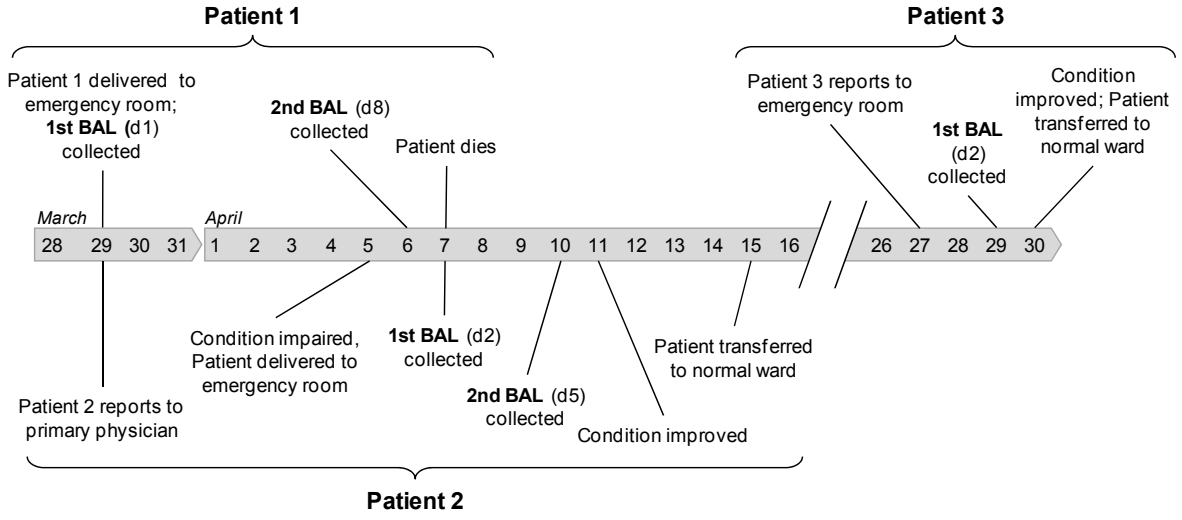
Rapid metagenomic diagnostics of a suspected outbreak of severe pneumonia in Northern Germany

Technical Appendix

1. Timeline and clinical/patient information

A timeline of events and time points when BAL samples were collected is shown below (Supplemental Figure 1). Supplemental Tables 1, 2 and 3 provide clinical data and results of conventional diagnostic tests performed for each of the patients.



Supplemental Figure 1: Timeline of events

Supplemental Table 1: Clinical information for patient 1

Date	***Information***
03/29/13	• Patient 1 (49-year old police officer, immunocompetent) is admitted to the emergency room of the University Medical Centre Hamburg-Eppendorf after having suffered for one week from symptoms of pneumonia. • Patient is diagnosed with ARDS and mechanically ventilated in intensive care unit. • CRP (238mg/l) and CK (402) levels elevated, PCT: 71 • Leukocytosis (mainly neutrophils) • Antibitotic treatment with meropenem and clarithromycin initiated. • 1st bronchioalveolar lavage (BAL) collected, diagnostic tests performed on BAL: o PCR tests negative for: FLU A/B, PIV 1-4, RSV, CoV, Adenovirus, Bocavirus, Mycoplasma pneumonia, Chlamydophila pneumonia, Legionella o Cultivation negative after 48h o PCP negative
03/30/13	• CT shows bilateral laminar infiltrates; atypical pneumonia
04/01/13	• Bronchial secretion cultivation, negative after 48h

04/03/13	• 2nd bronchial secretion cultivation, negative after 48h
04/06/13	• 2nd BAL collected, cultivation of BAL negative after 48h
04/07/13	• Patient dies from multiple-organ failure

Supplemental Table 2: Clinical information for patient 2

Date	***Information***
03/29/13	• Patient 2, a 29-year old immunocompetent police officer from the same county as patient 1, presents to primary physician with malaise, tiredness, cough with discharge, diarrhoe • Patient receives Cefuroxim (Cephalosporin 2nd gen.; gram neg. bacteria)
04/05/13	• Impairment of medical condition, dispnoea, patient delivered to local hospital outside Hamburg; increased CO_2/decreased O_2 levels • Patient transferred to intensive care unit of the University Medical Centre Hamburg-Eppendorf, Patient is diagnosed with ARDS and mechanically ventilated • Antibiotic treatment with meropenem, doxycycline and clarithromycin initiated. • CRP 238mg/l; CK increased • CT shows bilateral laminar inhomogeneous infiltrates, mostly in the lower and middle lobe. • 1st BAL collected; bronchial secretion cultivation negative after 48h
04/06/13	• Serology tests indicate prior vaccination (IgG) against Flu, tests negative for: Legionella Ag, ELISA Chlamydiae , Leptospirosis, Mycoplasma, Q-fever
04/07/13	• 1st BAL collected, cultivation of BAL negative after 48h
04/10/13	• 2nd BAL collected, PCR tests negative for: FLU A/B, PIV 1-4, RSV, CoV, Adenovirus, Bocavirus, Mycoplasma pneumonia, Chlamydophila pneumonia, Legionella
04/11/13	• Condition improved; Patient is extubated and separated from ventilator
04/15/13	• Patient is moved from intensive care to normal ward

Supplemental Table 3: Clinical information for patient 3

Date	***Information***
04/27/13	• Patient 3, a 22-year old police officer (immunocompetent) who had been sharing office space with patient 2, reports to the emergency room with symptoms of pneumonia • Patient is admitted to intensive care unit without requiring mechanical ventilation • Antibiotic treatment with meropenem, doxycycline and clarithromycin initiated.
04/28/13	• CT shows bilateral laminar and inhomogeneous lung infiltrates • Leukocytosis (mainly neutrophils) • CRP: 255mg/l; CK increased; PCT: 11.35; Pro-BNP: 2959 • Patient suspected to suffer from systemic inflammatory response syndrome (SIRS) • Serological tests negative for: Legionella Ag, ELISA Chlamydiae , Leptospirosis, Mycoplasma, Q-fever
04/29/13	• 1st bronchioalveolar lavage (BAL) collected: o PCR tests negative for FLU A/B, PIV 1-4, RSV, CoV, Adenovirus, Bocavirus, Mycoplasma pneumonia, Chlamydophila pneumonia, Legionella o BAL cultivation negative after 48h

	○ PCP negative • Blood culture negative after 5d
04/30/13	• Condition improved; Patient moved from intensive care to normal ward
05/02/13	• Patient discharged

2. Next generation sequencing (NGS) of BAL samples

We subjected material from the 2nd BAL sample of patient 1 (collected 8 days after admittance and 1 day before the death of the patient) and the 1st BAL samples of patients 2 and 3 (collected 2 days after admittance; see timeline above) to unbiased RNA and DNA sequencing. The three BAL samples as well as the 1st BAL of patient 1 were furthermore analyzed by metagenomic 16S rRNA gene amplicon sequencing as described below.

2.1 Unbiased sequencing of total RNA and DNA

2.1.1 Preparation of RNA-seq and DNA-seq libraries

Total DNA and RNA were extracted from a total volume of 1ml BAL sample using the ZR DNA/RNA Kit (Zymo Research) as per the manufacturer's instructions. Illumina NGS libraries for RNA-seq were generated using the SCRIPT SEQ^{TM} v2 RNA Seq Kit (Epicentre Biotechnologies) using either 12.5 ng of DNAse treated total RNA or rRNA depleted RNA (RiboMinus™ Eukaryote Kit, Invitrogen) as an input. Adaptor-ligated fragments were purified using AMPure XP beads (Agencourt). For the preparation of DNA-seq libraries, RNAse treated DNA was subjected to sonication (15x30s duty cycles, 30s off phase between cycles) with a Bioruptor NGS sonication device (Diagenode) to generate fragments with an average size of 500 bp. DNA-seq libraries were subsequently generated with the NEXTFLEX ^{Tm} CHIPSeq (BIO Scientific) Kit using 50ng of sonicated DNA as input. Adaptor-ligated DNA-seq or RNA-seq libraries were amplified (12-14 cycles), purified and measured with a Qubit instrument (Invitrogen). The average fragment size of the libraries was determined with a BioAnalyzer High Sensitivity LabChip (Agilent Technologies). Diluted libraries were multiplex-sequenced on an Illumina MiSeq (2x250 bp paired-end run, 2-3 million reads per sample) or an HiSeq instrument (2x100 bp paired-end run, 80-120 million reads per sample).

2.1.2 Analysis of RNA-seq and DNA-seq data by homology-based database searches

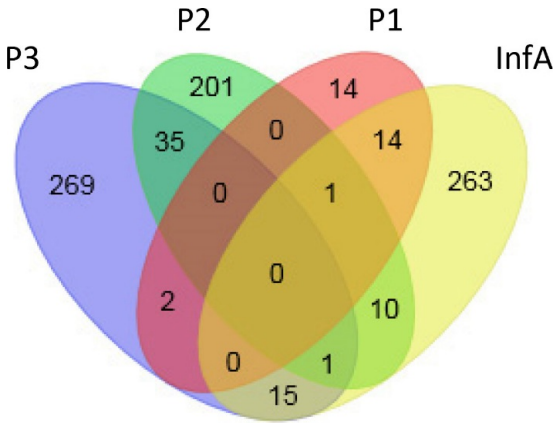
In order to eliminate sequences of host origin, reads were first aligned to the human reference assembly (NCBI 37.2) using Bowtie2 (v2.1.0) [1]. The Inchworm assembler (r2013-02-25) from the Trinity package [2] was subsequently used to assemble contigs from filtered reads. All contigs with a length of at least 300bp were retained for further analysis. For taxonomic classification, the filtered sequence contigs were aligned to a database of all known viruses, plasmids, bacteria and fungi using BLASTN and BLASTX (v2.2.28+) [3]. Contigs with significant hits were assigned the taxonomic rank of the highest scoring BLAST hit. To estimate read abundance for individual contigs and assigned taxonomic categories, filtered

reads were re-mapped to the assembled sequence contigs. All reads that did not map to any contig were furthermore directly mapped to the viral/bacterial/fungal database using Bowtie2. The results from the BLASTN analysis, which readily identified a C. psittaci infection in patient 2 but not patients 1 and 3, are depicted in Figure 1 of the manuscript. Similarly, neither BLASTX analysis nor direct mapping of reads not assembled into contigs did provide evidence for the presence of potentially pathogenic taxa that were shared among the three patients, as would be expected in case of an outbreak (data not shown).

Our metagenomic detection pipeline is primarily based on BLAST-based annotation of *de novo* assembled sequence contigs. Compared to direct mapping of reads to the all viruses, bacteria and fungi, this approach provides significantly higher discriminatory power due to increased length of the analyzed sequences. With regard to the identification of bacteria in complex diagnostic samples, however, it should be noted that the method performs much better with RNA-seq data than with DNA-seq data. This is due to the fact that DNA reads are distributed across the entire length of bacterial genomes, whereas the majority of RNA reads originates from the much smaller fraction of the genome that is highly transcribed. At a given non-saturating read depth, RNA-seq reads therefore achieve much higher local coverage and can be more efficiently assembled into longer sequence contigs compared to DNA reads. We therefore routinely perform DNA-seq only to increase the chances of detecting DNA viruses, for which assembly is facilitated due to small genome sizes combined with frequently high copy numbers. These considerations are supported by the fact that in the three cases analyzed here, *de novo* assembled DNA-seq contigs almost exclusively originated from anelloviruses, apathogenic DNA viruses that are present at high copy numbers in almost all bodily fluids (see Figure 1 in the manuscript).

2.1.3 Analysis of RNA-seq and DNA-seq data by pairwise BLAST

While the above pipeline allows identification of pathogens that bear significant BLASTN or BLASTX homology to known organisms, it cannot recognize novel pathogens which are not (or only very distantly) related to known bacteria, viruses or fungi. Among all contigs that did



Supplemental Figure 2: Pairwise BLASTClust alignment of sequence contigs that could not be assigned to known taxonomic units. The Venn diagram shows the number of clusters that were shared (sequence homology >= 97%) between each of the samples.

not exhibit significant homology to known sequences, we therefore additionally performed pairwise BLASTN alignments using the BLASTClust (v2.2.26) program. In addition to the three patients, we also included an unrelated control group of unclassified sequence contigs from 5 BAL samples of patients suffering from InfA infection. In case of an outbreak of a novel agent, this direct alignment strategy should reveal the presence of contigs that are shared among the suspected outbreak samples, but not the unrelated control samples. We did not detect the presence of any such contigs (Supplemental Figure 2), strongly arguing against an outbreak scenario.

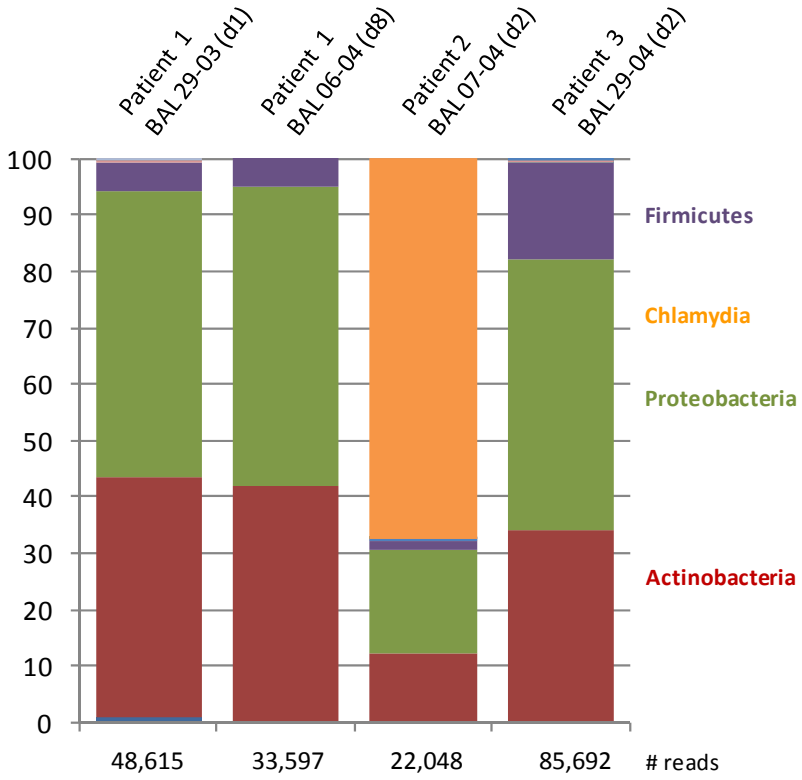
2.2 16S rRNA gene amplicon sequencing

2.2.1 Preparation of 16S amplicon libraries

16s rRNA amplicon sequencing was performed with 50ng DNA as starting material as recently described [4]. Briefly, the V4 region of the 16S rRNA gene was amplified using sequence specific primers which additionally included the Illumina flow cell adapters. In addition, the reverse primer contained a barcode sequence which allows pooling of different samples. PCR reaction was performed using a mastercycler EP gradient (Eppendorf) under the following conditions: 94°C 3min, 35 cycles at 94°C 30sec, 50°C 60sec, 72°C 90sec, followed by 5min 72°C. PCR products were gel purified using a NucleoSpin® Gel and PCR Clean-up Kit (Machery-Nagel). Sequencing was performed on a paired end (150-cycle) MiSeq run.

2.2.2 Analysis of 16S sequencing data

We employed the UPARSE analysis pipeline to cluster amplicon reads into Operational Taxonomy Units (OTUs) as previously described [5]. Only quality filtered and de-replicated fragment pairs with an overlap of at least 12bp between forward and reverse reads were considered for the analysis. Supplemental Figure 3 shows the results of the analysis on the phylum level. In accord with the results of our total RNA-seq data, the OTU Chlamyida was only detected in the BAL of patient 2, where it accounted for 67% of all assigned reads. Remaining phyla were mainly limited to Actinobacteria, Proteobacteria and Firmicutes, a distribution which is in line with recent comparative analyses of the bacterial microbiota in the respiratory tract of healthy patients or patients with viral infection [6, 7]. It is to be expected that quantitative assignments differ between 16S sequencing and total RNA-seq analysis, given that the former is based on amplification of the genomic 16S rRNA locus, whereas the latter involves sequencing of actual RNA transcripts. Nevertheless, 16S sequencing results were generally in good accord with total RNA-seq data, although members of the phylum Bacteriodetes (which is also part of the normal respiratory tract microbiome [6, 7]) where more prominently detected in the RNAseq analysis. As for the RNA-seq analysis, 16S analysis did not identify sequences in the samples from patients 1 and 3 that could be assigned to pathogenic bacteria.



Supplemental Figure 3: OTUs based on UPARSE analysis: Shown are percentages of reads within the taxonomy level phylum

3. Confirmation of *Chlamydophila psittaci* infection

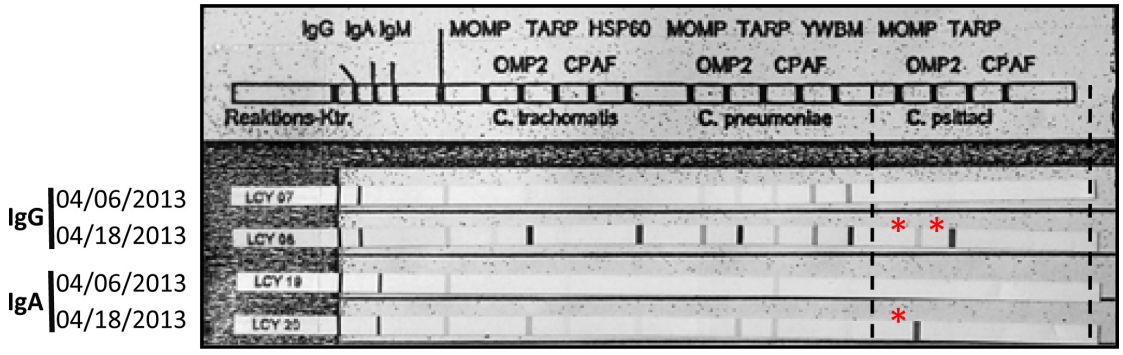
Diagnosis of *C. psittaci* infection is usually based on clinical symptoms and serological evidence [8, 9], or on direct detection of bacterial DNA. There are no commercial PCR tests for the detection of *C. psittaci*, and seroconversion can take up to 8 weeks. Following our initial metagenomic analysis, we performed quantitative real-time PCR (qPCR) using envB-specific primers as described before [10], as well as two primer pairs (c_13 and c_84) specific for 23S rRNA *C. psittaci* sequence contigs from our RNA-seq analysis (C_13 forward: 5'-GGTTGAGGGAGAGTCTATG; C_13 reverse: 5'-GAAGATTCCCTACTGCTG; C_84 forward 5'-GTTACGGGGATAATTTGC; C_84 reverse 5'-TTAGGTTGTGGTTAAGGG). Real time PCR was performed with the Qiagen Rotorgene Q 5plex. The reaction was carried out in 10µl 2x SyBr Green mix Thermo Life Sciences, 10µmol of each primer and 2µl extracted DNA. Cycling conditions were as follows: 10min 95°C; cycling (45 repeats) 15sec 95°C, 20sec 56°C, 20sec 72°C. The cycling was followed by a melting curve analysis.

In agreement with the metagenomic analysis, only the samples from patient 2 were positive. Ct values were significantly lower in the BAL collected at day 2 relative to that collected at day 5 post admittance, indicating successful antibiotic treatment (See Supplemental Table 4).

Supplemental Table 4: qPCR confirmation of C. psittaci infection

	PCR env B C. psittaci	PCR 23S rDNA C. psittaci c_13	PCR 23S rDNA C. psittaci c_84
Patient 1, 1st BAL 29.03 (d1)	Neg	Neg	Neg
Patient 1, 2nd BAL 06.04 (d8)	Neg	Neg	Neg
Patient 2, 1st BAL 07.04 (d2)	+ ct 26	+ ct 20.9	+ ct 21.29
Patient 2, 2nd BAL 10.04 (d5)	+ ct 29	+ ct 24.4	+ ct 27.77
Patient 3, 1st BAL 29.04 (d2)	Neg	Neg	Neg

Serum samples collected from patient 2 at day 5 and day 17 were also subjected to a LINEblot analysis to detect antibodies against *C. trachomatis*, *C. pneumoniae* and *C. psittaci*. While the serum from day 5 did not show any reactivity, a strong IgA- and IgG-mediated response to *C. psittaci* major outer membrane protein (MOMP) was detected at day 17, indicating the patient had seroconverted (Supplemental Figure 4). In contrast, serum samples from patient 3 collected 5 months after initial hospitalization remained negative, thus confirming the PCR results as shown above.



Supplemental Figure 4: The recomLine Chlamydia test system (Mikrogen Diagnostics) contains recombinant species-specific antigens separated on nitrocellulose strips. Shown are strips incubated with diluted serum samples from patient 2 at two different time points. Whereas no reactivity against *C. psittaci* antigens is visible in the first serum sample, both IgA (MOMP) and IgG antibodies (OMP2) were detected in the second serum sample taken 12 days later. Although IgG reactivity against several *C. pneumoniae*

and *C. trachomatis* antigens also appear in the second sample the serologic response is likely due to *C. psittaci* infection because a strong IgA reaction was seen against MOMP of *C. psittaci* only and OMP2, to which IgG antibodies were directed mainly, is known to induce cross-reacting antibodies.

References

1. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
2. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome.* Nat Biotechnol, 2011. **29**(7): p. 644-52.
3. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
4. Caporaso, J.G., et al., *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4516-22.
5. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nature methods, 2013. **10**(10): p. 996-8.
6. Chaban, B., et al., *Characterization of the upper respiratory tract microbiomes of patients with pandemic H1N1 influenza.* PLoS One, 2013. **8**(7): p. e69559.
7. Lemon, K.P., et al., *Comparative analyses of the bacterial microbiota of the human nostril and oropharynx.* mBio, 2010. **1**(3).
8. Smith, K.A., et al., *Compendium of measures to control Chlamydophila psittaci (formerly Chlamydia psittaci) infection among humans (psittacosis) and pet birds, 2005.* J Am Vet Med Assoc, 2005. **226**(4): p. 532-9.
9. Wong, K.H., S.K. Skelton, and H. Daugharty, *Utility of complement fixation and microimmunofluorescence assays for detecting serologic responses in patients with clinically diagnosed psittacosis.* J Clin Microbiol, 1994. **32**(10): p. 2417-21.
10. Okuda, H., et al., *Detection of Chlamydophila psittaci by using SYBR green real-time PCR.* J Vet Med Sci, 2011. **73**(2): p. 249-54.