µM of 
each dATP, dCTP, dGTP and dTTP, 0.1 µM of each primer, 1 unit of each of ExTaq 
polymerases (TaKaRa), and 10 ng of genomic DNA. The amplification conditions were as 
follows: 95°C for 5min, and then 30 cycles of 94°C for 40s, an appropriate annealing 
temperature (Table 1) for 40s, and 72°C for 1min. PCR products were analyzed by agarose 
gel electrophoresis and purified by ultrafiltration (Millipore). Both DNA strands of each 
amplicon were sequenced with the primers as used in PCR on an ABI-3700 sequencer. DNA 
sequences were aligned using MUSCLE Version 3.8 (10).

**Sequence diversity analyses**

The GC content was calculated as \(\frac{\text{Count}(G + C)}{\text{Count}(A + T + G + C)} \times 100\). The number 
of polymorphic sites, the average pairwise differences per-site (π), and the \(p\)-values of 
Tajima’s D test were calculated by DnaSP Version 5.10 (11). The average non-
synonymous/synonymous rate ratio (\(K_a/K_s\)) was calculated with KaKs Calculator Version 
2.0 (12) to estimate whether the group was under purifying selection or not. Distinct 
allelic profiles were assigned into different sequence types (STs), which were further 
clustered into clonal complexes (CCs), using eBURST Version 3 (13). Every two different 
STs who share six of the seven loci were connected with each other, and a connected 
graph with more than 2 STs was called a CC. If only two STs were connected together, the 
two were called doubleton. All other unconnected STs were named singletons. The 
founders (ancestry types) of CCs were predicted by eBURST with 1,000 re-samplings for 
bootstrap.
**Population structure analyses**

The sequences of the 7 MLST loci of each ST were concatenated, and the neighbor-joining tree was built by MEGA7 (14) with the concatenated sequences of all the detected STs. The percentages of replicate trees, in which the associated taxa clustered together in the bootstrap test (1000 replicates), were shown next to the branches. Linkage model was implemented by STRUCTURE Version 2.3 (15) to infer ancestries for each ST, which assumed that each ST was derived from fixed number of assuming ancestral subpopulations. The posterior probability $P(X|K)$ was calculated to determine the probability of ST $X$ derived from ancestral subpopulation. Individual runs (200,000 burn-in iterations and 1,200,000 iterations sampling iterations) per value of $K$ (ranged from 2 to 10) were performed. The splits network was generated by SplitsTree4 (16) using the neighbor-net method.

**Recombination analyses**

The pairwise homoplasy test (*Phi* test) was performed with SplitsTree4, which was used to detect evidence of recombination events in a set of aligned sequences. When the evidence for recombination was found, the $p$ value yielded from the *Phi* test was less than 0.05. The *LDhat* method implemented by RDP4 (17) was used to calculate the average per-site $\rho/\theta$ ratio based on the concatenated sequences of the seven loci with 1,000,000 MCMC updates, where $\rho$ and $\theta$ represented the rates of recombination and mutation respectively.

**Linkage equilibrium test**

The standardized index of association ($I_A$), testing linkage equilibrium, was calculated by Linkage Analysis Version 3.637 (18) with 10,000 iterations by Monte Carlo based on allelic profiles. If $I_A$ was statistically significant different from zero, alleles were suggested with genetic linkage; if not, the alleles were regarded with random associations.
Results

Considerable sequence diversity under purifying selection

Sequences of each of the 7 MLST loci had no insertions and deletions. The length of each of the 7 loci ranged from 263 bp (pryG) to 646 bp (fusA) (Table 2). The GC content of each of 7 loci ranged from 50.7 (dnaA) to 57.4 (gyrB), with a mean value of 54.5. Polymorphism was detected at 196 (5.58%) sites in total, of which only 7 had tri-allelic single nucleotide polymorphisms (SNPs) (Table 2). The diversity index π was 7.13e-3 for the concatenated sequences and ranged from 2.97e-3 (rplB) to 2.29e-2 (leuS) at different loci (Table 2). $K_a/K_s > 1$ or $< 1$ indicated positive or negative selection on the sequence tested, respectively. The $K_a/K_s$ ratio of the concatenated sequences was 0.019. fusA displayed a $K_a/K_s$ ratio of 0.104, while the $K_a/K_s$ ratios of all other loci were less than 0.043 (Table 2). These denoted that all these loci were under a strong purifying selection. The Tajima’s $D$ test of the concatenated sequences gave a value of -0.84, indicating that the sampled population was evolving randomly without evidence of positive selection (Table 2). However, rplB had a Tajima’s $D$ test value of -1.9, which was less than 0 with a statistical significance at the 0.05 level (Table 2). This suggested that rplB suffered a recent selective sweep or population expansion after a recent bottleneck.

Abundant STs and CCs

The allele number of each of the 7 loci varied from 10 (rplB) to 37 (leuS) (Table 2), generating 135 unique STs in total from the 213 strains tested, indicating that these 7 MLST loci exhibited sufficient ability for strain discrimination. The 135 STs could be assigned into 8 CCs (CC1 to CC8; 36 STs, 96 strains), 12 doubletons (24 STs, 28 strains) and 75 singletons (75 STs, 89 strains) (Fig. 1). The Chinese and non-Chinese isolates were clustered together, and no significant pattern could be found when inspecting the
relationship between STs/CCs and their years and geographic locations of isolation. The largest CC1 contained 7 STs, and the predicted founder of CC1 was ST4, containing 20 (68.97%) of the total 29 strains in CC1, which represented as the most predominant ST in this CC as commonly observed (13). However, ST50, the predicted founder of the second largest CC3, contained only one strain (4.55% of the total 22 strains in CC3). The fact that the CC3 founder ST50 was not the predominant ST in the complex might be resulted from the sampling bias.

**Clustering STs into several lineages**

A neighbor-joining phylogenetic tree was constructed from the concatenated sequences of the 135 STs with 1000 bootstraps (Fig. 2). Although these STs were clustered into several lineages, most of the bootstrap supporting values were lower than 50%. This suggested conflicting phylogenetic signals, due to frequent horizontal genetic transfer promoted by homologous recombination, leading to severe incongruence in phylogeny.

The Bayesian clustering approach implemented in STRUCTURE was used to infer the number of population units (denoted by K) within the 135 STs tested, yielding a maximal posterior value of $K=4$. This suggested that these 135 STs comprised 4 ancestral subpopulations. Accordingly, the 135 STs could be separated into 5 lineages (Fig. 2), designated L1 to L5, which contained 81, 29, 14, 3 and 8 STs respectively. STs assigned to each of the 8 CCs were interspersed throughout the five lineages, suggesting that there was no correlation between MLST and Bayesian analysis. The split network (Fig. 3) of the 135 STs contained rich parallel links, which supported the inference of 5 lineages with high frequency of recombination. As shown in Fig. 2 and 3, L4 to L5 were distantly separated from not only each other but also L1 to L3, while relatively L1 to L3 came close together. In summary, assignment of 135 STs into 5 lineages, as inferred by STRUCTURE-based Bayesian clustering, was supported by the neighbor-joining phylogenetic tree and
the split network.

**High levels of homologous recombination**

L3 to L5 were not included in estimation of recombination and linkage disequilibrium (see below) due to their limited numbers of STs contained. The $p$-values of *Phi* test for lineage L1 and L2 and the whole population (all the 135 STs) were less than 0.0001 (Table 3), indicating that homologous recombination occurred within and across lineages. The per-site $\rho/\theta$ values detected for L1, L2 and the whole population were 2.34, 3.44 and 1.189, respectively (Table 3): i) the recombination frequency was at least 2 times more likely to occur than recombination within L1 or L2; ii) the recombination rate was almost equal to the point mutation rate across the whole population; and iii) the recombination frequency within lineages was around 2 times higher than that across different lineages.

**Tendency to linkage disequilibrium**

Linkage disequilibrium was estimated with the $I_A$ parameter from STs, a statistic that was expected to be zero when the alleles were in the linkage equilibrium (free recombination). The $I_A$ values were 0.0146 ($p$-value=0.103), 0.0477 ($p$-value=0.021) and 0.0433 ($p$-value < 0.001) for L1, L2 and the whole population, respectively (Table 3), suggesting a tendency of linkage disequilibrium (non-random association of alleles at different loci) within and across lineages. Especially, the $I_A$ value detected for the whole population was significantly different from zero, indicating the credible linkage disequilibrium at the whole population level. Nevertheless, the above detecting $I_A$ values were yet relatively low, which was consistent with observation that evident recombination is occurring within the *K. aerogenes* population.

**Discussion**

*K. aerogenes* is widely distributed in the human gastrointestinal tract and in the
environment, where physiochemical properties greatly change. This requires frequent adaptation of *K. aerogenes* to its niches. *K. aerogenes* strains are genetically diverse and can be assigned into multiple CCs and lineages as detected by MLST and Bayesian analysis. Clinical *K. aerogenes* shows an epidemic population structure (displaying high levels of recombination occurred more frequently than point mutation), which was distinct from not only a clonal population (composed of strains derived from a common ancestor that diversified predominantly by mutation) but a panmictic population (in which recombination occur freely and polymorphic sites are all at linkage equilibrium) (19-21). This is the first report of MLST-based inference of genetic diversity of *K. aerogenes*. Further genome sequencing study on a large collection of *K. aerogenes* isolates will gain a deeper insight into evolution and epidemiology of this pathogen.

**Abbreviations**

MLST: multilocus sequence typing; PFGE: pulsed field gel electrophoresis; Ka/Ks: non-synonymous/synonymous rate ratio; STs: sequence types; CCs: clonal complexes; Phi test: pairwise homoplasy test; $I_A$: index of association; SNPs: single nucleotide polymorphisms

**Declarations**

**Acknowledgements**

*Klebsiella aerogenes* isolates were collected from different Hospitals. We are grateful to all collaborators that supported the studies presented in this publication.

**Authors’ contributions**

CG and FS conceived the study and designed experimental procedures. SC, YW, LZ, WL and WF performed the experiments. SC, YW, LZ, PL and DZ analyzed the data. QW and DZ contributed to reagents and materials. SC, CG and FS wrote this manuscript.

**Funding**
Conception of the study, design of the experimental procedures, sequencing, data collection, data analysis and writing of the manuscript was supported by the National Natural Science Foundation of China (81301397).

**Availability of data and materials**

All details of the strains and the sequences of multiple housekeeping genes are available in Additional file 1. All other data generated or analyzed during this study are included in this published article.

**Ethics approval and consent to participate**

All the bacterial isolates involved in this study were part of the routine hospital laboratory procedure, and ethics approval and informed consent were not required. The research involving biohazards and all related procedures were approved by the Biosafety Committee of Southwest Hospital.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Tindall BJ, Sutton G, Garrity GM. *Enterobacter aerogenes Hormaeche* and *Edwards* 1960 (Approved Lists 1980) and *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclatural type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980). Int J Syst Evol Microbiol. 2017;67(2):502-4.

2. Davin-Regli A, Pages JM. *Enterobacter aerogenes* and *Enterobacter cloacae*; versatile bacterial pathogens confronting antibiotic treatment. Front Microbiol. 2015;6:392.
3. Arpin C, Coze C, Rogues AM, Gachie JP, Bebear C, Quentin C. Epidemiological study of an outbreak due to multidrug-resistant *Enterobacter aerogenes* in a medical intensive care unit. J Clin Microbiol. 1996;34(9):2163-9.

4. Davin-Regli A, Monnet D, Saux P, Bosi C, Charrel R, Barthelemy A, et al. Molecular epidemiology of *Enterobacter aerogenes* acquisition: one-year prospective study in two intensive care units. J Clin Microbiol. 1996;34(6):1474-80.

5. Davin-Regli A, Saux P, Bollet C, Gouin F, De Micco P. Investigation of outbreaks of *Enterobacter aerogenes* colonisation and infection in intensive care units by random amplification of polymorphic DNA. J Med Microbiol. 1996;44(2):89-98.

6. Piagnerelli M, Kennes B, Brogniez Y, Deplano A, Govaerts D. Outbreak of nosocomial multidrug-resistant *Enterobacter aerogenes* in a geriatric unit: failure of isolation contact, analysis of risk factors, and use of pulsed-field gel electrophoresis. Infect Control Hosp Epidemiol. 2000;21(10):651-3.

7. Rottman M, Benzerara Y, Hanau-Bercot B, Bizet C, Philippon A, Arlet G. Chromosomal ampC genes in *Enterobacter* species other than *Enterobacter cloacae*, and ancestral association of the ACT-1 plasmid-encoded cephalosporinase to *Enterobacter asburiae*. FEMS Microbiol Lett. 2002;210(1):87-92.

8. Xiao X, Zhang J, Zhang Y, Wang L, Tan Y, Guo Z, et al. Two methods for extraction of high-purity genomic DNA from mucoid Gram-negative bacteria. Afric J Microbiol Res. 2011;5(23):4013-18.

9. Shin SH, Kim S, Kim JY, Lee S, Um Y, Oh MK, et al. Complete genome sequence of *Enterobacter aerogenes* KCTC 2190. J Bacteriol. 2012;194(9):2373-4.

10. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7.

11. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA
polymorphism data. Bioinformatics. 2009;25(11):1451-2.

12. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics. 2006;4(4):259-63.

13. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004;186(5):1518-30.

14. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 2016;33(7):1870-4.

15. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-59.

16. Bryant D, Moulton V. NeighborNet: An agglomerative method for the construction of planar phylogenetic networks. Algorithms in Bioinformatics: Springer; 2002. p. 375-91.

17. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution. 2015;1(1).

18. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics. 2000;16(9):847-8.

19. Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? Proc Natl Acad Sci U S A. 1993;90(10):4334-6.

20. Smith JM, Feil EJ, Smith NH. Population structure and evolutionary dynamics of pathogenic bacteria. Bioessays. 2000;22(12):1115-22.

21. Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, et al. Free recombination within Helicobacter pylori. Proc Natl Acad Sci U S A. 1998;95(21):12619-24.
Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242-5.

Tables

Table 1

| Gene | Product | Primer 1 | Primer 2 |
|------|---------|----------|----------|
| dnaA | Chromosomal replication initiator protein | KAE-dnaA-F1 | CGTA |
|      |         | KAE-dnaA-R1 | GGC |
| fusA | Elongation factor G | KAE-F1usA-F1 | GTTC |
|      |         | KAE-F1usA-R1 | CAT |
| gyrB | DNA gyrase subunit B | KAE-gyrB-F1 | GCC |
|      |         | KAE-gyrB-R1 | CCT |
| leuS | Leucyl-tRNA synthetase | KAE-leuS-F1 | CAA |
|      |         | KAE-leuS-R1 | CGG |
| pyrG | CTP synthase | KAE-pyrG-F2 | CGTC |
|      |         | KAE-pyrG-R2 | CTTT |
| rplB | 50S ribosomal protein L2 | KAE-rplB-F1 | ATGG |
|      |         | KAE-rplB-R1 | TCA |
| rpoB | DNA-directed RNA polymerase subunit beta | KAE-rpoB-F1 | GCT |
|      |         | KAE-rpoB-R1 | TCG |

@, annealing temperature; #, length (bp) of PCR amplicon; $, length (bp) of MLST locus.

Refer to http://pubmlst.org/kaerogenes/ for the MLST scheme and data of K. aerogenes.

Table 2

Nucleotide and allelic sequence diversity
| Locus | Length (bp) | Alleles Number | GC Content (%) | No. of SNPs | No. of tri-allelic sites | No. of synonymous sites | No. of non-synonymous sites |
|-------|-------------|----------------|----------------|-------------|--------------------------|-------------------------|---------------------------|
| dnaA  | 442         | 31             | 54.8           | 36          | 1                        | 35                      | 0                         |
| fusA  | 646         | 17             | 53.1           | 27          | 0                        | 20                      | 7                         |
| gyrB  | 434         | 30             | 57.4           | 35          | 1                        | 33                      | 1                         |
| leuS  | 578         | 37             | 56.2           | 61          | 5                        | 54                      | 2                         |
| pryG  | 263         | 13             | 57.3           | 12          | 0                        | 12                      | 0                         |
| rplB  | 607         | 10             | 53.6           | 9           | 0                        | 8                       | 1                         |
| rpoB  | 545         | 18             | 50.7           | 16          | 0                        | 15                      | 1                         |
| Concatenated | 3515 | 135 | 54.5 | 196 | 7 | 177 | 12 |

Table 3

Recombination test and estimation

| Lineage (number of STs included) | Recombination | Linkage disequilibrium |
|----------------------------------|---------------|-------------------------|
|                                  | phi          | θ                       | ρ             | ρ LB 95%   | ρ UP 95%   | p/θ   | I_A         |
| Whole population (135)          | 0            | 8.97E-03                | 1.07E-02      | 7.46E-03  | 1.42E-02  | 1.189 | 0.0433      |
| L1 (81)                         | 0            | 4.65E-03                | 1.09E-02      | 9.35E-03  | 1.24E-02  | 2.336 | 0.0146      |
| L2 (29)                         | 0            | 4.39E-03                | 1.51E-02      | 7.06E-03  | 2.25E-02  | 3.441 | 0.0477      |

Additional File

Additional file 1. List of all the strains. Show are the background details of all the bacterial strains analyzed in this study.
Population snapshot. The population snapshot of the 219 isolates is diagrammed by eBURSTv3 based on their allelic profiles (STs; designated Arabic numerals beside the circles). The sizes of the circles are related to the numbers of strains within each ST. STs with SLV relationship are linked together to form 8 CCs (CC1 to CC8), 12 doubletons and 75 singletons. The founder and cofounder genotypes are colored blue and black, respectively.
Figure 2

Diagrams denoting population structure. This figure is depicted by using iTOL3 (22). Shown are the unrooted neighbor-joining tree (with the ST numbers) and the STRUCTURE-inferred population structure (the circle outside of the tree) of the 135 STs based on their concatenated sequences of the 7 MLST loci. The percentages (≥50%, i.e. bootstrap values) of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the tree branches. The branch lengths are ignored for better visualization of bootstrap values. The background colors for the ST numbers
represent the 5 lineages. For each ST, the proportions of each of the 4 subpopulations yielded by STRUCTURE are represented by different colors.

Figure 3
Phylogenetic split network. Shown is the split network of the 135 STs based on their concatenated sequences of the 7 MLST loci, generated by the neighbor-net method using SplitsTree4.

Supplementary Files
This is a list of supplementary files associated with the primary manuscript. Click to download.
3-Additional file 1.xlsx