

Analysis of a long-term outbreak of extensively drug-resistant *Pseudomonas aeruginosa* using whole-genome sequencing and a novel Time-Place-Sequence algorithm: a molecular epidemiological study

Authors: Matthias Willmann,^{1,2§} Daniela Bezdán,^{3,4} Luis Zapata,^{3,4} Hana Susak,^{3,4} Wichard Vogel,⁵ Klaus Schröppel,¹ Jan Liese,¹ Christopher Weidenmaier,^{1,2} Ingo B Autenrieth,^{1,2} Stephan Ossowski,^{3,4} and Silke Peter.^{1,2}

¹Institute of Medical Microbiology and Hygiene, University of Tübingen, Tübingen, Germany.

²German Center for Infection Research (DZIF), partner site Tübingen, Tübingen, Germany.

³Genomic and Epigenomic Variation in Disease Group, Centre for Genomic Regulation (CRG), Barcelona, Spain.

⁴Universitat Pompeu Fabra (UPF), Barcelona, Spain.

⁵Medical Center, Department of Hematology, Oncology, Immunology, Rheumatology & Pulmonology, University of Tübingen, Tübingen, Germany.

§Corresponding author

Contact details of corresponding author:

Dr. Matthias Willmann, MSc, DTM&H
Institute of Medical Microbiology and Hygiene
Elfriede-Aulhorn-Str. 6
72076, Tübingen
Germany

will80@gmx.de

Phone: +49 7071/29-81527

Fax: +49 7071/29-5440

Abstract

Background

Pseudomonas aeruginosa is a frequent nosocomial pathogen that causes serious infections particularly in immunocompromised patients. Here we report on a long-term outbreak from 2009 to 2012 with an extensively drug-resistant *P. aeruginosa* on two wards at a university hospital in southern Germany.

Methods and Findings

By using conventional infection control methods we identified 49 *P. aeruginosa* strains including eight environmental isolates that belonged to the multi-locus sequence type 308 and carried the metallo- β -lactamase IMP-8. Whole genome sequencing (WGS) was used to gain a higher discriminatory power for the reconstruction of the outbreak. Phylogenetic analysis on the basis of a non-recombinant core genome that contained 22 outbreak-specific single nucleotide polymorphisms (SNPs) revealed a pattern of four dominant clades with a strong phylogeographic structure and allowed us to determine the potential temporal origin of the outbreak to July 2008, one year before the index case was diagnosed. The introduction of a novel Time-Place-Sequence (TPS) algorithm to estimate the probability of transmission routes enabled us to identify potential superspreaders at the root of major clades. Furthermore, environmental sources such as washing basins were found to be important for sustaining the outbreak.

Conclusions

WGS provided an increased genetic resolution to gain further insights in outbreak dynamics. Our new TPS algorithm improved estimation of transmission probabilities.

TPS results suggest that the initial expansion of dominant sublineages was driven by a few superspreaders, while environmental contamination seemed to sustain the outbreak for a long period despite regular environmental control measures. Our study indicates that active screening cultures for *P. aeruginosa* in haematological-oncological patients can be a measure to prevent outbreaks that are difficult to contain once *P. aeruginosa* has established a stable biofilm in the hospital water system.

Background

Pseudomonas aeruginosa is an opportunistic human pathogen and among the five most common bacteria in health-care associated infections in Europe [1]. Of particular concern are the rising rates of worldwide observed antibiotic resistance. In 2012, almost one third of invasive clinical *P. aeruginosa* isolates in European hospitals were reported carbapenem resistant [1]. In the same year, the EARS surveillance network report from thirty European countries stated that 13.8% of *P. aeruginosa* isolates exhibited a multidrug-resistant phenotype (MDR-PA) [2,3]. A number of studies reported high mortality rates associated with blood stream infections due to MDR-PA [4-6] and an elevated total mean economic cost for nosocomial MDR-PA acquisition (15,256 Euro MDR-PA vs. 4,933 Euros Non-MDR-PA per patient admission) [7].

This poses a serious problem particularly in hospitals that are facing long-term outbreaks with *P. aeruginosa*, most likely caused by widespread environmental contamination [8-10]. Therefore, it is crucial to fully understand the relevant in-hospital transmission routes in order to apply efficient and effective infection control measures.

One approach to identify transmission routes could be to sample and genotype *P. aeruginosa* isolates during an outbreak and assess genetic relatedness and epidemiological correlations. Common typing methods such as pulsed-field gel electrophoresis (PFGE) or PCR amplification of non-coding repetitive sequences over the bacterial genome (rep-PCR) cannot provide a sufficient discriminatory power for a detailed outbreak description. In contrast, whole genome sequencing (WGS) provides sufficient genetic resolution for an extended investigation as shown by

recent studies on methicillin-resistant *Staphylococcus aureus* (MRSA), *Mycobacterium tuberculosis* and KPC-producing *Klebsiella pneumonia* [11-15]. Additionally, the rapid decrease in costs and turnaround time of WGS will facilitate an integration of this technology into routine diagnostic procedures in the near future and may establish a new standard for infection control measures.

In the present study, we describe a long-term outbreak of extensively drug-resistant *P. aeruginosa* (XDR-PA) on two wards at a university hospital in southern Germany. Information gained by WGS was used to trace previously unsuspected transmission routes, to identify superspreaders and was compared with an outbreak reconstruction obtained by using conventional surveillance methods.

Material and Methods

Study design

Our outbreak reconstruction is based on epidemiological and genetic data. The study has been approved by the local research ethics committee of the University of Tübingen (reference number: 077/2014R). Clinical and environmental isolates of *P. aeruginosa* collected between July 2009 and March 2012 were submitted to sequencing and further analysis if they were considered to belong to the outbreak by exhibiting a XDR phenotype and carrying an IMP-8 gene. Multiple isolates from one patient were included if they were obtained on different days and from different clinical materials.

Bacterial isolation, identification and drug-susceptibility testing

Species were identified using a linear MALDI-TOF mass spectrometer (AXIMA Assurance, bioMérieux, Marcy l'Etoile, France), supplemented by Vitek 2 system identification (bioMérieux, Marcy l'Etoile, France). *In vitro* bacterial susceptibility testing was conducted by the Vitek 2 system (bioMérieux, Marcy l'Etoile, France) and interpreted according to the EUCAST guidelines [16-18]. Colistin susceptibility was measured using the Etest gradient diffusion method and interpreted following CLSI breakpoints [19]. Further Etests were performed to determine the minimum inhibitory concentrations (MICs) for meropenem, doripenem, fosfomycin, and amikacin. Strains were stored at -80°C until analysis.

Library preparation, whole genome sequencing and genome assembly

Genomic DNA was sheared by Covaris (Covaris, Woburn, USA) to obtain 300 bp – 400 bp fragments. DNA libraries were prepared with TruSeq DNA LT Sample Prep Kit (Illumina, San Diego, USA) with 24 different barcodes using the standard protocol and were sequenced at 2x50 bp on an Illumina HiSeq2000 (Illumina, San Diego, USA). *SOAPdenovo version r1.05* was selected as assembly tool [20].

Genome assembly, variant and transmission analysis

A core genome alignment for all 49 isolates was generated using the progressiveMauve algorithm [21]. The isolate WGS-E4 exhibited a high sequencing and assembly quality and was defined as outbreak reference genome. It was sampled from an environmental source (siphon water) and considered a circulating hospital strain. Sequences were aligned to the WGS-E4 reference genome by bowtie2 [22] and core genome SNPs were called using the samtools package [23].

A transmission was assumed to have potentially occurred between patient A and patient B, if the detection of the outbreak strain in patient A preceded the one of

patient B in time. The probability of a transmission between patient A and B was based on four criteria, including information about time and place of both patients as well as on the information about variations in the genome sequences of their isolates. This new analysis method was named Time-Place-Sequence (TPS) algorithm. 1) Criterion one was considered fulfilled if both patients resided at the same ward with a minimum timely overlap of 24 hours before the outbreak strain was detected in patient B. 2) Criterion two was fulfilled if patient B resided in the same room that patient A occupied up to two weeks prior to patient B. 3) Criterion three was fulfilled when patient A and B stayed in the same room with a minimum timely overlap of 24 hours before the outbreak strain was detected in patient B. 4) Close genetical relatedness between strains was used as criterion four (see supplementary Appendix).

A transmission was considered *possible* when criterion one was fulfilled. In contrast, a transmission was considered *probable* when either criterion two or criterion three was fulfilled. Fulfilment of criterion 4 alone was regarded as a *probable* transmission. A transmission between patient A and B was rated as *predicted* when criterion 4 was fulfilled in combination with any of the three epidemiological criteria. If no criterion was fulfilled the transmission probability was considered unknown. The TPS algorithm was applied by two independent investigators.

Results and Discussion

Outbreak description and interventions

The University Hospital Tübingen is a 1500-bed tertiary teaching hospital, which has two major wards for the treatment of patients with haematological-oncological

diseases. The outbreak occurred on both wards. Ward 1 is a general ward with double rooms. Ward 2 is an intensive care unit for patients receiving stem cell transplantations and consists of single rooms.

The index patient (patient 1) was diagnosed with a *P. aeruginosa* blood stream infection in July 2009 on ward 1 (Figure 1B). The isolate was resistant to all antimicrobial agents tested except colistin, revealing an XDR phenotype according to the ECDC/CDC definition [3]. A second patient suffered from bacteraemia in November 2009 at the same ward, exhibiting again an XDR-PA with the same susceptibility pattern. At that time, the two cases were considered epidemiologically unrelated. However, new cases in the second quarter of 2010 and the increasing incidence in the third quarter pointed to an outbreak. Therefore, an investigation by the infection control team was launched (supplementary Appendix, Figure S1). The initial program of targeted measures included an enhanced emphasis on hand hygiene, a program of deep room cleaning, the introduction of a weekly screening program as well as a screening at admission, and the installation of an isolation zone with three rooms at ward 1 where infected or colonized patients were treated by medical staff who was not allowed to work in other ward areas at the same day.

The incidence of XDR-PA remained high in the beginning of 2011 despite the implementation of the initial infection control program (supplementary Appendix, Figure S1), suggesting that contaminated environmental sources could sustain the outbreak. Isolation of XDR-PA from siphon water at that time indicated that the outbreak was associated with colonisation of washing basins as reported in a previous outbreak [9]. This led to a replacement of all siphons under washing basins and to the simultaneous installation of a fully automated cleaning device for thermal disinfection (85 – 93°C) and vibration cleaning (50 Hz) of the new siphons to prevent

biofilm formation. The number of new XDR-PA cases dropped following these interventions and remained at an average baseline level of one new case per quarter at the time of submission, reflecting the transformation from an epidemic into an endemic state. A total number of 34 patients became colonized or infected with the outbreak strain between July 2009 and December 2013. A full list of interventions can be found in the supplementary Appendix (Table S2).

Molecular detection allowed the identification of IMP-8 metallo- β -lactamase in all XDR-PA outbreak isolates, suggesting a clonal occurrence. Metallo- β -lactamases (MBLs) confer resistance to all beta-lactams except aztreonam and can be considered as a major mechanism of resistance in the outbreak isolates. IMP-8 had been found in clinical isolates of *Enterobacteriaceae* and *Acinetobacter baumannii* in China and Taiwan [24-26] and in an environmental isolate of *Pseudomonas medocina* in a Portuguese hospital [27], but has not been described in clinical isolates of *P. aeruginosa*. To further confirm that the IMP-8 carrying XDR-PA isolates are closely related we performed multi locus sequence typing (MLST). All IMP-8 XDR-PA isolates belonged to the same sequence type 308 which has not been previously reported to be associated with a resistance phenotype. However, neither analysis of epidemiological data nor MLST provided further information regarding the temporal origin of the outbreak, transmission routes or factors responsible for the sustainment of the outbreak. In order to gain new insights into the nature of the outbreak we performed whole genome sequencing to increase the genetic resolution.

Core genome investigation

Forty-one IMP-8 XDR-PA isolates from 26 patients were collected for genome sequencing between July 2009 and March 2012, spanning the peak of the outbreak

in the first quarter of 2011. Additionally, we sequenced eight outbreak strains isolated from siphon water, washing basins and toilet siphons on both wards.

Twenty-four outbreak-specific core genome SNPs were detected when using WGS-E4 as reference. A total of 17 SNPs were located in gene regions, 11 of which cause amino acid changes. Seven SNPs were located in intergenic regions (Table 1).

Phylogeny of the outbreak strains

Prior to phylogenetic analysis we generated a non-recombinant core genome by detecting regions of recombinant origin using BratNextGen [28] and regions with prophage content using PHAST [29]. The removal of regions of recombinant origin resulted in the loss of one SNP at genome position 226,279. Another SNP at genome position 632,368 was additionally removed because of its location in a phage tail protein. The final non-recombinant core genome with a length of 5,823,688 bp contained 22 SNPs and was used for phylogenetic analysis.

We applied Bayesian statistics to infer a time-measured phylogeny (Figure 1A). The analysis revealed a strong phylogeographic structure with its root location at ward 1 with a 90.7% probability. In the beginning of 2010 the tree splits into four major clades spreading mainly over one ward, except clade three which extends over both wards. These results are in agreement with the hypothesis that sublineages of outbreak strains develop independently and are transmitted within their predominant location, although some isolates were occasionally introduced to the non-clade-related ward, most likely due to frequent patient transferrals between both wards as reflected in the epidemiological map (Figure 1B). Of note, multiple isolates sampled from the same patient over a short interval could belong to different clades, indicating intra-individual strain variation that can hamper transmission analysis since one

patient could transmit more than one variant simultaneously at any given time point. In addition, analysis is hampered by the possibility of multiple transmissions to one patient, which is the case in patient 3. In May 2010, this patient was initially detected positive on ward 1 with an isolate that belonged to clade one (P3A). In February 2011, an isolate (P3D) was sampled that was associated with clade three and was possibly transmitted in the outpatient department. Maximum-likelihood phylogeny (supplementary Appendix, Figure S2), the minimum spanning tree of the outbreak (supplementary Appendix, Figure S3) as well as the *SeqTrack* algorithm (supplementary Appendix, Figure S4) illustrate similar results.

Based on the Bayesian analysis, the median mutation rate was estimated to be 2.02×10^{-7} SNPs/bp/year (95% highest posterior density interval (HPD) $9.83 \times 10^{-8} - 3.18 \times 10^{-7}$ SNPs/bp/year). This is equivalent to 1.18 SNPs/year (95% HPD 0.57 – 1.85 SNPs/year) which is in the range of previous estimates for *P. aeruginosa* [30,31] but significantly lower than for *Staphylococcus aureus* (1 SNP per ~6 weeks) [11]. The low mutation rate is likely the reason why many nodes in the maximum-likelihood phylogeny remain weakly supported. It is uncertain whether this lower discriminatory power would provide enough resolution for tracing transmission routes in *P. aeruginosa* outbreaks that span a shorter time period or examine fewer isolates than our study.

Applying Bayesian statistics, we have estimated that the most recent common ancestor of the outbreak dates to July 2008 (95% HPD: October 2006 - May 2009).

Temporal origin of the outbreak

Assuming that an individual is exposed to only one variant, July 2008 could represent the time point when the unknown primary patient became a carrier. It is possible that

this primary patient has stayed regularly on ward 1 for up to one year before the index case emerged, posing a permanent but hidden threat to immunocompromised high risk patients. Another scenario is also possible. Most metallo- β -lactamases genes like IMP-8 are inserted in integron regions along with other determinants of resistance and can thus transfer along with mobile DNA elements associated with such integrons (plasmids and transposons) [32]. In fact, IMP-8 has recently been discovered on a plasmid in a clinical strain of *Citrobacter freundii* which potentially enables it to be horizontally transferred to other species in our hospital [33]. It is thus conceivable that the ST308 strain has acquired the IMP-8 gene from an environmental, perhaps unculturable strain approximately in July 2008 or within the time range of the 95% HPD. The new resistant strain could have remained hidden in the waste water system, and some patients might have been already colonized before the index case became apparent with a severe blood stream infection in July 2009. Witney et al. have found a susceptible *P. aeruginosa* clinical isolate highly similar to XDR isolates that appeared six years later in the context of an outbreak [34], showing that such a scenario is possible.

Reconstruction of transmission routes

We have investigated whether information about the genetic relatedness based on SNP data could improve tracing of transmission routes in addition to epidemiological information gained by conventional surveillance methods. The probability of a transmission was estimated by the novel TPS algorithm. Using the epidemiological criteria alone, transmission between patients was considered possible in 23 cases (Figure 2A, blue arrows) and probable in three cases (Figure 2A, brown arrows). Information gained from SNP data considerably advanced the understanding of transmission routes at outbreak onset (Figure 2B, orange arrows). It revealed seven

additional cases of probable transmissions, where no transmission was suspected based on the epidemiological criteria alone. Figure 2C shows transmission probabilities when using the TPS algorithm by combining epidemiological and genetical criteria. Three transmissions that had been scored as possible or probable based on epidemiological criteria alone could be predicted with a high probability by additionally applying the genetic criterion (Figure 2C, pink arrows). Looking at two of these cases, it can be noted that the patients stayed at ward 1 at the same time but not in the same room (P3 -> P7 and P9 -> P12). This is indicative for a transmission via medical personnel. In the third case two patients shared the same room (criterion three) for which the transmission could have occurred via personnel, direct patient contact or through environmental contamination (P8 -> P10). Surprisingly, the probable transmission of patient 9 to patient 11, who have stayed in the same room in a consecutive order (criterion two), was not confirmed by genetic data, suggesting that environmental contamination cannot always be assumed as transmission route in such a situation. However, a transmission might have taken place despite the genetic distance of the isolates. Sampling more than one isolate of patient 9 could have revealed a strain more closely related to the strain of patient 11. Thus, we still score the transmission as probable due to the valid epidemiological criterion.

Another interesting discovery was the genetic distance between the isolates of patient 1 and 2 which differed by three SNPs. Given the estimated mutation rate, it is unlikely that the potential descendent strain of patient 1 could have evolved so fast in the short period between the sampling of both strains. We hypothesise that more than one variant was already circulating during the early phase of the outbreak, potentially introduced by an unknown primary patient who became colonized some time before the admission to the hospital. The circulation of different variants in the

early phase of the outbreak could account for the weak correlation between the accumulation of SNPs since the first outbreak isolate and time (Pearson correlation = 0.34, $p = 0.017$).

Figure 3 illustrates probabilities of transmission routes for all outbreak patients between 2009 and 2012 based on the TPS algorithm. In contrast to the isolate of patient 2, which was not further transmitted, we identified three potential superspreaders with a high number of probable or predicted transmissions (patient 3, 7 and 9). The reasons for this divergence are unclear. It can be speculated whether a low neutrophil count in combination with the administration of broad spectrum antibiotics or a high frequency of room transfers could facilitate an enhanced pathogen shedding. However, we have not found significant differences between patient 2 and the superspreaders regarding these potential risk factors. The presence of superspreader was confirmed by using the *SeqTrack* algorithm that has recently been proposed for tracing transmission routes with genomic data [35] (supplement Appendix, Figure S4).

An example of particular interest is the probable transmission between patient 1 and 21. In 2009, patient 21 stayed in a room for three days directly after patient 1. In 2011, an outbreak strain was isolated from a rectal swab of patient 21 only two days after readmission (Figure 1B). The isolate differed from the strain of patient 1 in only one SNP. We have also obtained isolates from patient 15 and 16 that were genetically identical to the strain of patient 21, but the absence of any epidemiological link makes a transmission via patient 15 and 16 less likely. This indicates that a transmission could have already occurred in 2009 and that the outbreak strain established a stable colonisation in a host for almost two years.

Another predicted relation was found between patient 15 and 17. The single isolate of patient 17 has an identical core genome with a late isolate of patient 15 (P15D). Both isolates were obtained within three days. The small time gap caused uncertainty regarding the direction of the transmission but points to the possibility that patients tested positive for the outbreak strain can have transmitted their strain variants to patients whose carrier status was earlier detected. This assumption appears even more likely in the light of long and hidden colonisations as it has been suggested for patient 21 and shows that patients can be exposed multiple times and can be colonized with different variants of an outbreak strain. This complexity of transmissions is missed when only one isolate per individual is obtained.

Factors involved in sustainment of the outbreak

The recovery of eight outbreak isolates from the waste water system and washing basins proved a contamination of the hospital environment and outlined the presence of potential reservoirs for IMP-8 XDR-PA. Outbreak strains were also sampled from the inner surface of siphons after their removal and replacement, demonstrating biofilm formation. Especially alarming was the finding of environmental strain E4 in siphon water from a room where two genetically identical clinical isolates were recovered from patient 12 five weeks earlier while the patient stayed in this room (P12 A and P12C). At that time a deep cleaning program with regular application of tube cleaners was already initialized. This highlights the problems associated with inefficient eradication of biofilms and indicates that environmental contamination can act as reservoir for recurrent exposure of patients and sustain hospital outbreaks with *P. aeruginosa*.

This is supported by the fact that the core nucleotide diversity of the environmental samples ($\pi = 4.48 \times 10^{-7}$, $SE = 1.28 \times 10^{-7}$) was 97% of the total for all isolates ($\pi =$

4.62×10^{-7} , SE = 1.46×10^{-7}). Finding that the majority of the overall diversity is present in the small environmental subset suggests a dynamic exchange between patient and environmental isolates, supporting the hypothesis that the outbreak was sustained by environmental contamination. This was confirmed by the low degree of population differentiation between the patient and environmental subset (coefficient of differentiation = 0.017, SE = 0.079).

Outbreak-specific SNPs represent strain evolution over the duration of the outbreak and possibly an adaptation to hospital-related factors. The gene sequence changes found in *oprD* and *nrdD* were of particular interest as they occurred in more than one isolate (Table 1). OprD is an outer membrane protein and functions as a common channel for the intake of amino acids and peptides as well as carbapenem antibiotics [36,37]. Structural changes in OprD have been associated with a decrease in carbapenem susceptibility [38,39]. Another study investigated the fitness of 300,000 *P. aeruginosa* PA14 transposon insertion mutants in mice. *OprD* mutants were identified that did not only reveal an increased carbapenem resistance, but were also associated with an enhanced fitness for mucosal colonization and with systemic spread. Other *oprD* mutants demonstrated an elevated resistance to acid pH and an increased cytotoxicity against macrophages. Furthermore, exposure to human serum killed some *oprD* mutants to a lesser extent, which indicated that OprD might act as a crucial factor for the establishment of bacteremia [40]. It has also been reported that OprD can be sialylated, leading to a reduction of complement deposition on the bacterium when incubated with human serum [41].

Another variation has been observed in the *nrdD* gene, which encodes a ribonucleotide reductase enzyme (RNR). The three classes of RNRs are responsible for the conversion of nucleoside 5'-di- or triphosphates to deoxyribonucleotides which

are essential for DNA synthesis and DNA repair. NrdD belongs to the class III RNRs, and it has been shown that NrdD is of crucial importance for anaerobic growth of *P. aeruginosa*. Growth under such anaerobic conditions results in elongated shaped *P. aeruginosa* rods, which feature robust biofilm formation [42,43].

Since OprD plays a role in strain survival and pathogenic capability and NrdD facilitates anaerobic growth and is potentially involved in biofilm establishment, mutations in those genes could reflect an evolutionary development in response to clinical or environmental stress within the hospital.

Effectiveness of infection control measures

In spite of the success of our infection control interventions that transformed the epidemic into an endemic situation, it is debatable whether regular active screening cultures (ASCs) for *P. aeruginosa* in haematological-oncological patients could be a recommended practice. Recently, we have observed an overall in-hospital mortality in 38% of patients with *P. aeruginosa* blood stream infections but we could not verify that metallo- β -lactamase production or an MDR phenotype are independent predictors of mortality [44]. Our study indicates that, although the impact of multidrug-resistant *P. aeruginosa* on mortality in serious infections is still a controversial topic [45], it would be of great interest to focus infection control measures on the prevention of any infection with *P. aeruginosa*, regardless of the susceptibility pattern. Despite the lack of evidence that ASCs are an effective measure in endemic settings with MDR-PA [46], it is safe to assume that early detection of increasing carrier rates, including the individual antibiotic susceptibility profiles, enables to take action before high risk patients would suffer from severe infections, possibly during an outbreak situation.

Conclusions

We have investigated a long-term outbreak with IMP-8 XDR *P. aeruginosa* between 2009 and 2012. In our study we integrated SNP data and conventional surveillance methodology which allowed us to gain insights into the outbreak that would have been otherwise not accessible. Our approach enabled us to not only clarify transmission routes and identify possible superspreaders, but also provided details regarding the temporal origin of the outbreak. We therefore propose whole genome sequencing as a new standard in infection control. However, genetic data can be difficult to interpret in the absence of a clear epidemiological link, particularly when the true length of a colonisation status and the entire genetic intra-host variability of the outbreak strain are unknown. For that reason we scored transmission probabilities by implementing a new TPS algorithm that takes epidemiological correlation and genetic relatedness into account and can thus provide an appropriate estimation.

Our results demonstrate the importance of an appropriate surveillance system for patients on haematological-oncological wards which would enable an early observation of a rising incidence in otherwise hidden carriers. Missing such carriers increases the risk of an outbreak which is difficult to contain once *P. aeruginosa* has been established in a hospital water system.

Acknowledgments

We thank the director, physicians, laboratory and nursing staff of the involved medical wards. We further extend our gratitude to Kerstin Fischer and Nadine Hoffmann for their assistance in laboratory testing and analysis.

References

1. European Centre for Disease Prevention and Control. Point prevalence survey of health care associated infections and antimicrobial use in European acute care hospitals. Stockholm: ECDC; 2013.
2. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2012. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm: ECDC; 2013.
3. Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, et al. (2012) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect* 18: 268-281.
4. Aloush V, Navon-Venezia S, Seigman-Igra Y, Cabili S, Carmeli Y (2006) Multidrug-resistant *Pseudomonas aeruginosa*: risk factors and clinical impact. *Antimicrob Agents Chemother* 50: 43-48.
5. Tam VH, Rogers CA, Chang KT, Weston JS, Caeiro JP, et al. (2010) Impact of multidrug-resistant *Pseudomonas aeruginosa* bacteremia on patient outcomes. *Antimicrob Agents Chemother* 54: 3717-3722.
6. Tumbarello M, Repetto E, Treccarichi EM, Bernardini C, De Pascale G, et al. (2011) Multidrug-resistant *Pseudomonas aeruginosa* bloodstream infections: risk factors and mortality. *Epidemiol Infect* 139: 1740-1749.
7. Morales E, Cots F, Sala M, Comas M, Belvis F, et al. (2012) Hospital costs of nosocomial multi-drug resistant *Pseudomonas aeruginosa* acquisition. *BMC Health Serv Res* 12: 122.

8. Corvec S, Poirel L, Espaze E, Giraudeau C, Drugeon H, et al. (2008) Long-term evolution of a nosocomial outbreak of *Pseudomonas aeruginosa* producing VIM-2 metallo-enzyme. *J Hosp Infect* 68: 73-82.
9. Snyder LA, Loman NJ, Faraj LA, Levi K, Weinstock G, et al. (2013) Epidemiological investigation of *Pseudomonas aeruginosa* isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. *Euro Surveill* 18.
10. Loveday HP, Wilson JA, Kerr K, Pitchers R, Walker JT, et al. (2014) Association between healthcare water systems and *Pseudomonas aeruginosa* infections: a rapid systematic review. *J Hosp Infect* 86: 7-15.
11. Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, et al. (2010) Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 327: 469-474.
12. Koser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 366: 2267-2275.
13. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, et al. (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 364: 730-739.
14. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, et al. (2012) Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* 4: 148ra116.
15. Nubel U, Roumagnac P, Feldkamp M, Song JH, Ko KS, et al. (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 105: 14130-14135.
16. EUCAST (2012) Breakpoint tables for interpretation of MICs and zone diameters.

17. Leclercq R, Canton R, Brown DF, Giske CG, Heisig P, et al. (2011) EUCAST expert rules in antimicrobial susceptibility testing. *Clin Microbiol Infect.*
18. EUCAST (2012) EUCAST Disk Diffusion Method for Antimicrobial Susceptibility Testing.
19. CLSI (2012) Performance Standards for Antimicrobial Susceptibility Testing. Wayne, PA: Clinical and Laboratory Standards Institute.
20. Luo R, Liu B, Xie Y, Li Z, Huang W, et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1: 18.
21. Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5: e11147.
22. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357-359.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079.
24. Yan JJ, Ko WC, Wu JJ (2001) Identification of a plasmid encoding SHV-12, TEM-1, and a variant of IMP-2 metallo-beta-lactamase, IMP-8, from a clinical isolate of *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 45: 2368-2371.
25. Wang H, Guo P, Sun H, Wang H, Yang Q, et al. (2007) Molecular epidemiology of clinical isolates of carbapenem-resistant *Acinetobacter* spp. from Chinese hospitals. *Antimicrob Agents Chemother* 51: 4022-4028.
26. Yan JJ, Lee NY, Chen HM, Wang MC, Ko WC, et al. (2013) Bloodstream infections caused by IMP-8-producing Enterobacteriaceae isolates: the need for clinical laboratory detection of metallo-beta-lactamases? *Eur J Clin Microbiol Infect Dis* 32: 345-352.

27. Santos C, Caetano T, Ferreira S, Mendo S (2010) First description of bla IMP-8 in a *Pseudomonas mendocina* isolated at the Hospital Infante D. Pedro, Aveiro, Portugal. *Res Microbiol* 161: 305-307.
28. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, et al. (2012) Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Res* 40: e6.
29. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: a fast phage search tool. *Nucleic Acids Res* 39: W347-352.
30. Marvig RL, Johansen HK, Molin S, Jelsbak L (2013) Genome analysis of a transmissible lineage of *pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* 9: e1003741.
31. Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, et al. (2011) Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A* 108: 7481-7486.
32. Cornaglia G, Giamarellou H, Rossolini GM (2011) Metallo-beta-lactamases: a last frontier for beta-lactams? *Lancet Infect Dis* 11: 381-393.
33. Peter S, Wolz C, Kaase M, Marschal M, Schulte B, et al. (2014) Emergence of *Citrobacter freundii* carrying IMP-8 metallo- β -lactamase in Germany. *New Microbes and New Infections* 2: 42-45.
34. Witney AA, Gould KA, Pope CF, Bolt F, Stoker NG, et al. (2014) Genome sequencing and characterization of an extensively drug-resistant sequence type 111 serotype O12 hospital outbreak strain of *Pseudomonas aeruginosa*. *Clin Microbiol Infect.*

35. Jombart T, Eggo RM, Dodd PJ, Balloux F (2011) Reconstructing disease outbreaks from genetic data: a graph approach. *Heredity (Edinb)* 106: 383-390.
36. Trias J, Nikaido H (1990) Outer membrane protein D2 catalyzes facilitated diffusion of carbapenems and penems through the outer membrane of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 34: 52-57.
37. Trias J, Nikaido H (1990) Protein D2 channel of the *Pseudomonas aeruginosa* outer membrane has a binding site for basic amino acids and peptides. *J Biol Chem* 265: 15680-15684.
38. Ochs MM, Bains M, Hancock RE (2000) Role of putative loops 2 and 3 in imipenem passage through the specific porin OprD of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 44: 1983-1985.
39. Huang H, Jeanteur D, Pattus F, Hancock RE (1995) Membrane topology and site-specific mutagenesis of *Pseudomonas aeruginosa* porin OprD. *Mol Microbiol* 16: 931-941.
40. Skurnik D, Roux D, Cattoir V, Danilchanka O, Lu X, et al. (2013) Enhanced in vivo fitness of carbapenem-resistant oprD mutants of *Pseudomonas aeruginosa* revealed through high-throughput sequencing. *Proc Natl Acad Sci U S A* 110: 20747-20752.
41. Khatua B, Vleet JV, Choudhury BP, Chaudhry R, Mandal C (2014) Sialylation of outer membrane porin protein D: a mechanistic basis of antibiotic uptake in *Pseudomonas aeruginosa*. *Mol Cell Proteomics*.
42. Sjoberg BM, Torrents E (2011) Shift in ribonucleotide reductase gene expression in *Pseudomonas aeruginosa* during infection. *Infect Immun* 79: 2663-2669.

43. Lee KM, Go J, Yoon MY, Park Y, Kim SC, et al. (2012) Vitamin B12-mediated restoration of defective anaerobic growth leads to reduced biofilm formation in *Pseudomonas aeruginosa*. *Infect Immun* 80: 1639-1649.
44. Willmann M, Kuebart I, Marschal M, Schroppel K, Vogel W, et al. (2013) Effect of metallo-beta-lactamase production and multidrug resistance on clinical outcomes in patients with *Pseudomonas aeruginosa* bloodstream infection: a retrospective cohort study. *BMC Infect Dis* 13: 515.
45. Vardakas KZ, Rafailidis PI, Konstantelias AA, Falagas ME (2013) Predictors of mortality in patients with infections due to multi-drug resistant Gram negative bacteria: the study, the patient, the bug or the drug? *J Infect* 66: 401-414.
46. Tacconelli E, Cataldo MA, Dancer SJ, De Angelis G, Falcone M, et al. (2014) ESCMID guidelines for the management of the infection control measures to reduce transmission of multidrug-resistant Gram-negative bacteria in hospitalized patients. *Clin Microbiol Infect* 20 Suppl 1: 1-55.

Tables

Table 1. Twenty-four core genome single nucleotide polymorphisms (SNPs) of 49 *Pseudomonas aeruginosa* outbreak isolates. The specified SNP substitutions relate to the reference genome WGS-E4. bp indicates base pairs.

Core SNP position	Region	Enzyme code	Isolates	SNP	Substitution
226279*	-	-	P4	T->C	Intergenic
393592	Permease	-	P4	A->G	Phe187Ser
437952	CheW-like domain protein	EC:2.7.3	E3	C->T	Synonymous
632368*	Phage tail protein	-	P3A, P6, P11, P3B, P13A, P3C, P3D, P14A, P15A, E2, P15B, P14B, P16A, P15C, P15D, P15E, P3E, P16B, P13B	G->A	Synonymous
1000582	Porin d (oprD)	-	P1, P2, P3A, P4, P5, P6, P7, P8, P10, P11, P3B, P13A, P3C, P3D, P14A, P15A, E2, P15B, P14B, P16A, P15C, P17, P15D, P15E, P3E, P16B, P20, P21, P13B, P23, E6, E7	T->C	Asp106Gly
1063891	-	-	P6	G->A	Intergenic
1069103	Transcriptional regulator	EC:3.6.1.15	P16B	T->G	Val270Gly
1104673	Mfs family transporter	-	P14A, P14B	T->G	His160Pro
1315012	ABC-type antimicrobial peptide transport ATPase component	-	P11, E6	C->G	Synonymous
1537804	Lipoprotein	-	P25	G->A	Ser57Leu
1913494	Anaerobic ribonucleoside-triphosphate reductase (nrdD)	EC:1.17.4.2	P1, P2, P4, P9A, P11, P12B, P9B, P3D, P15A, E2, P15B, P16A, P16B, P18, P19A, P21, P22, E5, E7, P24, P26	C->A	Gly124Val
2389760	ABC transporter ATP-binding protein	EC:3.6.1.3	P19B	A->G	Glu148Gly
2684713	Chemotaxis transducer	-	P15C	C->G	Ile376Met
3239794	-	-	P13A, P13B	T->C	Intergenic
4049373	Pyochelin synthetase (pchF)	-	P1, P3B, P3C, E5	C->A	Synonymous
4103474	DNA-directed RNA polymerase subunit beta (RpoB)	EC:2.7.7.6	P16B	T->A	Gln152Leu
4867436	-	-	P25	G->T	Intergenic
4972016	Spermidine putrescine abc transporter substrate-binding protein	-	P3C, P3D	G->A	Synonymous
5236482	-	-	P2	G->C	Intergenic
5272104	Exodeoxyribonuclease III	EC:3.1.11.2	E5, P24, P25, E8	A->C	Thr225Pro
5577991	-	-	P12B	C->A	Intergenic
5660814	Transcriptional regulator	-	P26	T->C	Leu69Pro
5740900	-	-	P11	A->T	Intergenic
5749062	Hypothetical protein	-	P2, P5, P11, E1, P19A	G->T	Synonymous

* These SNPs were removed for phylogenetic analysis.

Figure legends

Figure 1. Bayesian phylogenetic reconstruction with location, divergence date estimation and epidemiological data of 49 *Pseudomonas aeruginosa* outbreak isolates

(A) Tree structure is based on a Bayesian phylogenetic statistical approach constructed from the core genome alignment that includes 22 outbreak specific SNPs. The node and branch colours present the known (leaves) and predicted (internal branches) location of the isolates. OPD stands for outpatient department. The brown numbers indicate the four major clades that emerged in the beginning of 2010 with a brown dashed line at the beginning of the clades. The orange dashed line marks the temporal origin of the tree in July 2008. Strains of patients with more than two study isolates were coloured. The material from which the strain was cultured is displayed at the right-sided box. Screening material for *P.aeruginosa* were pharyngeal and rectal swabs as well as stool samples. Patient isolates start with a "P" while environmental isolates start with an "E".

(B) Epidemiological map of 27 patients on two wards in the haematology department. The strain of patient x (PX) was not available for sequencing. The green and red colour indicates the stay at ward one or two. The asterisk indicates the first record of a patients outbreak isolate. The blue dashed line represents the start of the outbreak. Grey vertical blocks display times when there were no known carriers of the outbreak strain on both wards.

Figure 2. Estimated probability of transmission routes for the first 12 patients involved in the outbreak

Each patient is outlined as a circle. The colour of the circle indicates the patient's location at the first finding of the outbreak isolate. Red colour stands for ward one and green colour stands for ward two. The line colour and thickness reflect the estimated strength of transmission probability between two patients. Criteria 1 – 4 are in detail described in the material and methods section as well as in the supplement Appendix. Briefly, criterion 1: same ward, same time; criterion 2: consecutively same room within 14 days; criterion 3: same room, same time; criterion 4: close genetic relatedness. The presence of criteria is indicated by symbols in close proximity to the line of transmission between two patients. (A) Probability of transmission based on the three epidemiological criteria alone. (B) Probability of transmission based on the genetic relatedness criterion alone. (C) Probability of transmission based on the three epidemiological criteria in combination with the genetic relatedness criterion (Time-Place-Sequence algorithm). Only *probable* and *predicted* transmissions are shown. *Possible* transmissions were graphically omitted to focus on the major transmission probabilities.

Figure 3. Estimated probability of transmission routes for 27 outbreak patients between 2009 and 2012 using the Time-Place-Sequence (TPS) algorithm

Each patient is outlined as a circle. The colour of the circle indicates the patient's location at the first finding of the outbreak isolate. Red colour stands for ward one, green colour stands for ward two and blue colour stands for the outpatient department. The line colour and thickness reflect the estimated strength of the transmission probability between two patients. Criteria 1 – 4 are in detail described in the material and methods section as well as in the supplement Appendix. Briefly, criterion 1: same ward, same time; criterion 2: consecutively same room within 14 days; criterion 3: same room, same time; criterion 4: close genetic relatedness. The presence of criteria is indicated by symbols in close proximity to the line of transmission between two patients.