Supplementary information for "Parallel bacterial evolution within multiple patients ties novel genes to pathogenesis."

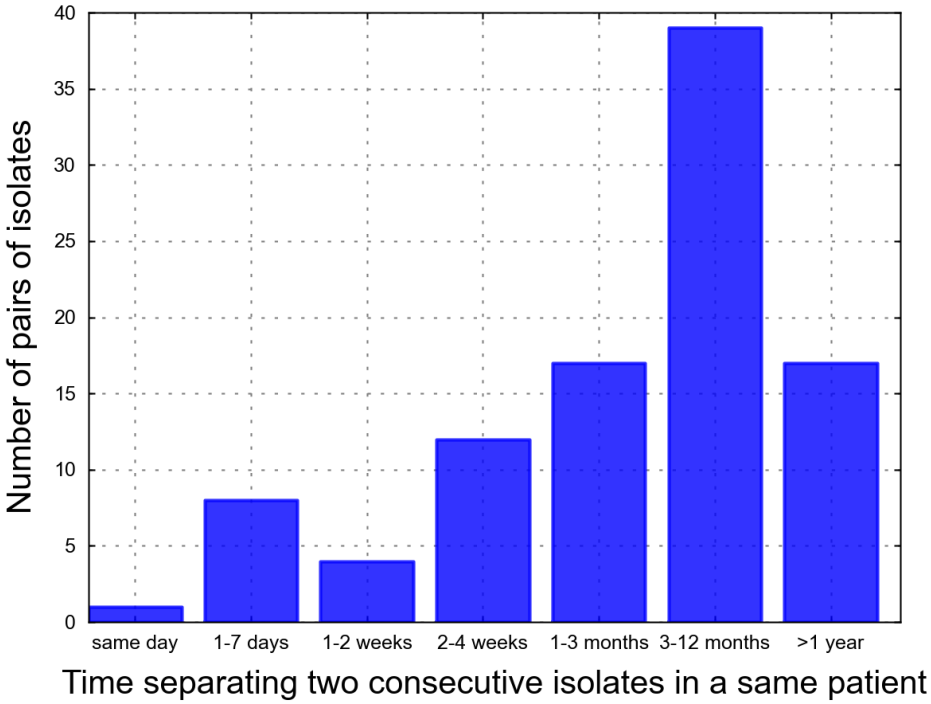
Tami D. Lieberman[1,*], Jean-Baptiste Michel[1,2,*], Mythili Aingaran[3], Gail Potter-Bynoe[4], Damien Roux[5], Michael R. Davis, Jr.[6], David Skurnik[5], Nicholas Leiby[1], John J. LiPuma[7,8], Joanna B. Goldberg[6], Alexander J. McAdam[9], Gregory P. Priebe[3,5,10], Roy Kishony[1,11].

[1] Department of Systems Biology, Harvard Medical School, Boston, MA, USA.
[2] Program for Evolutionary Dynamics, Harvard University, Cambridge, MA, USA.
[3] Department of Medicine, Division of Infectious Diseases, Children's Hospital Boston, Boston, MA, USA.
[4] Infection Prevention & Control, Children's Hospital Boston, Boston, MA, USA.
[5] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA.
[6] Department of Microbiology, University of Virginia Health System, Charlottesville, VA, USA.
[7] Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI, USA.
[8] Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA.
[9] Department of Laboratory Medicine, Children's Hospital Boston, Boston, MA, USA.
[10] Department of Anesthesia, Division of Critical Care Medicine, Children's Hospital Boston, Boston MA, USA.
[11] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.
[*] These authors contributed equally to this work.

Correspondence and requests for materials regarding genotyping, phenotyping and data analysis should be sent to RK (Roy_Kishony@hms.harvard.edu); correspondence and requests for materials regarding clinical samples, medical information and LPS pathogenesis should be sent to AJM (Alexander.McAdam@childrens.harvard.edu) and GPP (Gregory_Priebe@childrens.harvard.edu)

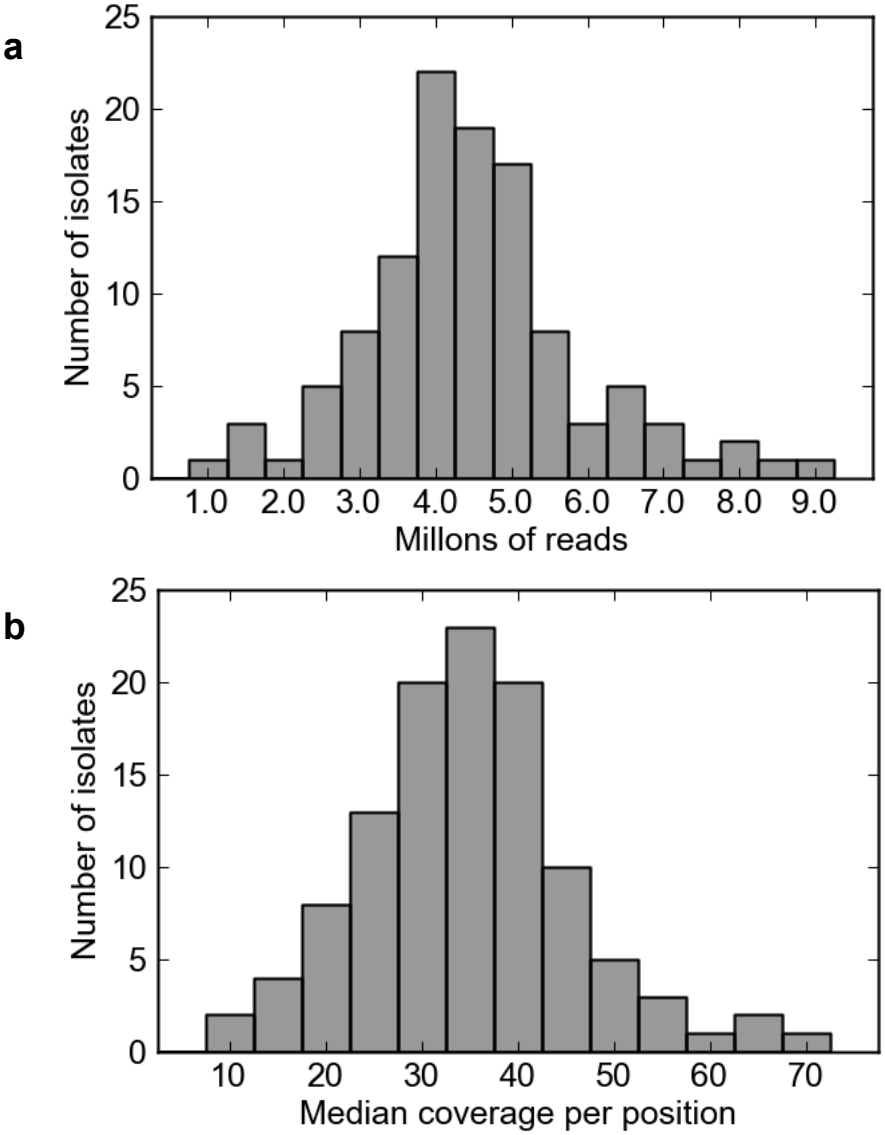
Supplementary Figures

Supplementary Figure 1



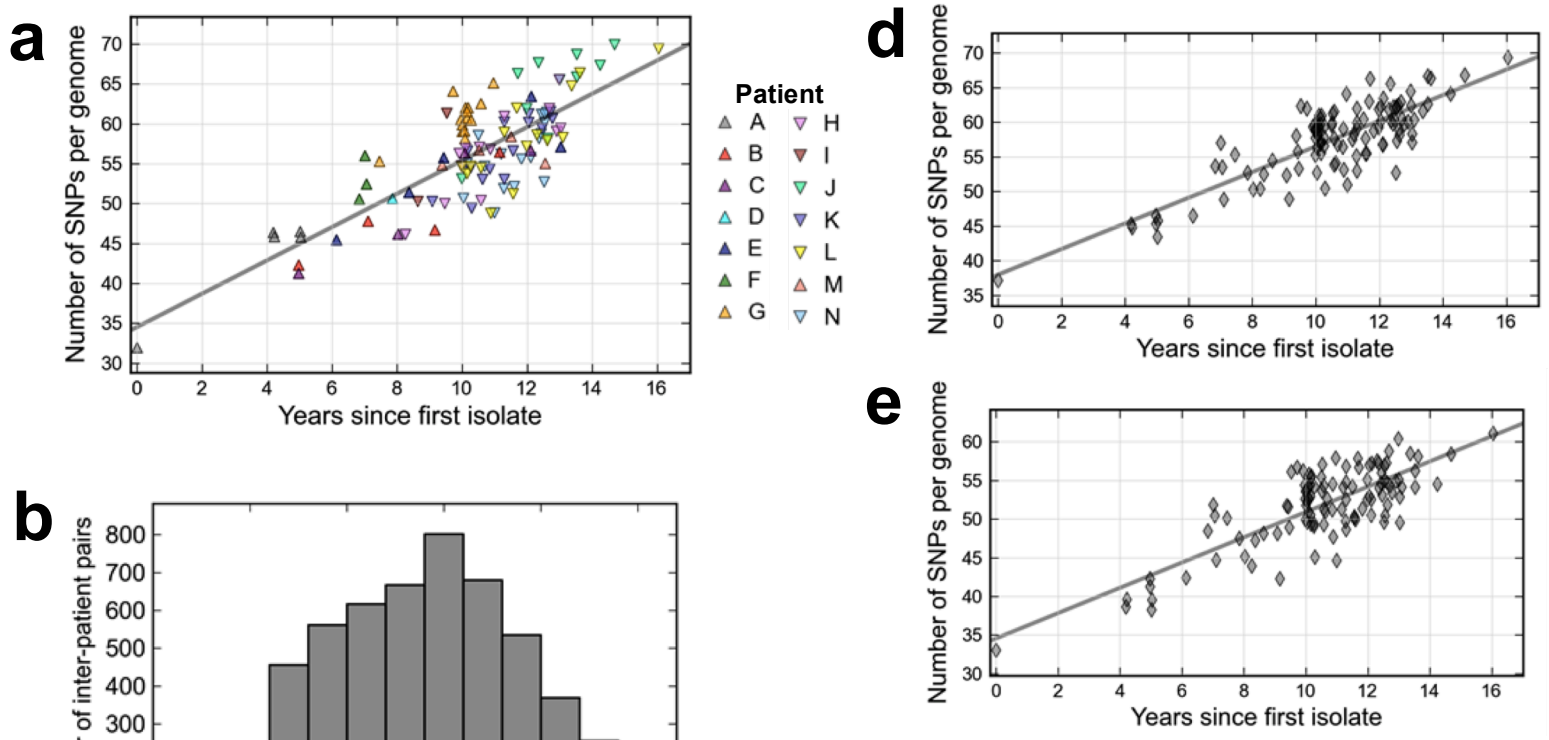
Supplementary Figure 1. Distribution of time separating two consecutive isolates from the same patient. We plot the number of pairs of isolates taken from the same patient within the time windows shown on the x-axis. On average, two consecutive isolates were separated by 3.5 months, but in 9 cases isolates were recovered during the same week.

Supplementary Figure 2



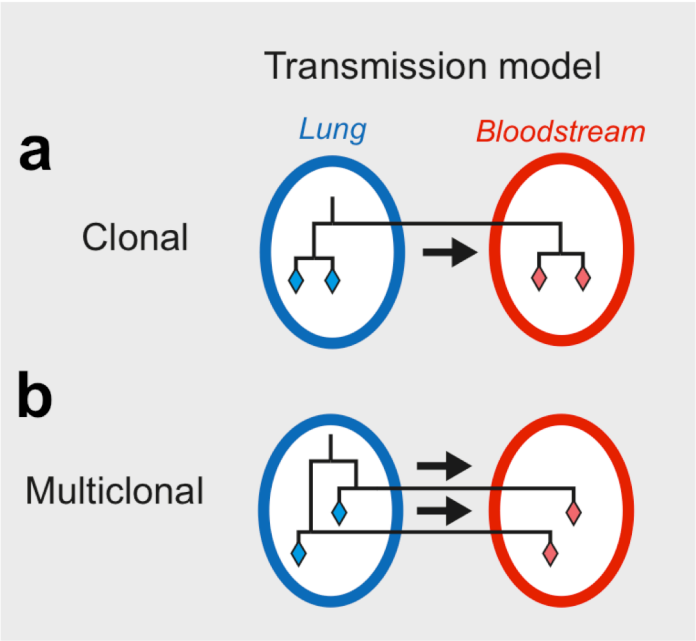
Supplementary Figure 2. Whole genome sequencing of 112 *B. dolosa* isolates at 37x depth. a, Distribution of the number of 75-nucleotide single end reads obtained per bacterial isolate. On average, 4.6 million reads were generated for each isolate. **b,** Distribution of median read depth per position for each isolate. On average, genomes were read to a depth of 37x, providing high quality reads for 93% of the genome.

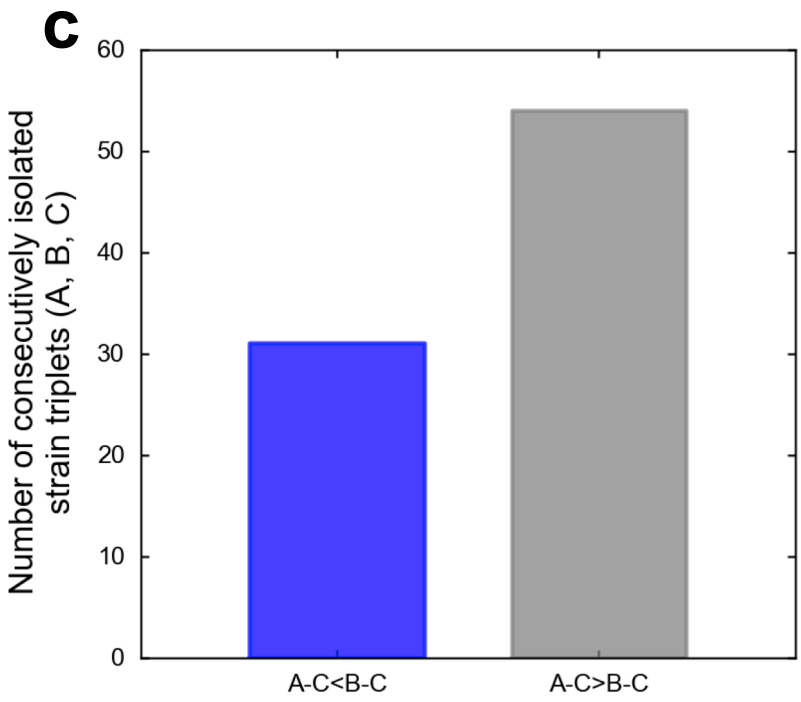
Supplementary Figure 3



Supplementary Figure 3. Genetic distance between bacterial isolates and rate of evolution. a, The rate at which bacterial mutations accumulate is consistent across patients. We plot the genetic distance of sequenced strains to the outgroup as a function of the time of isolation (years since first isolate in this study). Isolates from the same patient are represented in the same color (see legend). Within patients, isolates accumulate mutations at a rate similar to that observed for all 112 isolates. **Isolates recovered from the same patient are genetically closer.** Distributions of SNP between pairs of isolates taken from different patients **(B)** and from the same patient **(c).** On average, 28 SNPs separate two isolates from different patients; isolates from the same patient differ by an average 9.5 SNPs. **Mutations on genes which are not found under selection accumulate linearly with time. d** We exclude from the calculation of genetic distance those 17 genes under strong selection (which received 3 mutations or more overall). Mutations accumulate linearly with time (Pearson correlation: .83, fixation rate: 1.9 mutations per year). **e,** We exclude from the calculation of genetic distance the 45 genes which received more than one mutation. Mutations accumulate linearly with time (Pearson correlation: .79, fixation rate: 1.6 mutations per year).

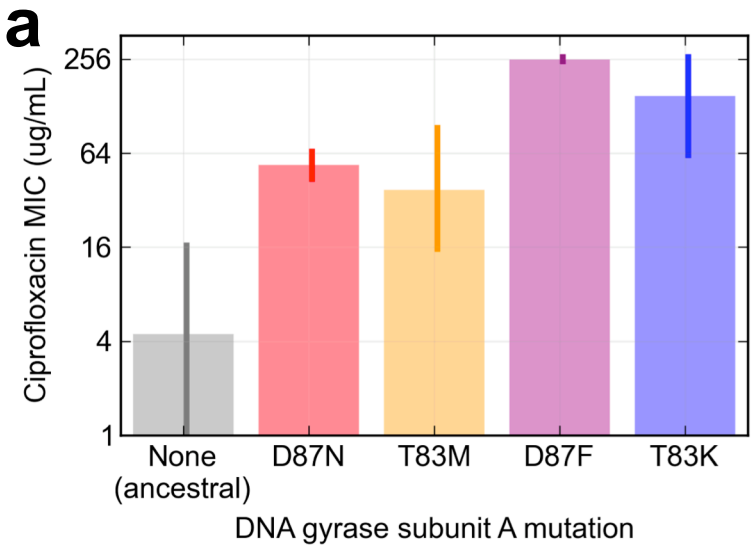
Supplementary Figure 4

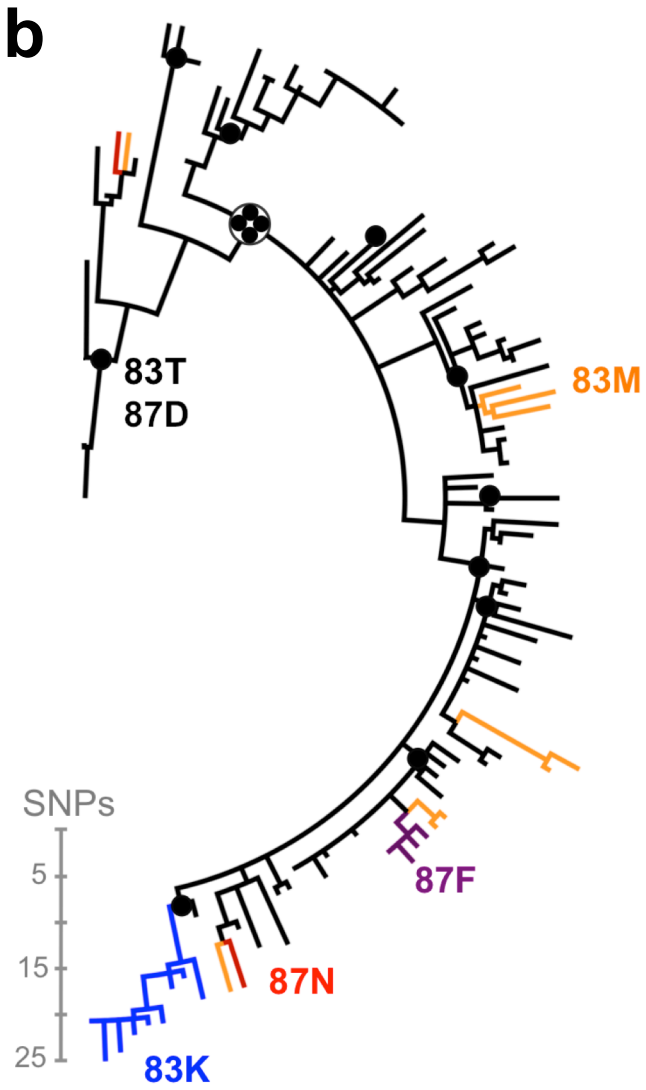




Supplementary Figure 4. Within-patient bacterial populations. a, b, Models of transmission from the lungs to the blood during bacteremia. Monoclonal transmission (**a**): a single clone from the lung is passed to the bloodstream (single arrow). Multiclonal transmission (**b**): multiple clones are transmitted from the lung to the blood, during one or several events (several arrows). **c, Phylogenetic analysis provides indirect evidence that the *B. dolosa* population in the CF lung is polymorphic.** We considered all the triplets of isolates (A,B,C) recovered successively from a patient. For each triplet, we asked whether C is genetically closer to A than to B (A-C < B-C), based on the phylogenetic tree in **Figure 2b**. A positive answer is not expected if the lung was a homogeneous environment for bacteria. We represent in gray the number of triplets that are in line with a homogeneous lung and in blue the number of triplets that contradict this expectation. We find discordance between phylogeny and time in 37% of cases (31 / 85 triplets). This supports a polymorphic model of the lung, where several distinct genotypes coexist.

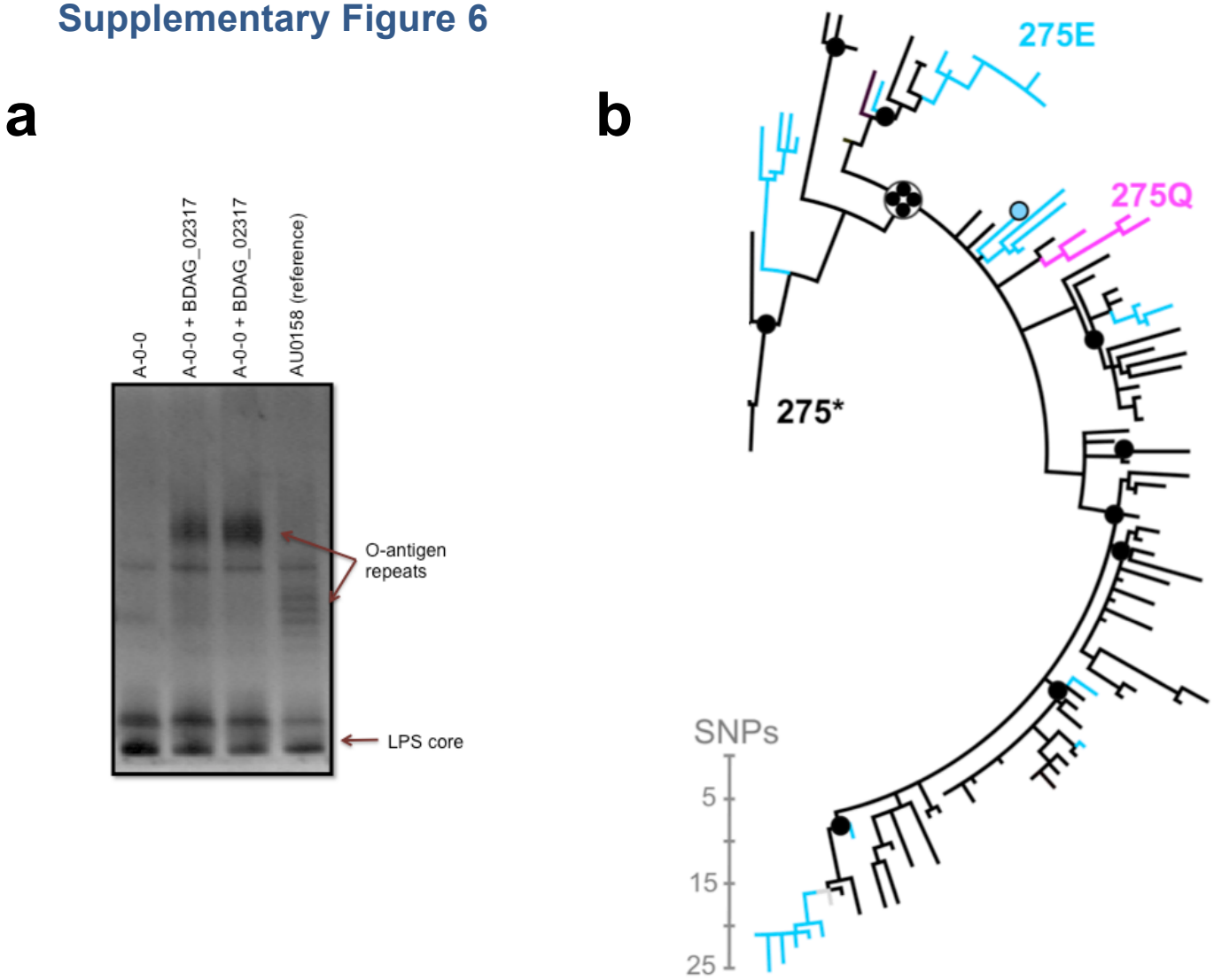
Supplementary Figure 5





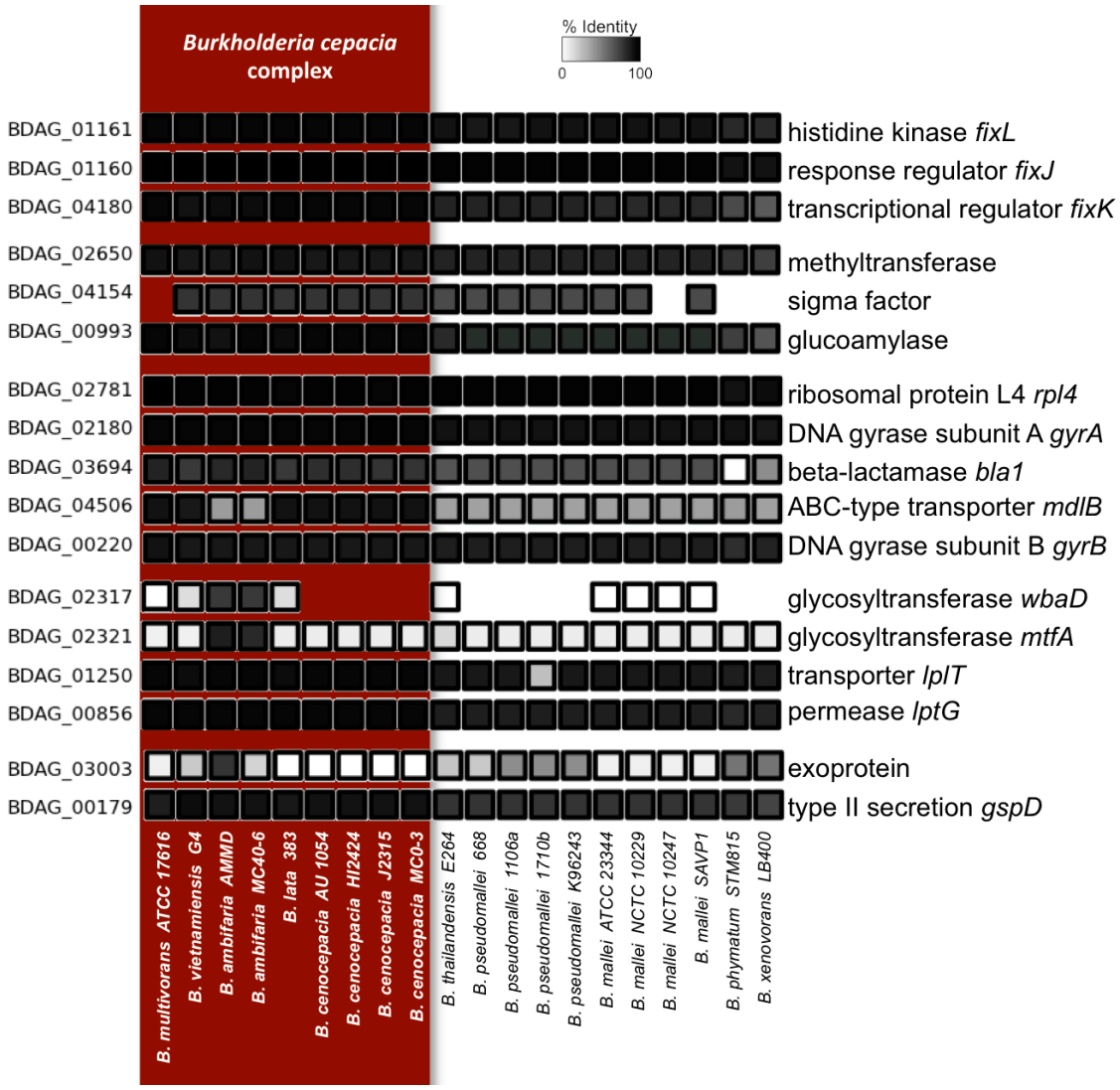
Supplementary Figure 5. Resistance to ciprofloxacin as a function of mutations in BDAG_02180 (gyrA). a, We assayed the resistance to ciprofloxacin to all 112 isolates (precision: 2-fold). We plot on a logarithmic scale the average resistance of these isolates as a function of the mutation observed in BDAG_02180.Thin bars indicate the range of MIC observed for mutation. The number of isolates corresponding to each mutation is as follows: wildtype: 88, D87N: 2, T83M: 9, D87F: 4, T83K: 9. **b, The last common ancestors of each patient have drug-sensitive genotype.** We display the phylogeny of the 112 strains. Branches are colored as a function of the inferred genotype in the gene BDAG_02180 at codons 83 and 87: black corresponds to the wild type—a threonine (T) at position 83 and aspartate (D) at position 87 (D), blue indicates a lysine (K) at position 83, orange indicates an methionine (M) at position 83, red indicates a asparagine (N) at position 87, and purple indicates a phenylalanine (F) at position 87. The last common ancestors of strains from each patient are shown as dots; they are colored with their inferred BDAG_02180 genotype. Even though mutations occurred in 6 patients, all the last common ancestors bear the wild-type BDAG_02180, which is associated with gene sensitivity (Figure 3A). This indicates that the fluoroquinolone resistance evolves within patients, rather than being transmitted from patient-to-patient.

Supplementary Figure 6



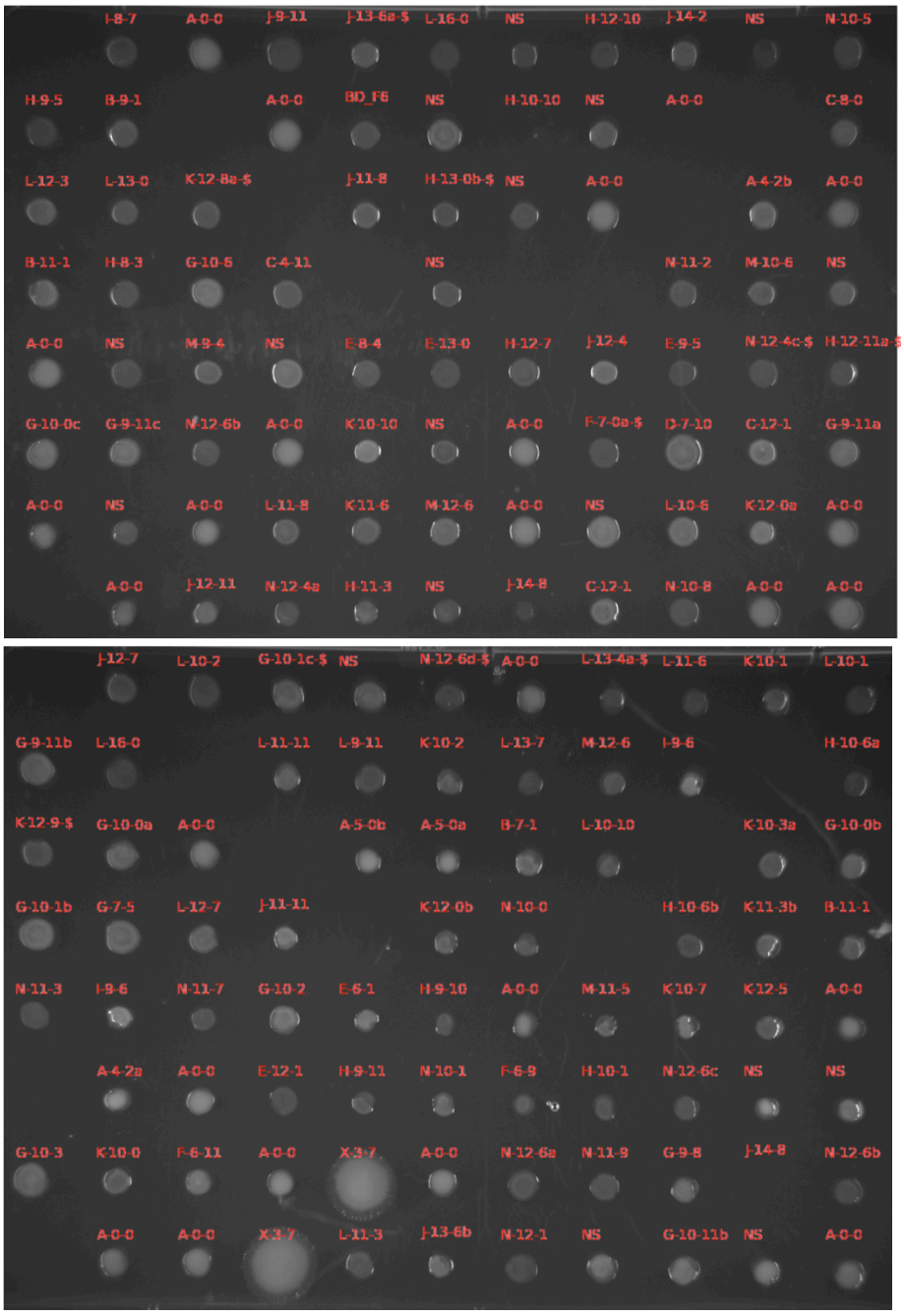
Supplementary Figure 6. O-antigen repeats as a function of BDAG_02317. a, Complementation with a full-length BDAG_02317 restores O-antigen presentation. Full length glycosyltransferase BDAG_02317 was amplified from AU0158 (reference strain) and inserted into a plasmid under the control of a constitutive promoter (**Supplementary Note**). Transformation of strain A-0-0 with this plasmid restored O-antigen presentation. We note variation in O-antigen chain molecular weight, as we had also observed amongst our isolates (as seen in **Figure 3b**). **b, Most last common ancestors display a truncated glycosyltransferase.** We display the phylogeny of the 112 strains. Branches are colored as a function of the inferred genotype in the gene BDAG_02137 at codon 275 (black: stop codon, pink: glutamine , blue: glutamate, gray: unknown). The last common ancestors of strains from each patient are shown as dots; they are colored with their inferred BDAG_02137 genotype. Even though mutations occurred in 9 patients, all the last common ancestors but one (patient D, for whom we sequenced a single isolate) bear the stop codon. We hypothesize that the truncated genotype may provide an advantage during patient-to-patient transmission.

Supplementary Figure 7



Supplementary Figure 7: Most genes under positive selection are conserved across *Burkholderia*. For each gene under positive selection, blastp was used to find the nearest homologs in each of the 20 *Burkholderia* genomic sequences (9 of which are also members of the *Burkholderia cepacia* complex which colonize individuals with CF, highlighted in red). For each comparison, a box is drawn if a match was found with an E value of less than 10^{-8}, and that box is shaded in according to the percent amino-acid identity between the two sequences (see key). *B. mallei and B. pseudomallei* are important pathogens infecting healthy individuals.

Supplementary Figure 8



Supplementary Figure 8. Isolates from the outbreak do not show significant variation in mucoidy. A fresh library of isolates was obtained by pinning from the thawed frozen stock into LB and growing overnight at 37°C. A 0.3 microliter aliquot of each overnight culture was pinned onto omnitrays containing YEM agar (0.5g Yeast Extract, 4 g mannitol, and 15g agar per liter) and grown at 37°C. After one day of growth, the plates were imaged. The outgroup strain, X-3-7, appears mucoid (right panel, two replicates). All strains from the outbreak are significantly less mucoid. Strains not sequenced for this study are indicated with “NS.” Not all sequenced strains are grown because some failed to grow on YEM agar

Supplementary Table 1:
Burkholderia dolosa isolates used in this study

Study name	Sample origin	SRA Accession
Patient A		
A-0-0	Airway	SRS242477
A-4-2a	Airway	SRS242478
A-4-2b	Airway	SRS242497
A-5-0a	Airway	SRS242395
A-5-0b	Airway	SRS242479
Patient B		
B-4-11	Airway	SRS242400
B-7-1	Airway	SRS242414
B-9-1	Airway	SRS242401
B-11-1	Airway	SRS242412
Patient C		
C-4-11	Airway	SRS242415
C-8-0	Airway	SRS242484
C-10-0	Airway	SRS242416
C-12-1	Airway	SRS242485
Patient D		
D-7-10	Airway	SRS242441
Patient E		
E-6-1	Airway	SRS242402
E-8-4	Airway	SRS242417
E-9-5	Airway	SRS242428
E-12-1	Airway	SRS242430
E-13-0	Airway	SRS242480
Patient F		
F-6-9	Airway	SRS242481
F-6-11	Airway	SRS242482
F-7-0a-$	Blood	SRS242483
Patient G		
G-7-5	Airway	SRS242403
G-9-8	Airway	SRS242418
G-9-11a	Airway	SRS242419
G-9-11b	Airway	SRS242487
G-9-11c	Airway	SRS242420
G-10-0a	Airway	SRS242421
G-10-0b	Airway	SRS242422
G-10-0c	Airway	SRS242423
G-10-1b	Airway	SRS242424
G-10-2	Airway	SRS242488
G-10-1c-$	Blood	SRS242426
G-10-3	Airway	SRS242425
G-10-6	Airway	SRS242427
G-10-11b	Airway	SRS242404
Patient H		
H-8-3	Airway	SRS242386
H-9-5	Airway	SRS242410
H-9-10	Airway	SRS242493
H-9-11	Airway	SRS242442
H-10-1	Airway	SRS242443
H-10-6a	Airway	SRS242385
H-10-6b	Airway	SRS242444
H-10-10	Airway	SRS242445
H-11-3	Airway	SRS242446
H-12-7	Airway	SRS242447
H-12-10	Airway	SRS242472
H-12-11a-$	Blood	SRS242492
H-13-0b-$	Blood	SRS242471
Patient I		
I-8-7	Airway	SRS242411
I-9-6	Airway	SRS242413

Study name	Sample origin	SRA Accession
Patient J		
J-9-11	Airway	SRS242466
J-11-8	Airway	SRS242439
J-11-11	Airway	SRS242406
J-12-4	Airway	SRS242440
J-12-7	Airway	SRS242407
J-13-6a-$	Blood	SRS242468
J-13-6b	Airway	SRS242467
J-12-11	Airway	SRS242469
J-14-8	Airway	SRS242408
J-14-2	Airway	SRS242470
Patient K		
K-9-0	Airway	SRS242463
K-10-0	Airway	SRS242489
K-10-1	Airway	SRS242432
K-10-2	Airway	SRS242433
K-10-3a	Airway	SRS242434
K-10-7	Airway	SRS242435
K-10-10	Airway	SRS242464
K-11-3a	Airway	SRS242431
K-11-3b	Airway	SRS242436
K-11-6	Airway	SRS242437
K-12-0a	Airway	SRS242405
K-12-0b	Airway	SRS242438
K-12-5	Airway	SRS242465
K-12-8a-$	Blood	SRS242490
K-12-9-$	Blood	SRS242491
Patient L		
L-9-11	Airway	SRS242473
L-10-1	Airway	SRS242448
L-10-2	Airway	SRS242449
L-10-6	Airway	SRS242409
L-10-10	Airway	SRS242387
L-11-3	Airway	SRS242388
L-11-6	Airway	SRS242389
L-11-8	Airway	SRS242390
L-11-11	Airway	SRS242391
L-12-3	Airway	SRS242392
L-12-7	Airway	SRS242393
L-13-0	Airway	SRS242394
L-13-4a-$	Blood	SRS242474
L-13-7	Airway	SRS242475
L-16-0	Airway	SRS242476
Patient M		
M-9-4	Airway	SRS242396
M-10-6	Airway	SRS242397
M-11-5	Airway	SRS242398
M-12-6	Airway	SRS242399
Patient N		
N-10-0	Airway	SRS242494
N-10-1	Airway	SRS242450
N-10-5	Airway	SRS242451
N-10-8	Airway	SRS242452
N-10-11	Airway	SRS242453
N-11-2	Airway	SRS242454
N-11-3	Airway	SRS242455
N-11-7	Airway	SRS242456
N-11-9	Airway	SRS242457
N-12-1	Airway	SRS242458
N-12-4a	Airway	SRS242459
N-12-4c-$	Blood	SRS242496
N-12-5-$	Blood	SRS242429
N-12-6a	Fluid, Right Inferior Hematoma	SRS242460
N-12-6b	Pleural fluid	SRS242461
N-12-6c	Tissue, Right Lung	SRS242462
N-12-6d-$	Blood	SRS242495
Patient X (outgroup)		
X-3-7	Airway	SRS242486

Study names designate patient and time. For example, 'A-2-4a' was recovered 4 years and 2 months after the date of the first isolate in our study, and it was the first isolate recovered from Patient A during this month. The second column designates origin of this isolate (blood isolates are also indicated in the study name by '$'. The third column indicates the Sequence Read Archive (at The National Center for Biotechnology Information, USA) sample number for the publicly available primary sequencing data

Supplementary Table 4:
Full annotation and homology for genes under parallel, positive selection used to assess biological relevance.

Gene ID	Broad annotation	Number of mutations	Biological relevance	Annotated homologs: organism (query coverage)	Additional annotation
Genes not previously implicated in pathogenesis					
BDAG_01161	PAS	17	Oxygen-related gene regulation	*fixL*: *B. rhizoxinica* (98)	Two-component system histidine kinase
BDAG_01160	regulatory protein LuxR	4	Oxygen-related gene regulation	*fixJ*: *B. rhizoxinica* (99)	Two-component system response regulator
BDAG_04180	transcriptional regulator Crp/Fnr family	3	Oxygen-related gene regulation	*anr*: *B. multivorans* (99), *fixK*: *C. crescentus* (69)	
BDAG_02650	SAM-dependent methyltransferase	8	?	ZP_02908193.1: *B. ambifaria* (99)	Methyltransferase
BDAG_04154	DNA-directed RNA polymerase specialized sigma subunit	4	?	YP_001116246.1: *B. vietnamiensis* (86)	Gene regulation, ECF subfamily
BDAG_00993	Glucoamylase	3	?	ZP_02893540.1: *B. ambifaria* (99)	Glycoside hydrolase
Genes previously implicated in pathogenesis					
BDAG_02781	Ribosomal protein L4	14	Antibiotic resistance	*rplD*	Macrolide resistance
BDAG_02180	DNA gyrase subunit A	11	Antibiotic resistance	*gyrA*	Quinolone resistance (see **Figure 3A**)
BDAG_03694	Beta-lactamase	8	Antibiotic resistance	YP_001811541.1: *B. ambifaria* (99), *bla1*: *B. cereus* (88)	Beta-lactam resistance
BDAG_04506	ABC-type multidrug transport system ATPase and permease components	5	Antibiotic resistance	*mdlB*: *M. succiniciproducens* (95)	Hypothetical role in AB resistance
BDAG_00220	Type IIA topoisomerase	3	Antibiotic resistance	*gyrB*	Quinolone resistance
BDAG_02317	Glycosyltransferase	10	Membrane Synthesis	*wbaD*: *E. coli* (88)	(see **Figure 3B**)
BDAG_02321	Glycosyltransferase	6	Membrane Synthesis	*mtfA*: *A. fulgidus* (99)	Mannitol transferase
BDAG_01250	Major facilitator superfamily (MFS_1) transporter	3	Membrane Synthesis	*lplT*: *B. cenocepacia* (99)	Lisophospholipid transporter
BDAG_00856	hypothetical protein	3	Membrane Synthesis	YP_002231781.1: *B. cenocepacia* (99), *lptG*: *E. clocae* (93)	Permease YjgP/YjgQ family
BDAG_03003	Large exoprotein involved in heme utilization or adhesion	4	Secretion	ZP_02889734: *B. ambifaria* (97), *fhaB*: *P. ananatis* (80)	Secreted product
BDAG_00179	Type II secretory pathway component PulD	3	Secretion	*gspD*: *B. cenocepacia* (99)	Secretion system component

Broad annotations are automatically assigned in the reference genome. Most relevant known homologs, additional annotation, and biological relevance were determined by manual annotation.

Supplementary Table 5: Primers used in this study

Name	**Primers 5'->3'**
GlycTcompF	TTTTTGAGCTCATACGACGTCGATGCCGGAGATCG
GlycTcompR	TTTTTTCTAGACCGTCGCTCCGGAGTCTCAACC
pUCP18up	GGCTCGTATGTTGTGTGGAATTGTGAGCGG

Supplementary Note

Burkholderia dolosa

Burkholderia dolosa is a member of the *Burkholderia cepacia* complex*,* a group of related species capable of infecting immune-compromised hosts. Members of the *B. cepacia* complex were initially categorized as *Pseudomonas cepacia.* Members of this complex are found frequently in the soil, and are associated with the roots of plants (they were first identified as the cause of onion rot in 1950).

However, in the 1980s they were recognized as an emerging cause of infection in patients with cystic fibrosis[1]. Some members of the *B. cepacia* complex*,* including *B. dolosa,* cause deadly 'cepacia syndrome' – a rapid health decline characterized by fevers, necrotizing pneumonia and bacteremia (presence of bacteria in the bloodstream)[2].

B. dolosa is one of the rarest members of the *B. cepacia* complex. It was first established as a distinct species in 2004[3].

Patient cohort

We retrospectively studied a small epidemic of *Burkholderia dolosa* that affected 39 people with cystic fibrosis (CF) who were treated at a Boston hospital. The average age of these patients at date of first *B. dolosa* detection was 19 years.

Treatment of these patients for *B. dolosa* and other bacterial infections of CF patients includes frequent administration of antibiotics, usually in combination[4]*.* While antibiotics are often helpful in quelling an acute flare-up, they are generally ineffective in eradicating the infection. Of the 34 patients who continued care at the Boston hospital, only 3 eventually cleared the infection. Some of the patients still colonized remain asymptomatic, while others developed cepacia syndrome[5]. At the time of submission, 9 patients had received lung transplants; 17 patients had died, all but one from their illness.

We chose to study 14-deidentified patients, including patient zero of the epidemic. Seven patients were included in this study on the basis of availability of isolates from the blood; the remaining six were chosen at random among patients who did not have bacteremia. At the time of writing, 5 of these patients had received lung transplants; 8 patients had died, all but one from their illness. None of these patients cleared their airways of *Burkholderia dolosa*.

As mentioned above, these patients receive frequent antibiotic therapy. However, the patients' records do not allow us to know their history of antibiotic usage with high enough accuracy to correlate this data with antibiotic-resistance or data related to antibiotic-resistance.

Bacterial isolate collection

Samples from these patients were collected during normal care. The 97 isolates labeled "airways" were obtained from either sputum or bronchoalveolar lavage (BAL) fluid. Another 11 isolates labeled "blood" were sampled from the blood cultures. The remaining 4 isolates labeled "other" were sampled from pleural tissue pleural or mediastinal tissue from fluid obtained during surgical procedures. Details are in

Supplementary Table 1. In most cases, the frozen clinical stocks were prepared from a single colony. The timing between isolates is described in **Supplementary Fig. 1**.

Frozen clinical stocks were streaked on solid media. A single colony from each plate was chosen at random and frozen in 15% glycerol to create a working library.

We also obtained the isolate LMG18943 (also known as AU0645), shown elsewhere to be of the same strain[6] (Strain SLC6). This isolate was taken from a different CF patient in a different location in the USA during the duration of the epidemic. We used this isolate as the outgroup for phylogenetic analyses.

Throughout this work, time of isolation is reported relative to the collection of an isolate from patient zero (isolate A-0-0). The time of collection of this isolate is unrelated to the first detection of *B. dolosa* in this patient. Letters distinguish isolates collected on the same month from the same patient. In some cases, the lettering is not continuous; some sequenced isolates were omitted from the study because they were either duplicates or recovered during autopsy.

Illumina sequencing

DNA was extracted from single colonies using standard procedures, following instructions from MoBio UltraClean® Microbial DNA Isolation Kit (cat # 12224-50, MO BIO Laboratories, Inc., USA).

Genomic libraries were constructed using the Illumina-compatible Nextera™ DNA Sample Prep Kit (cat# GA091120, EPICENTRE Biotechnologies, USA). Each genomic library was made up of four to six multiplexed genomes, barcoded with Nextera™ adapters. The libraries were sequenced using single-end, 75 bp reads on Illumina Genome Analyzer GAII by Partners HealthCare Center for Personalized Genetic Medicine (PCPGM). On average, 4.6 million reads were obtained per isolate (**Supplementary Fig. 2a**).

Reads were aligned using the *B. dolosa* genome AU0158 as a reference[7]; this reference genome belongs to an isolate recovered from patient zero. The reference genome is 6.42 megabases (Mb) long and comprises 3 chromosomes of lengths 3.41 Mb, 2.18 Mb, and .83 Mb. Alignment was performed by PCPGM using the standard Illumina pipeline[8] (CASAVA v1.6). We used SAMtools 0.1.12a to manipulate consensus sequences and find SNP between isolates[9].

For each strain we calculated the average number of reads aligned to each position in the reference genome (**Supplementary Fig. 2b**). We found the average read depth to be 37x. These reads produced high coverage: on average we obtained confident calls for 93% of the genome (Phred score>25).

Polymorphic loci between the 112 isolates

We first obtained the list of 4,375 loci where at least one of the 113 isolates (including the outgroup) was different than the reference genome. Next, we filtered out data to obtain a high-quality list of 511 polymorphic loci, where we expect less than one false positive.

Specifically, we used a two-step process. First, we retained loci where polymorphism was observed between at least two strains called with high confidence scores (Phred score >35). Then, at each of these 511 loci, we included all calls that were made with high enough confidence (Phred score >25; because fewer calls were made compared with the previous list, the Phred score could be lowered).

With this procedure, we obtain 56,811 calls at these 511 polymorphic loci. This two-step process and choice of scores maximizes the number of isolates with information per location, under the constraint of obtaining at most one miscalled base among all 56,811 calls. On average, 111 out of 113 isolates were called at these loci.

19 of these loci are unique to the outgroup; we resolve 492 polymorphic loci within the isolates of this epidemic.

Intragenic bias

We calculated that 426 of the 561 mutations occurred in intragenic regions (using the 5012 genes automatically annotated on the reference genome): the proportion of intragenic mutations is 75% (95% CI=72%-79%, Clopper-Pearson binomial confidence). This estimate is not statistically distinct from the intragenic proportion of the B. dolosa genome (79%): we do not detect a significant intragenic bias (p=0.04, binomial sampling with n=561).

Strains used for O-antigen assay

The strains used in the O-antigen assay presented in **Figure 3b** were (from left to right): A-0-0, AU0158, M-9-4, G-9-8, G-9-11b, C-12-1, B-11-1, E-9-5, E-12-1, C-10-0, K-10-0, K-12-5, K-12-8a-$, F-6-9, I-9-6, N-12-1, J-13-6a-$, J-14-2, J-14-8, and J-12-4.

Glycosyltransferase complementation

Escherichia coli Sm10[10] and *B. dolosa* were routinely grown on Luria agar (LA) containing 1.5% Bacto Agar (Difco, Franklin Lakes, N.J.). The medium was supplemented with 10 µg of tetracycline per ml for *E. coli* and 75 µg of tetracycline and 100µg of ampicillin per ml for *B. dolosa*, as appropriate. The glycosyltransferase gene BDAG_02317 was amplified by PCR with primers GlycTcompF and GlycTcompR (see **Supplementary Table 5**) and genomic DNA of B. dolosa AU0158. This gene was ligated into the broad-host-range vector pUCP18Tc creating expression plasmid pUCP18Tc-GlycT. Biparental matings were performed to transfer pUCP18Tc-GlycT from E. coli SM10 to *B. dolosa* as previously described[11]. Transconjugants were confirmed by PCR using a forward primer specific of the complementing plasmid (pUCP18up) and primer GlycTcompR. O-antigen presentation was assayed as described in Online Methods.

fix genes

Burkholderia dolosa lacks the genes necessary for nitrogen fixation[1] suggesting that the *fix* system is utilized for gene-regulation of a different nature, as has been seen in other species[12]. All of the 24 mutations seen in this pathway are nonsynonymous and none result in a stop codon, suggesting a tuning of the pathway rather than its interruption. For this reason, we annotate these genes as involved in oxygen-related gene regulation.

Supplementary References

1. Coenye, T., Vandamme, P., *Burkholderia: molecular microbiology and genomics*. Horizon Scientific Press (2007).

2. Kalish, L. A. *et al.* Impact of *Burkholderia dolosa* on lung function and survival in cystic fibrosis. *Am J Respir Crit Care Med* **173**, 421-425 (2006).

3. Vermis, K. *et al.* Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. *Int J Syst Evol Microbiol* **54**, 689-691 (2004).

4. Lyczak, J.B., C.L. Cannon, and G.B. Pier, Lung infections associated with cystic fibrosis. *Clinical microbiology reviews* **15**, 194-222 (2002).

5. Lipuma, J. J. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev* **23**, 299-323 (2010).

6. Biddick, R., Spilker, T., Martin, A. & LiPuma, J. J. Evidence of transmission of *Burkholderia cepacia*, *Burkholderia multivorans* and *Burkholderia dolosa* among persons with cystic fibrosis. *FEMS Microbiol Lett* **228**, 57-62 (2003).

7. *Burkholderia dolosa* Sequencing Project. Broad Institute of Harvard and MIT (http://www.broadinstitute.org/).

8. Illumina, I., San Diego, CA., *Complete Secondary Analysis Workflow for the Genome Analyzer,* Pub. No. 770-2009-033 Current as of 19 October 2009.

9. Li, H., et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-9 (2009).

10. Simon, R., P. U., and A. Puhler. A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Biotechnology* **1,** 784–91 (1983).

11. Urban, T. A. *et al.* Contribution of *Burkholderia cenocepacia* flagella to Infectivity and Inflammation. *Infection and Immunity* **72,** 5126-34 (2004).

12. Crosson, S., McGrath, P. T., Stephens, C., McAdams, H. H. & Shapiro, L. Conserved modular design of an oxygen sensory/signaling network with species- specific output. *Proc Natl Acad Sci U S A* **102**, 8018-8023 (2005).