PLOS PATHOGENS

Pathogen to commensal? Longitudinal within-host population dynamics, evolution, and adaptation during a chronic >16-year *Burkholderia pseudomallei* infection

Talima Pearson[1], **Jason W. Sahl**[1], **Crystal M. Hepp**[1,2], **Karthik Handady**[1], **Heidie Hornstra**[1], **Adam J. Vazquez**[1], **Erik Settles**[1], **Mark Mayo**[3], **Mirjam Kaestli**[3], **Charles H. D. Williamson**[1], **Erin P. Price**[1¤a], **Derek S. Sarovich**[1¤a], **James M. Cook**[1], **Spenser R. Wolken**[1], **Richard A. Bowen**[4], **Apichai Tuanyok**[1¤b], **Jeffrey T. Foster**[1], **Kevin P. Drees**[1¤c], **Timothy J. Kidd**[5], **Scott C. Bell**[6], **Bart J. Currie**[3,7], **Paul Keim**[1]*
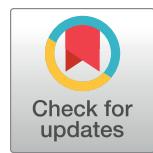
1 Pathogen & Microbiome Institute, Northern Arizona University, Flagstaff, Arizona, United States of America, **2** School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, United States of America, **3** Global and Tropical Health Division, Menzies School of Health Research, Darwin, Northern Territory, Australia, **4** Department of Biomedical Sciences, Colorado State University, Colorado, United States of America, **5** School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia, **6** Department of Thoracic Medicine, The Prince Charles Hospital, Chermside, and QIMR Berghofer Medical Research Institute, Queensland, Australia, **7** Infectious Diseases Department and Northern Territory Medical Program, Royal Darwin Hospital, Darwin, Northern Territory, Australia

¤a Current address: GeneCology Research Centre and Sunshine Coast Health Institute, University of the Sunshine Coast, Queensland, Australia
¤b Current address: Department of Infectious Diseases & Immunology, Emerging Pathogens Institute, University of Florida, Florida, United States of America
¤c Current address: Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire, United States of America
* Paul.Keim@nau.edu

Abstract

Although acute melioidosis is the most common outcome of *Burkholderia pseudomallei* infection, we have documented a case, P314, where disease severity lessened with time, and the pathogen evolved towards a commensal relationship with the host. In the current study, we used whole-genome sequencing to monitor this long-term symbiotic relationship to better understand *B. pseudomallei* persistence in P314's sputum despite intensive initial therapeutic regimens. We collected and sequenced 118 *B. pseudomallei* isolates from P314's airways over a >16-year period, and also sampled the patient's home environment, recovering six closely related *B. pseudomallei* isolates from the household water system. Using comparative genomics, we identified 126 SNPs in the core genome of the 124 isolates or 162 SNPs/indels when the accessory genome was included. The core SNPs were used to construct a phylogenetic tree, which demonstrated a close relationship between environmental and clinical isolates and detailed within-host evolutionary patterns. The phylogeny had little homoplasy, consistent with a strictly clonal mode of genetic inheritance. Repeated sampling revealed evidence of genetic diversification, but frequent extinctions left only one successful lineage through the first four years and two lineages after that. Overall,

the evolution of this population is nonadaptive and best explained by genetic drift. However, some genetic and phenotypic changes are consistent with *in situ* adaptation. Using a mouse model, P314 isolates caused greatly reduced morbidity and mortality compared to the environmental isolates. Additionally, potentially adaptive phenotypes emerged and included differences in the O-antigen, capsular polysaccharide, motility, and colony morphology. The >13-year co-existence of two long-lived lineages presents interesting hypotheses that can be tested in future studies to provide additional insights into selective pressures, niche differentiation, and microbial adaptation. This unusual melioidosis case presents a rare example of the evolutionary progression towards commensalism by a highly virulent pathogen within a single human host.

Author summary

Pathogens frequently jump between different hosts, and associated adaptation may lead to the emergence of new infectious agents. Such host-jumping evolution is witnessed through endpoint analyses but these cannot capture genetic changes in lineages that have gone extinct. In this study, we have identified and monitored an example of the evolution of a bacterium often deadly to its mammalian host, in an unprecedented case whereby disease lessened through time and the pathogen became a part of the commensal human flora. We used genomic analyses to characterize more than 16 years of this evolutionary process and the stepwise mutations that control pathogen interactions with the patient. Soon after infection, mutational changes occurred that allowed the bacterium to remain in the airways without causing disease. This shift towards avirulence was determined based on clinical data and virulence testing in an animal model. In addition, mutations occurred that contributed to the persistence of the bacteria in the patient's lungs. Finally, we found evidence for the evolutionary emergence and persistence of two distinct lineages of the bacterium over the last 13 years, presenting interesting questions about niche utilization. Bacteria are ubiquitous in the human body and almost all are beneficial or benign. In this study, we document the evolutionary conversion of a normally deadly bacterium towards a commensal.

Introduction

Within-host evolution of bacterial populations over the course of chronic infections allows the rare opportunity to witness the emergence of bacterial fitness strategies in a new environment. These tactics to evade and overcome host immune responses in order to persist in the host environment may include increased replication efficiency, slower growth, attenuation of virulence, increased transmissibility, and the development of antimicrobial resistance [1]. Genetic diversity within the pathogen population is critical for adaptive evolution as it provides the genetic foundation upon which selection can act. Diversity in an infection may be due to heterogeneity in the inoculating population, lateral gene transfer, or may develop through intrinsic mutations over the course of the infection. The short-term fate of mutations depends on a range of factors such as the population size, environmental stability [2], and the fitness effects of these and other linked mutations. The long-term evolutionary persistence of mutations that emerge within a host depends both on host longevity and pathogen transmission to additional

hosts. In most cases, little is known about the emergence of ephemeral mutations that occur during the process of long-term carriage within a host. Occasionally, this evolutionary process leads to the emergence of novel species, but only successful lineages are studied and processes are retrospectively inferred. The low cost and high throughput of current sequencing technologies makes characterizing many isolates from single infections possible, enabling a more representative characterization of the population. Phylogenetic analyses of an infecting population and comparison to a close outgroup is necessary to determine the directionality of evolution and to provide a detailed understanding of the population dynamics, evolution, and possible adaptation.

Melioidosis is an infectious disease that primarily occurs in the tropics, particularly Southeast Asia and northern Australia, where it is hyperendemic [3–5]. Human infections with the melioidosis bacterium, *Burkholderia pseudomallei*, are not usually communicable, with most cases acquired via percutaneous inoculation, inhalation, or ingestion of contaminated soil and water [6]. The incubation period for acute disease following an infecting event is ~9 days, but can be as short as 24 hours with high-level exposure, such as aspiration following near-drowning [7]. Approximately 9% of cases present as a chronic infection, defined as symptoms being present for >2 months [8]. Melioidosis presentations are extremely diverse, ranging from self-resolving skin ulcers without evidence of sepsis to rapidly progressive septic shock with multiple internal organ abscesses [8]. Pneumonia is present in half of cases, overall bacteremia rates are 60%, and septic shock occurs in 21% of all cases [8]. Treatment involves intensive intravenous antibiotic therapy with ceftazidime, meropenem, or imipenem for at least 10 days, followed by eradication therapy with oral trimethoprim-sulfamethoxazole for 3–6 months [7]. Despite this regimen, recurrence occurs in up to 6% of cases through either relapse or reinfection.

Bacterial airway infections are serious illnesses that can progress to chronic infection if not eradicated in a timely manner. Acute and chronic infections can be highly morbid. A notable example is the cystic fibrosis (CF) airways, which contain bacterial populations that frequently adapt to evade host and treatment onslaught, enabling pathogens to undergo longitudinal genetic changes *in vivo* that facilitate persistence and drive respiratory decline [9–11]. Although melioidosis is rare in CF cases, ~76% of the documented chronic cases are in CF patients [12] compared with just 10% in non-CF cases. *B. pseudomallei* can undergo dramatic genomic and transcriptomic changes that may enable persistence in the airways of CF patients despite intensive antibiotic treatment regimens [13, 14], demonstrating that this bacterium is adept at long-term airway persistence under certain conditions. In contrast, acute pneumonia can be fulminant with subsequent sepsis. Melioidosis is notable for having both scenarios in the clinical spectrum [15]. The progression of inhalation/pulmonary melioidosis can be very rapid and severe. In animal models, acute melioidosis occurs within 1–2 days of aerosol exposure to *B. pseudomallei* [16].

Similar to many other saprophytic soil dwelling bacteria, *B. pseudomallei* hosts a large and highly plastic genome [17]. The ~7Mbp genome of wild-type *B. pseudomallei* isolates is carried on two chromosomes. However, the size of the pangenome of multiple conspecific strains is truly impressive. The accumulation of coding region sequences in the pan-genome [18] is largely associated with horizontal acquisition of sequences, including genomic islands from genetic near neighbors [19]. Despite frequent gene additions and selection for maintaining new genes, a bias towards deletions stabilizes the genome size of free-living bacteria [20–22]. The virulence of *B. pseudomallei* has been attributed to multiple genomic loci, including type III and type VI secretion systems, lipopolysaccharides, actin-based motility factors, toxins, adhesins, and capsular polysaccharide clusters [17, 23]. The large accessory genome likely contributes to the observed clinical diversity [24]. Also, the large effective population sizes typical

of free-living prokaryotes, results in efficient selection of characteristics that enable adaptation and persistence in dynamic and hostile environments.

In the Darwin Prospective Melioidosis Study (DPMS), 9% of the 1,105 culture-confirmed cases over the last 29 years presented as a chronic infection [8]. The majority of patients in these cases have survived, and all but one eventually cleared their infection. This individual, Patient 314 (P314), remains sputum culture-positive for *B. pseudomallei* since being diagnosed with melioidosis in 2000 as an in-patient following a bronchiectasis exacerbation event [25]. We have collected and cultured *B. pseudomallei* isolates from P314's airways since this initial diagnosis to better understand the population dynamics in this unprecedented case. We have previously shown genetic diversity in variable-number tandem repeat loci among these isolates [26] and subsequently compared three P314-derived genomes to document some of the evolutionary changes over an 11.5-year period [25]. Here, we analyze and compare the genomes and phenotypic characteristics of 118 *B. pseudomallei* isolates sampled over a >16-year chronic carriage period from P314's airways, along with six closely-related isolates sampled from P314's home between two and four years after diagnosis. These comparisons allow assessment of the evolutionary and temporal dynamics of this extraordinary case rather than strictly comparing evolutionary end-points, providing broader inferences on the evolution of host specificity.

Materials and methods

Clinical sample collection, culturing, and isolation

We collected and analyzed a 17-year longitudinal series of *B. pseudomallei* isolates for this study. The Microbiology Departments of the Royal Darwin Hospital, Darwin and The Prince Charles Hospital, Brisbane, and the Melioidosis laboratory at the Menzies School of Health Research, Darwin, performed bacterial isolations. *B. pseudomallei* was cultured from sputum, nasal and throat swabs, nasal mucus, or extracted lung tissue following a lobectomy in 2003 by using Ashdown's selective broth and Ashdown's selective medium [27], except for the initial clinical isolate MSHR1043, which was isolated using chocolate agar as it is sensitive to gentamicin [25]. *B. pseudomallei* was confirmed biochemically using the API 20E system (API System SA, Lyon, France) or, more recently, MicroScan WalkAway (Dade Diagnostics, West Sacramento, Calif.) and the TTS1 *B. pseudomallei* PCR assay [28].

Genome sequencing, assembly, and data deposition

Genomes were subjected to whole-genome sequencing (WGS) on the Illumina platform, followed by assembly with SPAdes v3.10.0 [29]; contigs that aligned to non-target species (e.g. PhiX) or that contained anomalously low coverage compared to the average genome coverage were manually removed. The environmental isolate MSHR1435 was sequenced on both the Illumina and PacBio platforms and assembled as described elsewhere [30]. This closed genome was previously deposited into GenBank under accessions CP025264.1 and CP025265.1. The raw read data for each strain were submitted to the Sequence Read Archive database (S1 Table) under BioProject accession PRJNA321854.

SNP calling

Short reads from all genomes were mapped across the complete genome of the environmental isolate MSHR1435 using BWA-MEM v0.7.7 [31], and SNPs were called with the UnifiedGenotyper method in GATK v3.4.6 GATK [32, 33]. SNPs were filtered from downstream analyses if they fell within duplicated regions in the reference genome, as determined by a NUCmer

self-alignment [34] or, if they fell within short tandem repeats, as determined by Tandem Repeats Finder v4.0.9 [35]. If SNPs had a coverage depth of <3 or an allele proportion <90%, they were manually investigated. All of these functions were wrapped within the NASP pipeline [36]. Corresponding SNPs between different reference genomes (MSHR1435, MSHR1043, K96243) were identified with the ISG pipeline [37]. Functional information was applied to each SNP with SnpEff v4.3r [38] using the PGAP annotation.

Phylogenetic reconstruction

The phylogenetic tree was constructed in two steps. First, 126 orthologous, core-genome SNPs (S2 Table) present in each isolate (except 2 missing SNP loci in the outgroup K96243, which did not affect the topology) were used to construct a maximum parsimony tree using a heuristic search with default options in PAUP 4.0b10 [39] to give an accurate estimation of phylogenetic grouping, albeit with limited resolution. This tree was one of four equally parsimonious trees. Two homoplastic SNPs among clinical strains caused a consistency index (CI) (excluding uninformative characters) of 0.9737 and a tree length of 128. The depicted tree is the single tree whose topology agrees with a tree inferred without the homoplastic SNPs. Secondly, other non-core characters (36 SNPs and 15 deletions) were added to the SNP matrix for added phylogenetic resolution to better capture all the character changes that occurred along all branches of the tree. To ensure independence of evolutionary events, each inferred deletion event was manually added as a single binary character to the data matrix. Similarly, for SNPs present in regions deleted in other isolates, the inferred immediate ancestral state for each indel was determined (from the phylogenetic placement of the sample in the first tree) and added to the matrix. For SNP calls that were missing in some genomes, we manually determined if the missing call was due to a deletion or poor sequencing coverage. For loci in low coverage areas, we assumed the immediate ancestral allele as a conservative estimation of the evolutionary state. The unambiguous optimization of these character states eliminates potential phylogenetic ambiguity due to missing data in tree reconstruction [40] and treats each deletion event as a single event rather than a separate event associated for every SNP within that region. This method is similar to uninode coding [41]. Sequences for samples MSHR1360 and MSHR1459 are from a population of cells with two discrete genotypes, which were not parsed during the subculturing step. For some SNPs in these sequences, both alleles are detectable and the deleted regions are present in a minority of reads, although each allele had a minimum coverage of 3x. Based on the presence of alleles, we could determine the phylogenetic position of both components and therefore include them in the final tree as they contribute to our understanding of the population dynamics. Other sequences also show evidence of fine-scale mixed genotypes, however minor alleles occur in very few reads and we could not confidently parse the genotypes.

Comparative genomics

To identify deletions, raw data were aligned against the genome of the environmental isolate MSHR1435 with BWA-MEM and the per base depth of coverage (DOC) was identified with SAMtools v0.1.19 [42]. Due to low levels of genome mixtures observed across many of the genomes, a dynamic filtering approach was employed. First, regions with <25% DOC compared with the average DOC for each genome was identified by summing the coverage across each base of the reference genome and then dividing by the total genome length. For each base, if the depth was less than 25% of the average DOC, then the position was considered absent. Visual confirmation of deletions was performed using Circos [43]. To further verify deletions, each coding region sequence (CDS) in MSHR1435 was screened against all genome

assemblies with the large-scale BLAST score ratio (LS-BSR) pipeline [44] using the BLAT aligner [45]. CDSs that fell within deletions identified with the read mapping were then tabulated (S3 Table). LS-BSR was also used to find CDSs that were part of the core or the accessory genome. For one region that was present in environmental samples but had signatures of degradation in clinical samples, as identified with LS-BSR, raw data were re-aligned to MSHR1435 using Bowtie2, and SAMtools was used to identify per base DOC. This approach allowed for the determination of deletion boundaries.

Molecular clock analysis

To estimate the rate of intra-patient evolution and time to most recent common ancestor for clinically relevant events, we performed a Bayesian molecular clock analysis in Bayesian Evolutionary Analysis Sampling Trees (BEAST) v1.8.4 [46] using the 126 core-genome, orthologous SNP dataset. Substitution model selection was carried out in MEGA 7.0.9 [47] for the 120 clinical genomes. The corrected Akaike Information Criterion and Bayesian Criterion results indicated that the HKY substitution model [48] best fit this dataset. To determine the best fitting clock and demographic model combinations for these data, we employed the generalized stepping-stone sampling marginal likelihood estimator [49]. Each of the eight model combinations was iterated 100,000,000 times, where Markov chains were sampled every 10,000 generations. We found that the strict clock combined with the Bayesian Skygrid demographic prior outperformed the other seven combinations (S4 Table). Using the Strict Skygrid model combination, we ran three additional chains for 100,000,000 generations, sampling every 10,000, and found convergence within and among chains using Tracer v1.6 [50]. Importantly, we included a correction for SNP ascertainment bias by indicating the number of invariant sites (A: 985225, C: 2103022, G: 2112938, T: 982967), allowing for a more accurate estimation of rates and timing. We used LogCombiner [46] to merge the four different chains, discarding the first 10% as burn in (40,000,000 generations), and then resampled every 40,000 generations. The resulting file was input to TreeAnnotator to produce a maximum clade credibility tree, and then visualized using FigTree v1.4.3 [51].

Selection analyses

We analyzed the protein coding regions of 74 concatenated genes from the clinical genomes, containing 88 SNPs, for evidence of selection using SLAC via HyPHY [52] and the codon-based test of neutrality described by Nei and Gojobori [53]. No intergenic or nonsense mutations were included in these analyses. We also used all 177 genetic characters in the dataset (SNPs and deletions) to test for departure from a neutral model of evolution using Tajima's Neutrality Test -D [54].

Phenotypic testing

We previously noted considerable differences in growth rate and colony morphology between P314's initial isolate and a 139-month isolate [25]. To determine if other isolates over the course of the infection also had observable changes in growth characteristics, 110 P314 isolates were streaked for isolation onto modified Ashdown's agar [27] minus gentamicin. Plates were incubated at 37˚C. At ~48 h and again at 72 h, plates were photographed. Photographs were used to examine and compare growth characteristics and to determine colony morphology among all isolates. Colony morphology was mapped onto the phylogenetic tree.

Motility testing

Flagellin associated with motility is a virulence factor that can be recognized by Toll- and NOD-like receptors [55]. Motility is easily tested in the laboratory by stab-inoculating a semi-solid media containing a bacterial growth indicator; if bacteria are motile then growth will be observed diffusing from the stab line. After ~48 h incubation at 37°C on Ashdown's agar (minus gentamicin), single colonies or cells from a visible region of uniform growth were selected and stab-inoculated into motility test medium with tetrazolium growth indicator (Hardy Diagnostics, California). A known motility-positive strain of *B. pseudomallei* unrelated to P314, Bp82, and *Staphylococcus aureus* were used as motility-positive and -negative controls, respectively. Controls were grown on standard Luria Bertani agar for ~48 h before use. Additionally, a sterile inoculating needle was used to stab a single vial of motility medium and used as an uninoculated control. Motility tests were incubated at 37°C; tubes were observed for growth at 48, 72, and 120 h and determined as positive or negative for motility based on comparison to the controls [56]. Results were mapped onto the phylogenetic tree.

Capsular polysaccharide (CPS) testing

Previously identified mutations in *wcbR* in the P314 isolates [25] probably led to decreased virulence of these isolates by reducing CPS-I production and expression, an essential virulence factor [57]. To test this hypothesis, we examined the CPS production using the InBiOS Active Melioidosis Detect Rapid Test (Seattle, WA), which specifically detects CPS produced by *B. pseudomallei* and *B. mallei* [58]. Tests were performed using the 110 ~48 h-old isolates described above and according to the manufacturer's instructions. Additionally, CPS and lipopolysaccharide (LPS) production was evaluated in seven isolates by Western blot analysis.

Antibiotic testing

Certain P314 isolates were screened using E-tests to determine their minimum inhibitory concentrations (MICs) to assorted clinically relevant antibiotics, including gentamicin, ceftazidime, doxycycline, meropenem, amoxicillin-clavulanate, and imipenem. For each antibiotic, test strips were placed on bacterial lawns and plates were incubated and interpreted as per manufacturer's instructions (bioMérieux, France).

Animal testing

BALB/c mice were purchased from Charles River laboratories and housed at Colorado State University. To examine the attenuation of the P314 and P314-associated environmental isolates, isolates were grown at 37°C in BHI medium to an OD_{600} of 1, supplemented with glycerol to 10% v/v, and frozen in multiple aliquots at -80°C. A vial of these stocks was thawed and serial dilutions plated on BHI agar plates to determine titer. Mice were intranasally challenged as previously described [59]. The amount of bacterial culture (CFU/dose) was verified by titration and plating the challenge inoculum on BHI agar plates. Mice were weighed and evaluated for clinical symptoms associated with *B. pseudomallei* infection prior to and after infection. This monitoring was performed daily for the first 7 days, and subsequently every 3–4 days to 21 days. The clinical scores indicated in the study were: 0 = normal, 1 = questionable illness, 2 = mild but definitive illness, 3 = moderate illness, 4 = severe illness, moribund—euthanized, 5 = found dead.

Ethics statement

This study involved the informed written consent of a single adult human subject (P314), with ethics approval obtained from the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research (approval number HREC 02/38). Mouse challenge experiments were performed under Select Agent and ABSL-3 containment practices. This work was conducted at Colorado State University and was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Every effort was made to minimize animal suffering. The protocol was approved by the Animal Care and Use Committee of Colorado State University (protocol number 17-7497A).
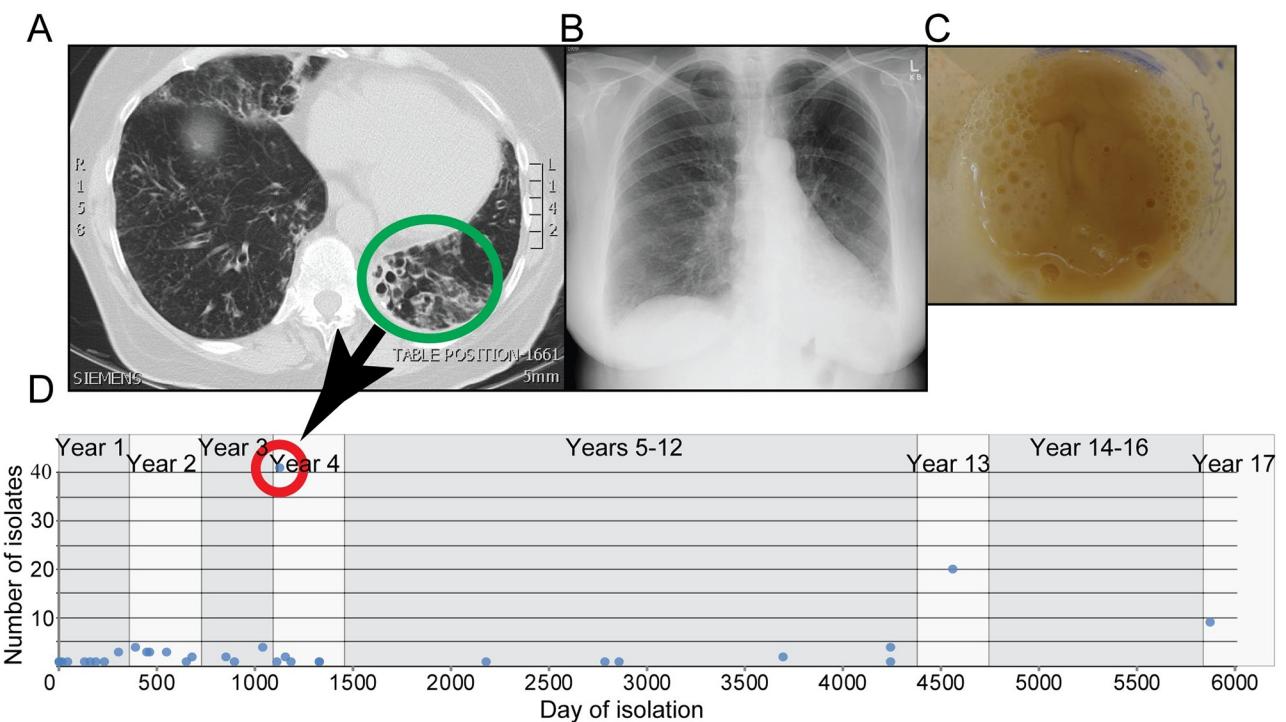
Results

A chronically infected patient

Melioidosis cases are unevenly distributed across northern Australia, with the pathogen found in soil and water in regions where cases occur [60]. Acute melioidosis dominates the clinical presentations and chronic infections that fail to resolve despite appropriate antimicrobial therapy, leading to long-term carriage, are very uncommon. Such cases have only been recognized in a very small number of melioidosis patients, most of whom have cystic fibrosis [12, 14]. However, this case is the only one where the patient survived, yet failed to clear her infection. In July 2000, a 61-year-old female patient (P314) with known bilateral non-cystic fibrosis bronchiectasis and past respiratory infections with *Mycobacterium avium* complex and *Pseudomonas aeruginosa*, presented with worsening cough and infiltrates on her chest X-ray consistent with a diagnosis of acute pneumonia. *B. pseudomallei* was cultured from her sputum as well as nose and throat swabs, confirming her first presentation of melioidosis. She was initially treated with 14 days of intravenous ceftazidime and her symptoms improved, but her sputum continued to culture *B. pseudomallei*. Subsequent multiple courses of antibiotics in various combinations over the next 10 years failed to clear her infection. These included intravenous ceftazidime, intravenous meropenem, nebulized ceftazidime, oral doxycycline and oral amoxicillin/clavulanate. She has a severe allergy to trimethoprim/sulfamethoxazole, which precluded its use. Respiratory symptoms and quantity of productive sputum production have fluctuated over subsequent years, but both improved substantially after she had a left lower lobectomy in August 2003, 37 months after initial diagnosis (Fig 1). With moderately severe bronchiectasis and intermittent *P. aeruginosa* infections (for which she receives intravenous meropenem), her daily sputum production is not unexpected. While she no does not exhibit symptoms of melioidosis, her sputum is often-culture positive for *B. pseudomallei*. P314 has not been on antibiotics with activity against *B. pseudomallei* for much of the last five years and is the only patient where anti-*B. pseudomallei* therapy has been discontinued without full clearance of the bacteria. We have collected and analyzed a >16-year longitudinal series of *B. pseudomallei* isolates for this study.

Environmental isolates

To identify the putative origin of her infection, P314's home environment was repeatedly sampled over an approximate two-year period (Aug 2002, Oct 2002, Jan 2003, May 2003 and Apr 2004). At the time, P314 lived in a semi-rural town in the tropical north region of the Northern Territory, Australia, and her home used unchlorinated groundwater from a private well for consumption and bathing. *B. pseudomallei* was cultured from the bore head (water prior to the tank), holding tank, pool, and household delivery pipes. In all, 21 environmental samples were

Fig 1. Patient timeline of treatment and _B. pseudomallei_ isolate collection. A) CT scan shows damaged, widened airways (green circle) in the lower left lobe that required surgery on day 1134. B) Chest X-ray after lobectomy. C) Patient's sputum in cup. D) Timeline of isolate collection. From excised lung tissue, we cultured 41 isolates on day 1134 (red circle).

investigated and 11 colonies were genotyped using multilocus sequencing typing and WGS. Six of the environmental isolates were multilocus sequence type (ST) 131, matching the patient isolates. The remaining isolates were from ST132 (n = 2), ST329 (n = 1), and ST29 (n = 2). One isolate was identified as _B. humptydooensis_. At the WGS level, all six ST131 environmental isolates were highly similar to the patient's clinical isolates. Genome sequences from the clinical and the six ST131 environmental isolates were compared to three other ST131 genomes (from other patients) to understand within-ST131 diversity (S1 Fig). We found that there was a maximum of 16 core genome SNPs among the different environmental isolates and the closest environmental isolate (from pool water) differed from the closest P314 clinical isolates by only 1 core genome SNP. However, no SNP separated all environmental isolates from the clinical isolates, consistent with probable evolution of environmental isolates from the original infecting ancestor in the ~2 years following the original infection event. In contrast, the phylogenetic distance of the P314-associated strains to ST131 isolates from other clinical cases was much greater. There was an average SNP distance of nearly 10,000, and in some cases 40,000 core genome SNPs, between isolates in this more diverse set. Thus, environmental sampling at P314's home point strongly towards infection occurring from her home environment and provided strains that closely represent the original infecting agent.
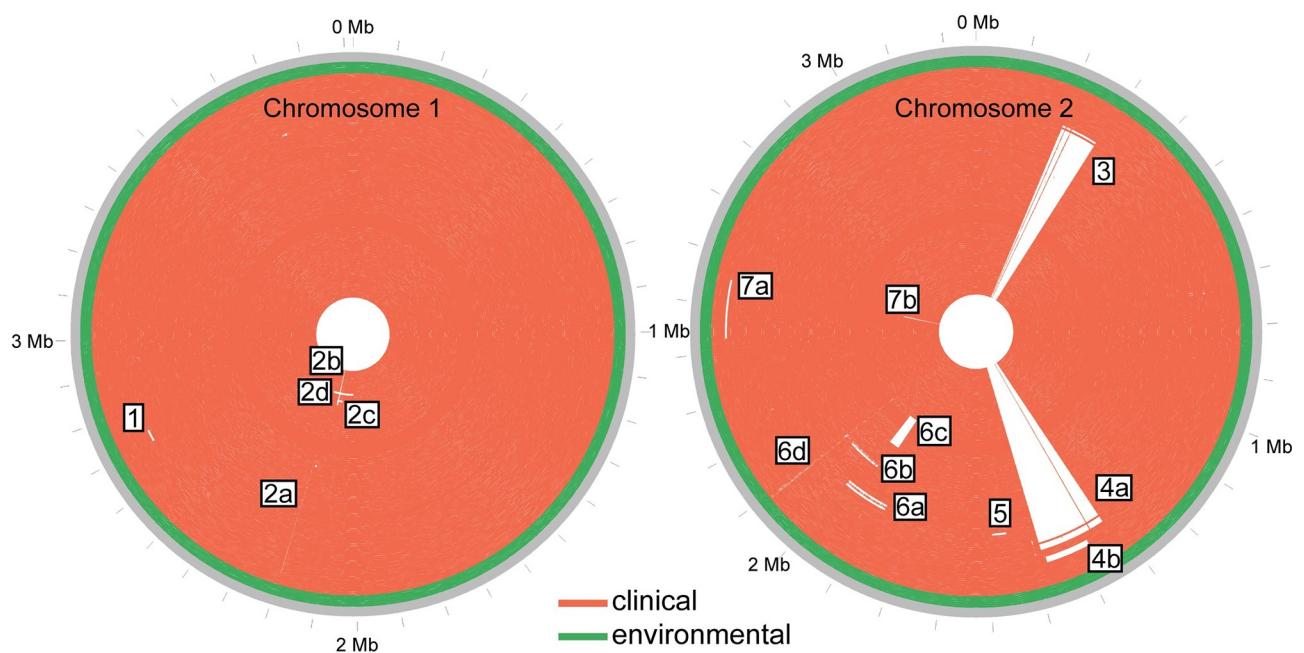
Genomic deletions

Seven regions that include large deletions were observed among the clinical isolates when compared to the reference genome of the environmental isolate, MSHR1435. Some of these deletions became fixed in the population, others occurred in lineages that went extinct, and others persisted in only a subset of the population. Overlapping deletions with different start

and end points, suggesting independent events, occurred in two regions. Subsequent decay surrounding previously deleted regions occurred in two additional regions. These deletions are depicted in Fig 2 while the phylogenetic location of each deletion event can be seen in Fig 3 and S3 Table. On Chromosome 1, we saw only two deleted regions. Deletion 1 (~31,000 bp) was observed in a single genome (MSHR1287) while Deletion 2a (1,416 bp) was identified in all, with the exception of four early clinical genomes. Three subsequent adjacent deletions (2b, 2c, 2d) occurred in a single lineage. All other large deletions were found on Chromosome 2. Deletions in regions 3 and 4 occurred relatively early over the evolutionary history of these genomes and while deletions 3 and 4a became fixed in the population, the 4b deletion event occurred independently on a lineage that subsequently went extinct. Regions found elsewhere in the genome show a high level of similarity to a small part of deletions 3 and 4a, giving the appearance in Fig 2 that each of these deletions are actually comprised of two separate deletions. Mixed-culture genomes (see below) also resulted in low levels of coverage in regions that were expected to be deleted. Examples of this can also be found in deletions 3, 4a, and 6a. Deletions in regions 4, 6, and 7 each involve the elimination of multiple overlapping genome segments, representing independent events and parallel evolution. Only one of these 15 deletions was found in all clinical isolates and no environmental isolates, suggesting that this initial deletion may be associated with an important early adaptation and the loss of virulence. This deletion is in the non-ribosomal protein synthetase gene cluster. This is a highly modular enzyme and the loss of this intragenic region may not eliminate its expression, but rather change the secondary products it produces [61].

Deleted genes

Some genes contained within annotated deletions were associated with known virulence mechanisms, including the type III secretion system (CXQ84_30725-CXQ84-30860) [62]. The single, and earliest, deleted region is found in all clinical isolates, and is thus likely associated with virulence attenuation. This region encompasses an annotated non-ribosomal peptide synthetase (NRPS) (CXQ84_31440), a large (3,290 aa), repetitive gene associated with deletions in regions 6c and 6d (Fig 2). This is shown in higher resolution in Fig 4. AntiSmash [63] results suggest that this region is annotated as a condensation domain-containing protein (SMCOG1127), although its potential product is unknown.

The CDSs from the large deletions shown in Fig 2 were also screened against 51 *B. mallei* genomes (S5 Table) with BLAT in conjunction with LS-BSR. Despite its different taxonomic nomenclature, *B. mallei* is a monophyletic clade within *B. pseudomallei* that has evolved to occupy a specialized niche that is generally associated with chronic respiratory infections (glanders) of equines, though it can also infect other mammals [64]. It has been shown [65] that the evolution of *B. mallei* is strictly clonal and all *B. mallei* are derived from a single ancestral *B. pseudomallei* strain. The results (Fig 5, S6 Table) demonstrate that many deletions in P314 have also been observed in *B. mallei*, suggesting parallel evolution, possibility associated with host adaptation in the respiratory tract. Many of these *B. mallei* isolates are still virulent, suggesting that the observed parallel evolution is not necessarily associated with pathogenesis. To get a better sense of whether any of these deleted regions are associated with adaptation to long-term carriage, we also screened them against pairs of sequential genomes from 7 long-term *B. pseudomallei* infections in CF patients [14], and a panel of 262 genomes from environmental and acute infection isolates (Fig 5, S6 Table). Parts of deleted regions 2d and 6 were also absent in some of the chronic CF genomes, however they were missing in both initial and subsequent isolates, indicating that their deletion was not specific to within-host evolution over the course of these infections. Some of these deleted regions were also absent in isolates

Fig 2. The location of large deletions in P314 genomes relative to MSHR1435. Whole genome sequencing of 120 *B. pseudomallei* collected over 17 years reveals large chromosomal deletions across both chromosomes. The genomes' order from the outside to inside concentric circles corresponds to the top to bottom order on the phylogenetic tree (Fig 3). This inside-to-out arrangement approximates the evolutionary trajectory of the infecting *B. pseudomallei* population. MSHR1435, which is used as the reference, is outermost and is a closed genome while all others are high quality drafts. All environmental isolate genomes (including MSHR1435) are basal in the phylogeny (Fig 3) and found in the outermost rings (green) of the concentric circles. Raw sequence reads were mapped to the reference chromosomes and any base with a depth of coverage of <25% of the average genome coverage was considered to be missing. The figure was constructed with Circos [43]. Coordinates for deleted regions are as follows: 1(2700183–2733804), 2a (2196965–2198380), 2b (2141162–2157867), 2c (2106537–2157867), 2d (2006273–2214755), 3 (204194–294524), 4a (1322983–1479924), 4b (1375399–1474558), 5 (1551804–1588830), 6a (1874062–1994243), 6b (1934891–1974776), 6c (1956279–2122568), 6d (2069785–2099797), 7a (2429295–2552683), 7b (2546831–2555624).

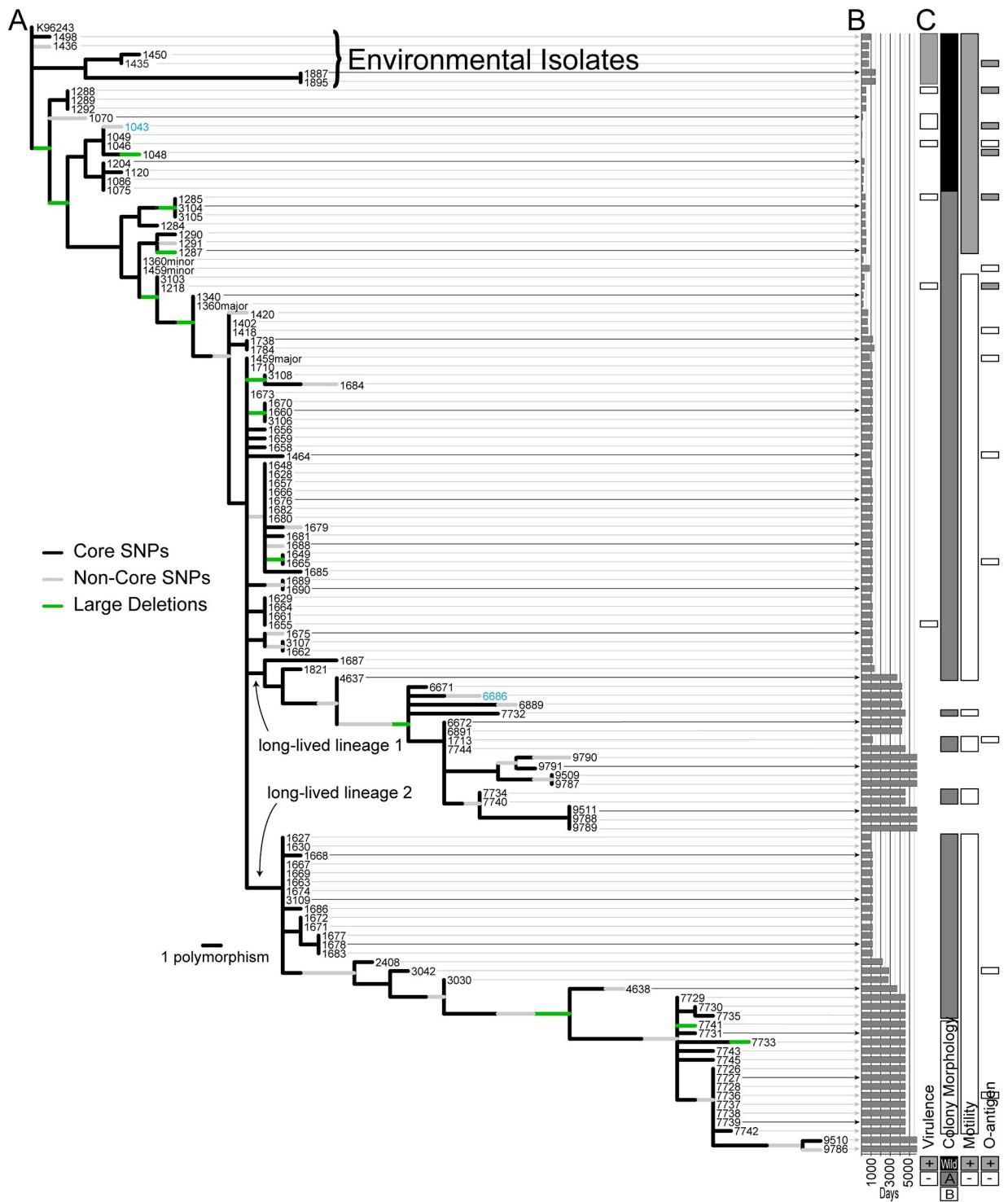https://doi.org/10.1371/journal.ppat.1008298.g002

from the environment and acute infections, providing additional evidence that these deletions are non-adaptive or neutral. Conversely, parallel deletions among only long-term carriage isolates provide evidence of adaptive evolution. Region 4b, for example, was lost independently twice in P314, is absent among *B. mallei*, was deleted over the course of one CF infection, yet is always present among the diverse panel of environmental and acute infection isolates. The deletion of genes in this part of region 4d is likely to be adaptive and include genes for energy and amino acid metabolism.

Small genomic deletions

We have previously described a small number of indels in P314 genomes [25]. These regions were screened against all P314 genomes and mapped across the phylogeny (S7 Table). The *wcbR* frameshift (CXQ84_04665), which results in an altered capsular polysaccharide synthesis protein, was observed in all but the four most basal genomes. Similarly, we have previously described a single nucleotide insertion that creates a frameshift in the *wbiI* gene of some P314 isolates, which is correlated with a loss of the O-antigen [66] and occurred on branch 6 (S2 Fig).
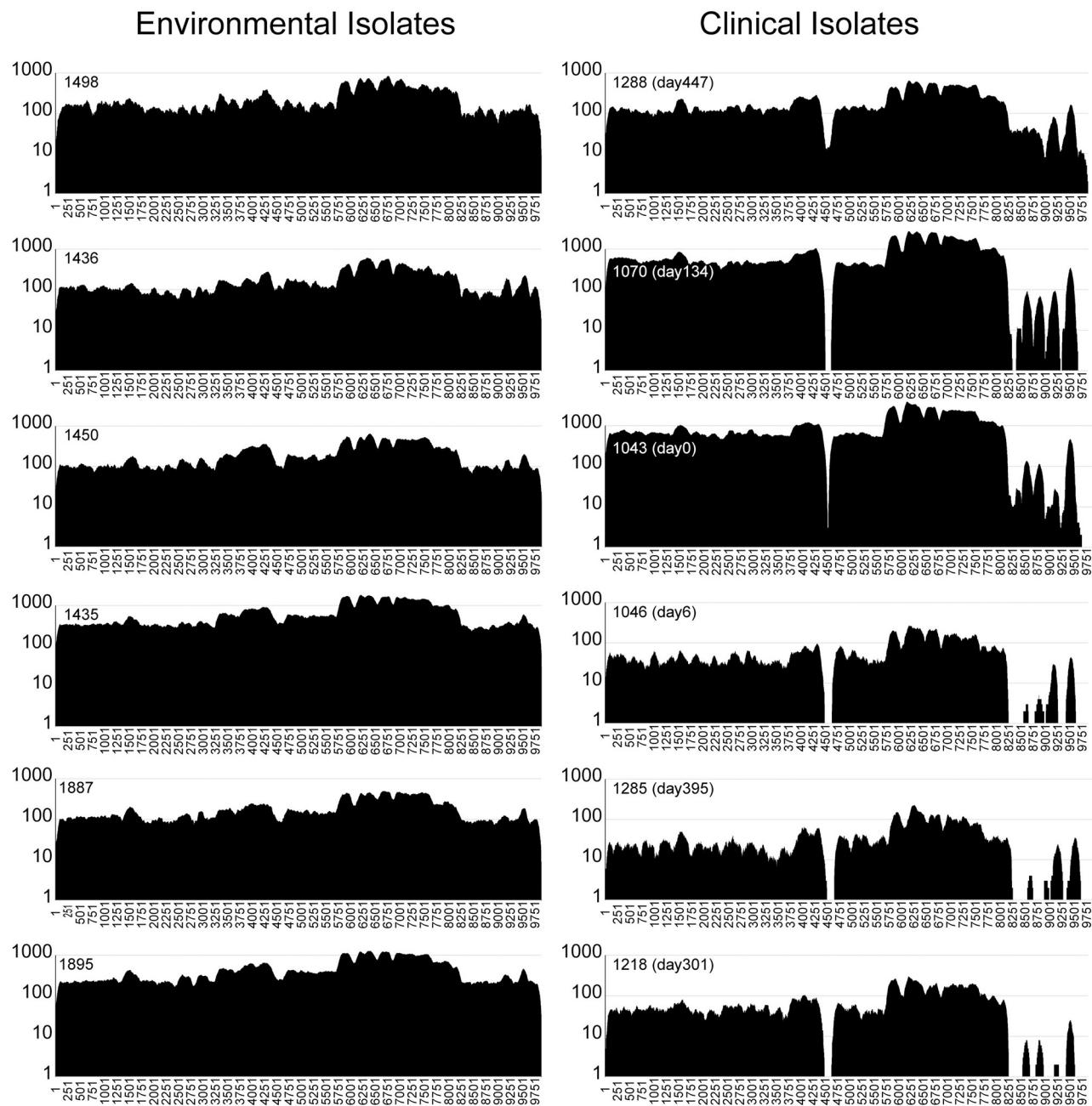
Phylogeny

The most parsimonious phylogenetic tree was constructed using 126 core SNPs and has a consistency index of 0.97, indicating character changes that occurred only once, with the

Fig 3. Evolutionary changes. (A) Phylogenetic tree showing evolutionary locations of polymorphism, (B) day of collection for each isolate, and (C) phenotypic characteristics (virulence, motility, and three colony morphotypes—W = wild type, A = small colony variants, B = fastidious). Core genome SNPs are represented as black branch segments, while non-core genome SNPs are gray. The large deletions shown in Fig 2 are green branch segments. We previously compared the genomes in blue [25]. Where mixtures are observed, the major and minor components are labeled. Methodological details for phylogenetic reconstruction are provided in the materials and methods section.

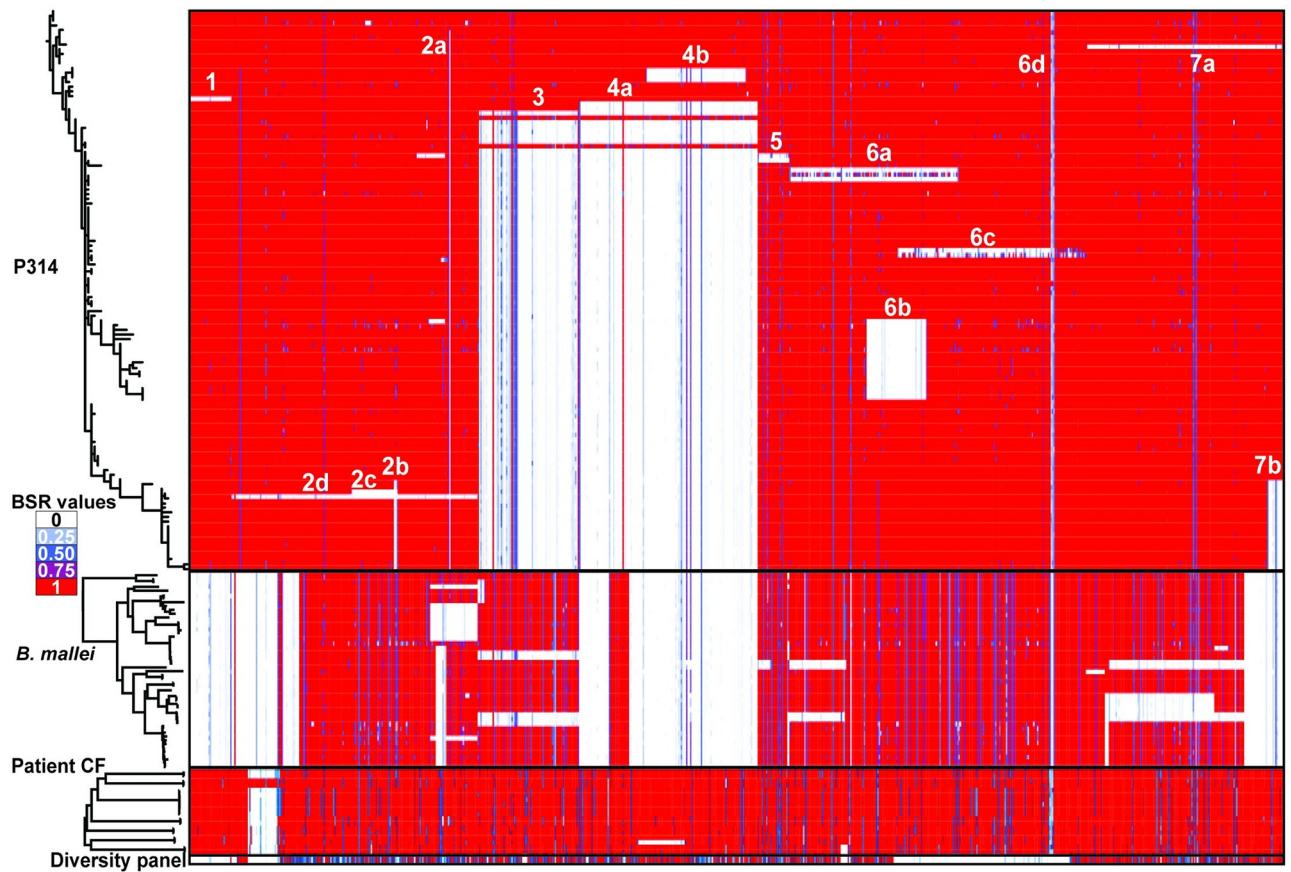https://doi.org/10.1371/journal.ppat.1008298.g003

Fig 4. Sequence read mapping to a non-ribosomal peptide synthetase (NRPS) gene from different isolates. The Illumina reads from each of the isolates in this study were aligned to the NRPS gene from the MSHR1435 reference (MSHR1435 locus CXQ84_31440). This figure is a higher resolution representation of the read mapping shown in Fig 2 and covers the 6c and 6d regions shown there. Twelve examples of this analysis are illustrated with six environmental and six clinical isolates, with the complete isolate set present in the supplemental material (S1 Table). The read mapping shows the early loss of the internal gene sequences around position 4500 and later deletions in the 3' region. Isolate are arranged from top to bottom in the same order as in the phylogenetic tree (Fig 3). These same isolates were all used for animal challenge experiments (see below).

https://doi.org/10.1371/journal.ppat.1008298.g004

exception of two homoplastic SNPs (Fig 3). Consideration of homoplasies and their location in the phylogeny is important for evaluating the degree to which genetic inheritance is clonal or involves lateral gene transfer (LGT) [67–70]. SNP phylogenies of clonal populations typically have little or no homoplasy, with any character state conflict caused by convergent,

Fig 5. Convergent gene deletions among long-term carriage and non-host-adapted genomes. Genes from the 7 deleted regions (Fig 2) found among P314 isolates are mapped against 51 *B. mallei* genomes, chronic *B. pseudomallei* genomes from 7 CF patients, and a diverse panel of acute/ environmental *B. pseudomallei* genomes. Parts of each of the seven regions deleted among P314 genomes were also deleted in *B. mallei* genomes. Portions of three of the seven regions were also deleted in some genomes associated with long-term CF infections, however two of these deletions were evident in both the initial and later isolates, suggesting that their absence is not necessarily linked to host adaptation. Indeed, these genes are also absent in the diverse panel of non-host-adapted (acute and environmental) genomes. Conversely, part of region 4b was deleted over the course of a CF infection and among all *B. mallei*. None of the genes from this region were deleted in any of the genomes from the acute and environmental panel, suggesting that their deletion may be a specific adaptation to long-term carriage. Deletions in the 6d region correspond to the NRPS gene (Fig 4) and are also observed in other genomes from these panels. Putative deletions seen among P314 genomes were assumed to be due to fragmented assemblies if they could not be verified with read mapping (Fig 2) and were thus not assigned a number/letter.

https://doi.org/10.1371/journal.ppat.1008298.g005

parallel, or reversal mutations rather than by LGT [71]. While even low levels of homoplasy can be an indicator of limited LGT, such recombination is not likely to have occurred over the course of the P314 infection. For one of the homoplastic SNPs, LGT would have required transferring an allele only found among isolates from years 1 and 2 to isolates found in year 17. While we cannot eliminate LGT as a cause for the second homoplastic SNP, no other polymorphisms would have been linked to any putative LGT event. Furthermore, a putative LGT event would not have played any evolutionary role in this population as the lineages containing both strains with the shared homoplastic allele immediately went extinct. This clonal mode of inheritance with no evolutionary role for recombination is in stark contrast with the overall *B. pseudomallei* population where diversity, and presumably adaptive evolution, is driven primarily by LGT [18, 72]. This tree is also likely to be a highly accurate reconstruction of the true evolutionary patterns as trees constructed with core genomic SNPs have been shown to be highly evolutionarily stable and yield accurate trees for clonally replicating populations [73–75]. We have high confidence in the accuracy of the phylogeny given the high consistency index and as

four equally parsimonious trees were produced that differed only in the topological assignment of two genomes that shared an allele at a homoplastic SNP. A maximum likelihood tree of the core SNPs also exhibited the exact topology and branch lengths as one of the four maximum parsimony trees. Bootstrapping is commonly used for assessing the robustness of a tree, however even when introducing the concept, Felsenstein recognized that it was a poor method for trees with little homoplasy as it underestimates confidence for nodes supported by a small number of SNPs [76] [75]. Given the phylogenetic stability of SNPs in low-homoplasy datasets, we can have high confidence in branches supported by even a single SNP. Measures of homoplasy are therefore a more appropriate direct measure of accuracy in such situations [77]. The addition of 36 non-core genome SNPs and 15 large deletions as evolutionary characters to the tree increased phylogenetic resolution (Fig 3). Most of these additional genomic characters can be mapped to a single point on the tree; the independence of each large deletion event was inferred by different breakpoints even where deleted regions overlap across strains. The single exception to this is a single nucleotide deletion in two distantly related genomes and must therefore involve two independent events.

We used the genome of K96243 as an outgroup for rooting this dataset. Given the genomic closeness of the environmental isolates, we could not assume that they were an outgroup as a diverse infecting inoculum could result in environmental strains nested among the clinical isolates. Indeed, diverse inoculating populations of *B. pseudomallei* have been documented [78, 79]. Likewise, the first isolate recovered from the infection (MSHR1043) cannot be assumed to represent an outgroup if the infecting inoculum was diverse, or if *in vivo* evolution had already generated diversity. The rooting of this phylogeny and monophyly of the clinical isolate clade suggests that the infection evolved from a recent single hypothetical ancestor that was likely present in the inoculum. This does not preclude the possibility of a diverse multiple bacteria inoculum where all but one lineage went extinct before the first isolates were collected. In fact, the initial isolate was not the most basal in this phylogeny and suggests that early diversity is not represented in this sample set. The rooting also shows that some of the sampled environmental isolates are closely related to the clinical isolates and differ from the initial infecting strain by only a single SNP. *In vivo* diversity was observed before the collection of the first clinical isolate (MSHR1043), which is not among the most basal isolates in the tree.

Throughout the tree, numerous SNPs and deletions (discussed above) are observed that are specific to lineages that went extinct. In most cases, these lineages only contain single or small groups of isolates, suggesting limited fitness. However, one clade that went extinct contains 13 isolates, suggesting some degree of ephemeral fitness for this subgroup. The large number of isolates sequenced, particularly at the beginning of year 4, allowed us to sample lineages and diversity that would otherwise be concealed. Only two of the many nascent lineages persisted beyond year 4 and continued to diversify. After >16 years, all the extant isolates are derived from these two long-lived lineages, and in the first, two sublineages persist (Fig 3).

Selection analysis

We found little evidence for selection effects in 88 SNPs from protein coding regions or from the entire suite of 177 phylogenetic characters. The SLAC comparison of synonymous and nonsynonymous nucleotide substitution rates reveal no sites where dN-dS was significantly different from 0. The codon-based neutrality test of Nei and Gojobori was dN-dS = 1.196 (p = 0.234) and failed to reject the null hypothesis of neutral evolution. Similarly, Tajima's D for the entire set of isolates (environmental and clinical) showed no significant departure from the null model of neutral evolution (D = -1.52). However, this relatively more negative Tajima's D is approaching the critical value for significance, and may indicate that either purifying

selection or population expansion is occurring. The latter is consistent with the Bayesian Sky-grid plot. Finally, when we inspected the earliest and latest time points, we did not find evidence for deviation from neutral evolution (D = -0.43 and 0.01, respectively).

Mixed genomes

During the course of SNP discovery and evaluation of deleted regions, we identified heterogeneity at various loci among genomes. The source of the mixture was unknown, but could indicate incomplete separation of cells during laboratory isolation, library preparation, or optical contamination during the sequencing process; in all cases, the contamination was low, was easy to identify, and didn't appear to affect the overall results due to the low level of homoplasy observed in the phylogeny. As such, we assumed that this heterogeneity reflected the sampled population and did not re-isolate and re-sequence the few isolates that we suspected contained multiple genotypes. Our level of analyses focused on rigorous identification of even the smallest differences among genomes and the ability to observe low-level read mixtures. In two cases, we were able to separate mixtures into minor and major populations based upon SNP frequency in the read data and determine their respective phylogenetic position (Fig 3) by considering which SNP loci showed heterogeneity among sequencing reads. Evidence of strain mixtures was most clear in the deletion matrix (Fig 2), where four instances of mixtures were identified. Even when a dynamic filtering approach was employed to reduce as much contamination noise as possible, intermittent and low-level coverage was still evident in deleted regions for some isolates. Such mixed isolates might have an impact on phenotypic attributes.

Phenotypic characteristics

Colony morphology of individual isolates could be described by one of three categories: wild-type, small colony variants (SCVs), and fastidious (S8 Table). As we have previously reported [25], wild type colonies show robust growth and are easily recognized as *B. pseudomallei*. SCVs, however, are difficult to recognize as *B. pseudomallei*; even after 48 h of growth, colonies were typically only <1 mm in diameter. Smaller still were isolates categorized as fastidious; after 72 h, these isolates could barely be discerned on agar plates, appearing as a nearly clear sheen in the primary zone of inoculation with few, if any, pin-prick sized colonies in other areas.

The change from wild type to SCVs arose during the evolutionary history of these isolates with SCVs being present in sputum samples as early as 301 days after the first isolate (MSHR1043) was collected (Fig 3, S8 Table). SCVs have persisted throughout P314's infection, they occur in both long-lived lineage 1 and 2 (Fig 3), and were isolated from different sample types (i.e. nose and throat swabs, various locations in the lung; S8 Table). Fastidious isolates appear to occur only in long-lived lineage 2 (Fig 3) and are derived from a single sputum sample produced 4,560 days post initial isolate. Interestingly, this sputum sample also produced SCVs in lineage 1 (MSHR7732, 7744, 7734, and 7740) and in lineage 2 (MSHR7729, 7730, and 7735; Fig 3) demonstrating the mixed phenotypic and genotypic nature of this population at this time point. It is unclear if fastidious isolates have developed in lineage 1 as the most recent isolates in this lineage (MSHR9790, 9791, 9509, 9787, 9511, 9788, and 9789) have not been characterized.

Loss of motility was observed in 85 of 110 clinical isolates which include 72 SCVs and all 13 fastidious isolates (S8 Table). All wild type morphology isolates (n = 18) were motile (Fig 3). A small subset of SCVs (n = 7) that were isolated from P314's nose and throat between days 395–461 post initial isolate were also motile (Fig 3 and S8 Table). Excluding the minor variants of isolates MSHR1360 and 1459 (Fig 3), loss of motility occurred on a single branch on the
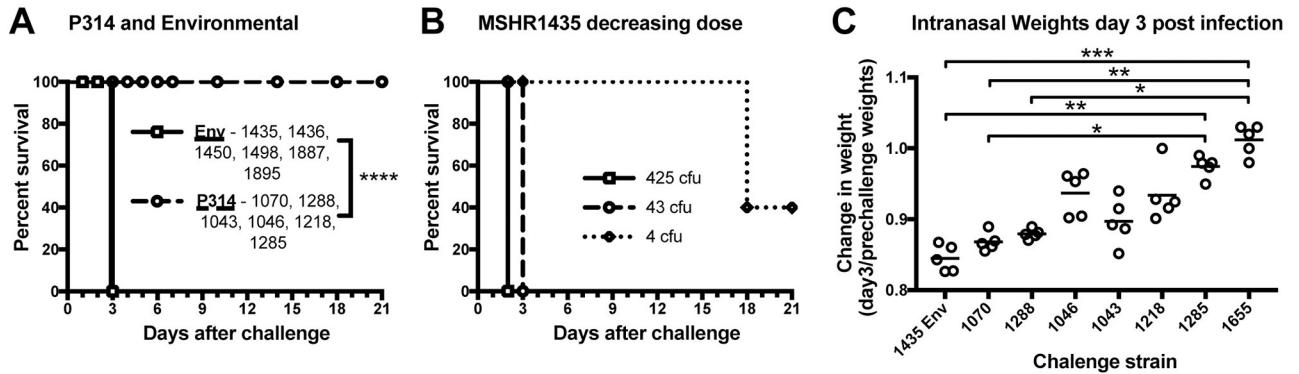
phylogenetic tree after the phenotypic change to SCVs. This indicates that as late as day 461 there were motile SCVs still present in the *B. pseudomallei* population (see isolates MSHR1285, 3104, 3105, 1284, 1287, 1290, and 1291; Fig 3). Anecdotally, we noted what appeared to be diminishing motility over evolutionary time in these seven isolates prior to full loss of motility in later isolates. This was indicated by the distance the bacteria appeared to diffuse from the stab line in motility media, with earlier isolates appearing to have a larger diffusion cloud than later isolates. This observation suggests that loss of motility may not be as simple as a single genetic deletion or mutation, which would be implied by motility loss being mapped to a single point on the phylogenetic tree. Instead this suggests a pattern of down-regulation to the mechanism for motility, perhaps interacting with global changes in morphology, which warrants further investigation.

Other phenotypic changes observed in these isolates include the decreased production of the carbohydrates CPS and LPS O-antigen (S8 Table, Fig 3), both of which are important virulence factors [66]. Antibiotic resistance has emerged in many of the clinical isolates and is no doubt selected by the therapeutic regimes used in the futile attempts to cure the infection. One of the most complex phenotypes to change, however, is virulence in animal models.

Animal studies

To determine if the patient isolates remain virulent in a mammalian host, BALB/c mice were intranasally infected with a select number of *B. pseudomallei* isolates collected from the patient or from the environment at the patient's home (Fig 6A). The patient isolates were selected to include representatives of the early clades as well as a few later strains, thus enabling us to examine temporal and genetic correlates to virulence changes (Fig 3a). All of the mice infected with the patient isolates survived to 21 days after infection, regardless of the inoculum level. However, all of the mice infected with the environmental isolates succumbed to infection or were euthanized due to clinical score and weight loss by day three after infection. The back-titrated doses of the environmental isolates were between 2600 and 6000 CFU/mouse depending on the isolate. In contrast, the doses of the clinical isolates were higher, between 9250 and 11,040 CFU/mouse, suggesting that the clinical isolates were attenuated even at higher intranasal doses. To better quantify the virulence of the environmental isolates, BALB/c mice were intranasally challenged with a decreasing dose of MSHR1435. All of the mice that were infected with >43 CFU succumbed to infection by day 3 after challenge. In addition, 50% of the mice that received 4 CFU succumbed to infection by 18 days after infection, suggesting that the 21-day LD50 for MSHR1435 is around 4 CFU (Fig 6B). In combination, these data suggest that the clinical isolates are highly attenuated in the BALB/c model even at an infective dose ~1000 fold greater than the estimated LD50 of the related environmental isolate. The phylogenetic position of the tested clinical isolates shows that attenuation evolved rapidly after the infection, as none of the clinical isolates had evidence of virulence in our study.

In addition to mortality, we examined morbidity in the mouse model by monitoring weight loss and clinical scores after infection. Infected mice were weighed and scored prior to and after challenge. All infected mice, except those infected with MSHR1655, lost weight until day 2 after challenge. At that point, they either succumbed to infection, had to be euthanized, or started to recover (S3A and S3B Fig). We analyzed the change in weight loss at day 2 to the phylogenetic distance from the environmental isolates (Fig 6C). There is a general decrease in weight loss as the isolates tested become less related to the environmental strain. The average percent weight loss is around 12–15% for MSHR1435 and MSHR1070 then decreases to a 0–2% weight loss for MSHR1285 and MSHR1655. This weight loss is significantly different between these strains (Fig 6). In addition, the weight loss, or the lack thereof, correlates with

Fig 6. Mortality and morbidity of P314 isolates and P314-associated environmental isolates. A-C) BALB/c mice were intranasally inoculated with *B. pseudomallei* isolates (n = 5/group). Mice were monitored for clinical score and were weighed daily for weight loss and if needed were euthanized (A and B). The inoculum for the single dose (A) and the 1435 decreasing dose (B) was back titrated and the CFU/mouse was determined and listed in the figure or below. Back titration results: (A) P314-associated environmental isolates, 1435–2,600 CFU, 1436–3,400 CFU, 1450–2,800 CFU, 1498–3,025 CFU, 1887–6,000 CFU, 1895–6,250 CFU. P314 isolates, 1070–9,250 CFU, 1288–9,550 CFU, 1043–11,040 CFU, 1046–10,350 CFU, 1218–10,150 CFU, 1285–10,450 CFU. Survival significance was determined by Mantel-Cox log rank analysis comparing the individual environmental isolates to the patient isolates. C) Individual percent change in mouse weights at day 2 are shown by dots (n = 5/group) and the mean weight is shown by the bar. The isolates are also listed by the relatedness to the environmental isolates cluster (right to left). The change in weight significance was determined with a Kruskal-Wallis test by comparing the change in weight at day 2 for each isolate listed. BALB/c mice were also challenged with 1655 with no mortality but the experiment was terminated early at 14 days (1655) and is not shown. * P<0.05; ** P<0.01; *** P<0.001; **** P<0.0001.
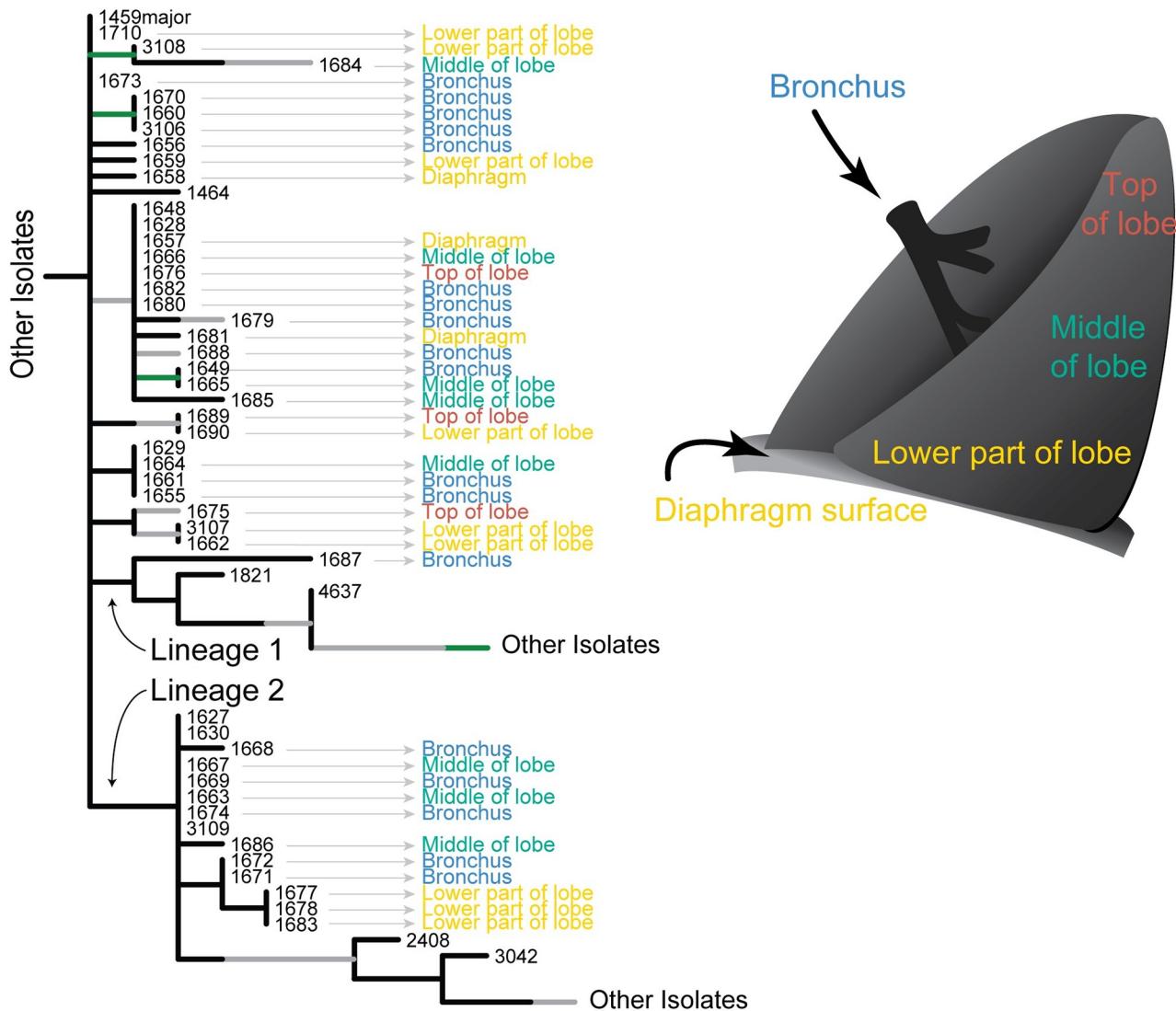
an increase or decrease in clinical scores (S2C Fig). In combination, these data clearly demonstrate clinical differences imparted by the environmental versus clinical isolates and additionally suggest that as clinical strains evolved in the host lung microenvironment, mammalian morbidity decreased. This evolutionary loss of virulence is likely to be highly adaptive and mutations and deletions of genes associated with virulence have been lost in other examples of long-term carriage [14, 80].

Virulence gene screen

A set of 551 genes associated with virulence in *B. pseudomallei* [81] were screened across P314 genomes with LS-BSR. Significant variation existed in the conservation of these regions across genomes (a subset of these genes is shown in S9 Table). However, none of these regions were exclusively present in the environmental genomes, suggesting that the differential distribution of these regions fails to explain the virulence phenotype observed in the mouse model. Additionally, no clear trend across isolates used in the animal challenge were observed, suggesting that variation in the conservation of these genes fails to explain differences in animal morbidity.

Spatial structuring

Removal of the lower left lung lobe on day 1134 afforded the opportunity to collect 41 isolates from various sites and test hypotheses regarding spatial segregation of genotypes. Isolates collected from similar sites are found throughout the tree and multiple clades within the tree contain isolates from multiple, and even distant pulmonary sites (Fig 7). Moreover, isolates from sputum collected before and after the lobectomy are placed in different positions on this tree, suggesting that not only was there a lack of clustering by genotype within the excised tissue, but also that isolates from the excised tissue are representative of the overall lung population. The emergence of the two long-lived lineages with phenotypic and genotypic characteristics raises the possibility of subsequent partitioning of this population. However, at the time of the
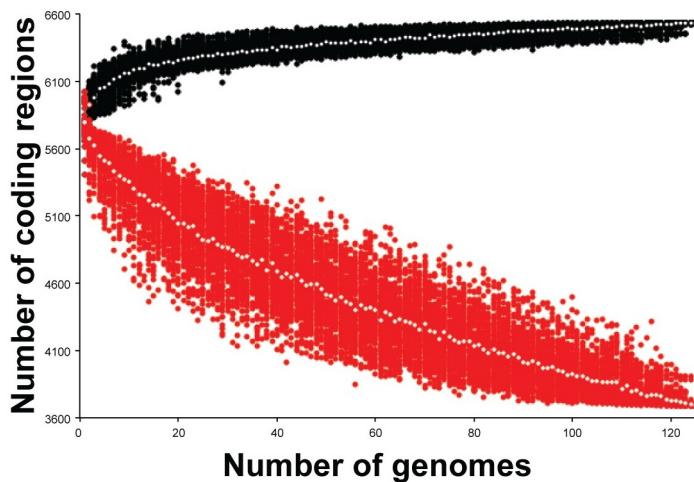
Fig 7. Spatial dynamics. Tree showing location of lobectomy samples suggest one non-segregated population. Samples from the same region are found across clades on the tree and multiple clades each contain samples from multiple sites.

lobectomy members of these two lineages were present in the excised tissue and had already begun to phylogenetically diverge, showing no evidence of physical separation.

Pan-genome size and diversity

B. pseudomallei genomic content is greatly influenced by LGT. However, in the P314 population, there is no evidence for the uptake of exogenous DNA from either conspecifics or allospecifics. The pan-genome accumulation curve (Fig 8) shows a closed pan-genome. Conversely, the number of CDSs shared in all clinical genomes decreases without reaching an asymptotic minimum. The initial increase in the pan-genome size as well as the continued decrease in shared gene content is due solely to the deletion of genomic regions. Consequently, with the switch from a free-living to an obligate intracellular lifestyle, the pan-genome size is limited to the gene content present at the time of infection and the generation of diversity is therefore

Fig 8. Pan and core genome dynamics. Size of pan (black) and core (red) genome with increased random sampling of genomes show a closed genome and the reduced size of some genomes due to genomic deletions.

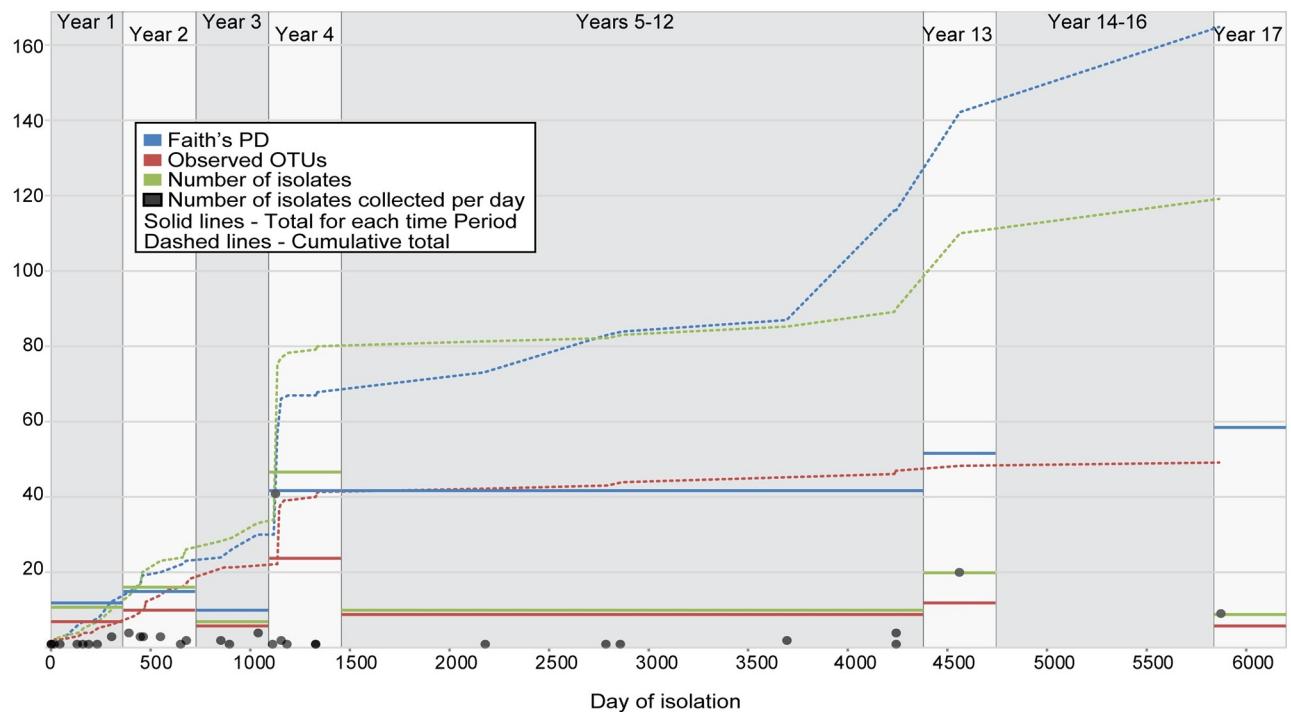https://doi.org/10.1371/journal.ppat.1008298.g008

tied to manipulation of this endogenous gene content. As the core genome curve does not reach an asymptotic minimum, continued evolution through novel deletions is likely for this population.

Evolution of diversity

While LGT has not played a role in the evolution of the P314 population thus far, mutation through deletion and point mutation has. While much of the diversity generated within the clinical population is ephemeral, cross sectional diversity increases longitudinally. The phylogeny (Fig 3) shows that even before the first isolate was collected, multiple lineages had diverged through the accumulation of SNPs and deletions. In the first year, 11 isolates were sampled, yielding 7 distinct genotypes (observed Operational Taxonomic Units (OTUs)) separated by 12 core and non-core SNPs (Faith's PD). In the 2nd and 3rd years, the number of isolates collected, the number of observed OTUs, and Faith's PD for each year remained similar to year 1 values (Fig 9). The population diversity however was not stagnant as the cumulative number of observed OTUs and Faith's PD increased over this time. In year 4, we collected 47 samples (41 of which were cultured from the lobectomy) but only observed 24 distinct OTUs. Faith's PD increased to 42 SNPs due to the sampling of multiple short lineages. Through the periods of years 5–12, year 13, and year 17, the number of isolates collected and observed OTUs was also similar to each of the first three years (Fig 9). Although many of the short lineages that were observed in year 4 did not persist, Faith's PD, however remained high for subsequent periods because of the continued diversification of the two long-lived lineages. The cumulative total of observed OTUs and Faith's PD increases dramatically at the time of the lobectomy (due to the large number of isolates collected) and continues to increase towards year 17. The difference between Faith's PD cumulative total and the value observed during year 17 is due to lineages that have gone extinct, although it is certainly possible that some presumably extinct lineages persist but were not sampled.

Molecular clock and evolutionary rates

The longitudinal sampling of P314 over a >16-year period and the whole genome data set provide the opportunity to understand the population evolutionary rates in the context of a
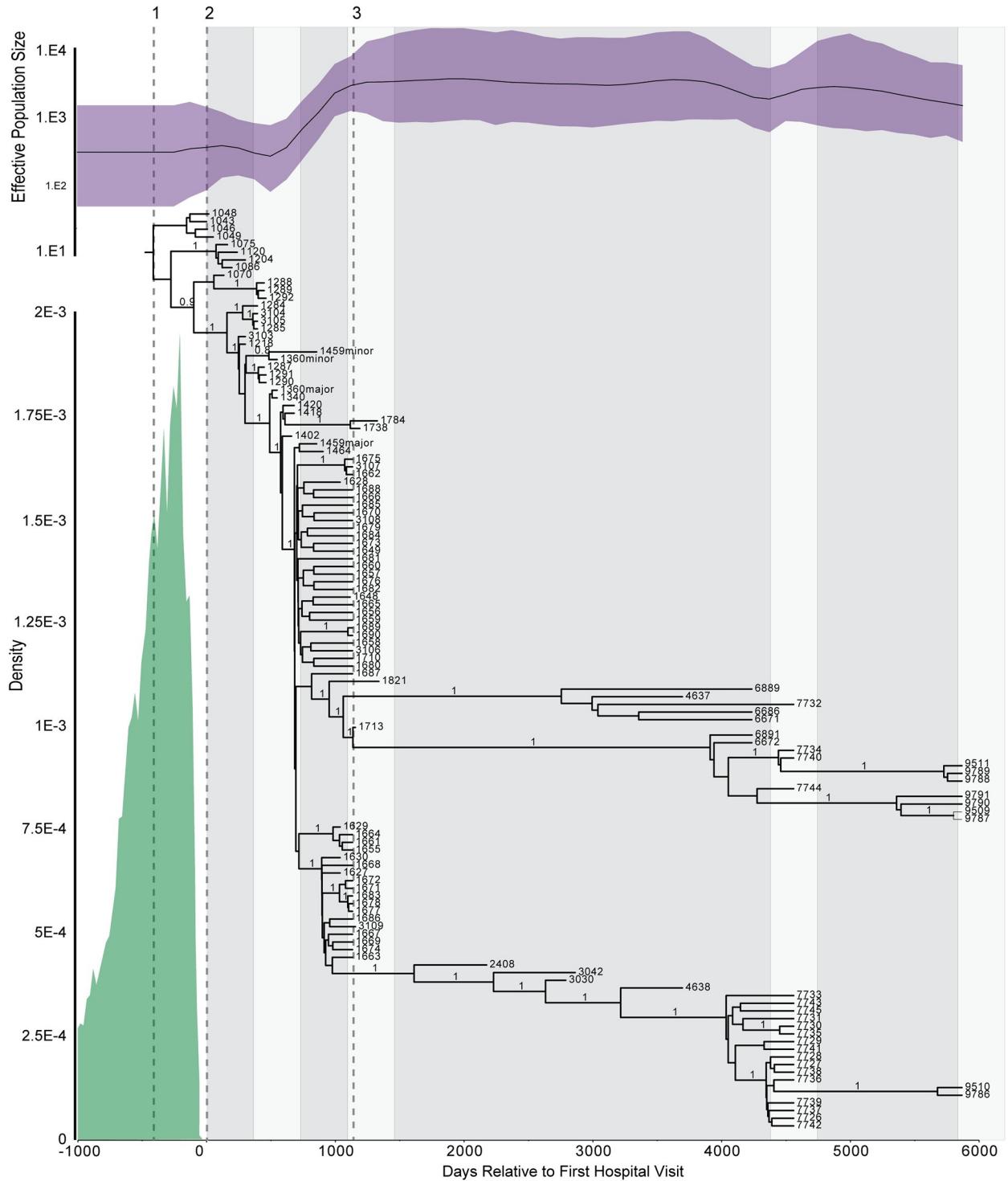
Fig 9. Population diversity dynamics. Estimates of diversity (Faith's PD and Observed OTUs) in each time period and accumulating over time. This shows that mutations in the first part of year 1 cause a rapid increase in accumulated diversity that continues with time. Continued extinction of earlier genotypes suppresses this accumulation of diversity at each time period. This changes in year 4 when diversity in this time period dramatically increases and, despite the extinction of most newly emerged lineages, is maintained thereafter due to the persistence and continued divergence of two clades. The number of observed OTUs and isolates in each time period remains mostly constant with the exception of year 4 due to the large number of isolates collected from the lobectomy.

molecular clock. The median evolutionary rate across the entire genomic dataset was 1.7×10^{-7} substitutions/site/year (95% HPD 1.3×10^{-7}–2.1×10^{-7}). In agreement with this estimate, it has been previously shown that within patient evolutionary rates of *B. pseudomallei* average 4.9×10^{-7} substitutions/site/year [14], *B. dolosa* averages 3.3×10^{-7} substitutions/site/year [80], and *B. multivorans* averages 3.6×10^{-7} substitutions/site/year [82]. We observed that the effective population size was the smallest at the time of infection, and increased until the partial lung lobectomy occurred in year 4, after which it was stable (Fig 10).

The two extant long-lived lineages differ in their evolutionary rates. This is visually apparent from a comparison of the branch lengths in Fig 3 where the first long-lived lineage is ~70% of the length of the second. In maximum parsimony analysis, the branch length is determined by the number of character state changes, which in this data set is the number SNPs that occur along each lineage. The SNP evolutionary rates can be estimated from the regression slopes of the branch lengths between the root and each genome against the day of isolation (S4 Fig). The regression slope for the isolates before the long-lived lineage bifurcation and the slope for long-lived lineage 2 are not significantly different from each other (F = 2.244, DFn = 1, DFd = 96, P = 0.1374), consistent with a similar evolutionary rate. In contrast, the slope of long-lived lineage 1is significantly different from the pre-bifurcation slope (F = 60.33, DFn = 1, DFd = 80, P<0.0001) and long-lived lineage 2 (F = 158.6, DFn = 1, DFd = 52, P<0.0001). Given these observed similarities and differences in evolutionary rates, we excluded the long-lived lineage 1 isolates to calculate the X-intercept (-307.6; 95% CI -240 to -378.2) as an estimation of the time

Fig 10. Bayesian estimation of infection date and population dynamics. SNPs from 120 clinical genomes longitudinally sampled over 17 years, were evaluated to determine the estimated time of infection (TMRCA for all clinical isolates) and population dynamics over the course of the infection. The 95% highest posterior density interval is in green and suggests a median time of infection of 343 (mean: 414, 95%HPD: 10–999 days) days prior to the first hospital visit. Posterior probabilities that are greater than 0.80 are indicated for all interior branches. The Bayesian Skygrid (top, purple) demonstrates the fluctuation of the effective population size over the course of the infection. Three important events are indicated: 1) Time of Infection, 2) Time that the first isolate was collected, and 3) Time when lung lobectomy was performed. Iterative downsampling of isolates at the time of the lobectomy show that effective population size estimates are not driven by the numerous samples collected at the time of the lobectomy (S5 Fig).

https://doi.org/10.1371/journal.ppat.1008298.g010

to the most recent common ancestor (MRCA). This represents an approximation of when the patient was infected (days prior to the patient's first clinical presentation).

We also used a Bayesian approach (Fig 10) to approximate the date of infection. Given the anomalous rate of long-lived lineage 1, calculating the time to MRCA with and without these isolates resulted in estimates of median times of 343 (mean: 414, 95%HPD: 10–999 days) and 179 days (mean: 223, 95%HPD 6–558 days), respectively, prior to the patient's first clinical presentation. The Bayesian TMRCA estimates are earlier than those estimates derived from the linear regression analysis. The regression estimates are based on root-to-tip distances calculated from a maximum parsimony tree, whereas the Bayesian estimates incorporate substitution, site, clock, and demographic models.

Genome decay

Extensive genome decay, in the form of large genomic deletions, has been documented in *B. mallei* and other long-term carriage isolates [14] and is a general characteristic of prokaryotic evolution. Recent work suggests that such decay is not an adaptive mechanism towards genome streamlining and higher replication efficiency, but rather due to an intrinsic bias towards deletion [21, 22]. Large genomes in free-living species are maintained by the balance of addition of new genes through LGT and deletion. Once inside a host, opportunities for LGT are limited, resulting in a neutral ratchet-like loss of genes. Such loss can be expected to be due to drift as a small effective population size makes selection highly inefficient. While some of the gene losses observed in P314 may be adaptive, the majority are likely to be driven by drift, despite the fact that these same genes may be lost in other data sets (Fig 5).

Discussion

Bacterial evolution in the natural environment is usually only witnessed through endpoint observations, which greatly limit our understanding of the process. For example, when a pathogen population emerges and moves from a reservoir to a susceptible host, a single such ecological jump likely involves a series of sequential adaptive, neutral, and deleterious steps as well as sublineage extinction, but only the emerged successful pathogen population is available for analysis. Endpoint comparisons to understand this process reveal only the mutations that persist in the population in the end state. The nature and extent of ephemeral mutations are lost and the temporal order of many mutations cannot be reconstructed. In this study, we have had the opportunity to sample and characterize a pathogen population in a chronic case of melioidosis across >16 years. We find that the pathogen persisted in the face of therapeutics, evolved, and propagated in a strictly clonal fashion, lost virulence, and still exists in its human host, likely as a commensal.

Chronic infections with *B. pseudomallei* are uncommon but almost always curable, and this patient's original pulmonary disease (non-cystic fibrosis bronchiectasis) is very likely to have contributed to colonization and persistence. Fortuitously, and despite high environmental diversity of *B. pseudomallei* in this region [83], sampling around the patient's home revealed extremely closely related isolates, providing a surrogate for the *bona fide* infective inoculum. This reference material allows for precise estimations of the likely ancestral genotype and subsequent *in vivo* evolution. While the patient was initially clinically unwell, the disease pathology progressively lessened to the point that she now has essentially no symptoms that can be attributed to her *B. pseudomallei* infection. Her daily sputum production is likely due to her moderately severe bronchiectasis and intermittent exacerbation of a *P. aeruginosa* infection. Yet she still harbors a non-pathogenic *B. pseudomallei* population in her respiratory tract. We

have documented sequential genomic and phenotypic changes in the pathogen that are associated with its conversion from a pathogen towards a commensal.

Phylogenetic reconstruction of the population's evolutionary trajectory reveals a highly structured topology where accumulated mutations both persisted and disappeared from the population. The ancestral genotype was rapidly purged from the population and existing genotypes increasingly diverged from the ancestral type. After >16 years, only two lineages (that began to appear in the 4th year), persisted. A neutral evolutionary model, driven primarily by drift rather than selection, best explains the overall changes in this population. This result is consistent with phylogenetic patterns: the purging of a fraction of the continuously diversifying population, a robust molecular clock, and the low homoplasy content in the character set as parallel evolution of adaptive phenotypes or reversion of deleterious mutations will increase homoplasy. An overall neutral model of evolution is also consistent with the change from a free-living to long-term intracellular lifestyle. With a bias towards genomic deletions and the inability to gain genetic content through lateral gene transfer due to the isolation of this population from conspecifics, a ratchet-like model of gene loss is expected. This change in lifestyle will also be accompanied by a massive reduction in the effective population size, causing selection to be less efficient and drift to play a major evolutionary role. An overall neutral model does not exclude the possible role of selection in the emergence of adaptive changes. Indeed, convergence of some of the large deleted regions provides evidence for selection for the loss of genes in these region. We can also assume that over 16 years, selective pressure has increased the frequency of phenotypic changes that appear to be adaptive: the loss of the O-antigen, antibiotic resistance, the decrease in CPS, and the non-expression of flagellin proteins. Likewise, the non-ribosomal peptide synthetase (NRPS) partial deletion, observed in all patient isolates, was rapidly fixed in the population consistent with strong selective pressure due to the likely link to a loss of virulence in this genetic background. Strong selection against virulence has been documented in experimental infections of *Caenorhabditis elegans* with *P. aeruginosa*, resulting in the evolutionary transition from pathogen to commensal [84].

Phenotypic changes are observable in even the earliest clinical isolates, as demonstrated through the comparison of early clinical isolates to environment isolates from the patient's home. The environmental isolates have nearly identical genomes and are separated from the clinical isolates in the phylogeny by as few as zero core genome SNPs and yet have different phenotypic characteristics. While virulence in animals is a complex phenotype, the contrast between the environmental isolates and those from the patient is profound. In the mouse model, the environmental isolates are universally lethal with an exceedingly low LD50 of ~4 CFU (Fig 6A and 6B). In contrast, the clinical isolates all failed to kill at doses >1,000 fold higher (Fig 6A). Our reference to early clinical isolates must be kept in context as we predict that the patient was infected ~1 year prior to clinical presentation and collection of our first isolate (Fig 10). However, from our phylogenetic reconstruction, we can discern the genomic changes and the order in which they occurred in this time period. While diversity was rapidly generated, a deletion in a non-ribosomal peptide synthetase (NRPS) gene was the single change shared by all clinical isolates (Fig 2), indicating that this was the first of all observed mutations. The exact role that this complex enzyme plays is unclear, but it is the sole mutation correlated with early decreased virulence.

Other phenotypic changes that could be related to adaptation continued to occur throughout the >16-year timeline. O-antigen is a known virulence factor and a major antigen that triggers intense host responses. We have previously shown that some of these clinical isolates do not produce this antigen and that this is associated with a mutation in the *wbiI* gene [66] (S7 Table). The capsular polysaccharide (CPS) production is greatly reduced in the clinical isolates, though not totally absent (S8 Table). Motility of the isolates is reduced and even eliminated in

some clinical isolates, although the genomic basis for this loss is not obvious. The gross colony morphology changes over time, with some of the most recent clinical isolates having very small colonies and slow growth. And finally, given that morbidity in the mouse model can be estimated by animal weights following infection, we observe that later clinical isolates cause less morbidity than earlier ones (Fig 6C). Despite an overall neutral evolutionary model, these examples provide evidence of continuing adaptive evolution even in the later years of this chronic persistence. Some phenotypic changes can be rationalized as avoidance of the host immune response (e.g., O-antigen, CPS, and flagellin antigens). Commensal bacteria are able to co-exist with their hosts in part by not stimulating the inflammatory responses, unlike many pathogens [85]. However, we have little understanding of this niche and its complexity, which limits our ability to understand the full extent and repertoire of selective pressures involved in this adaptation.

The maintenance of two long-lived lineages

While lineage extinction is very common in this population, in year 4, two lineages emerged and persisted for at least 12 years (Figs 3 and 10). Even within these two lineages, extinction continued to occur, consistent with drift and stochastic evolution. Relatively large genomic differences exist between these lineages, including differential genome deletion patterns: long-lived lineage 1 is missing region 6b, while long-lived lineage 2 has lost regions 2b and 7 (Figs 5A and 2). While genetic drift may be the dominant driver of phylogenetic characters along these two lineages, some changes are likely adaptive. For example, phenotypic changes in long-lived lineage 2 include very small colony morphology and slower growth. Small colony variants of *P. aeruginosa* and *S. aureus* in chronic lung infections are thought to be associated with persistence [86, 87]. Given these phenotypic and genotypic differences, how and why do members of both lineages persist? A strictly neutral model, would suggest that neither lineage is better adapted and their fitness is dependent on drift. However, the likely adaptive nature of phenotypic differences warrants the exploration of other possibilities. Perhaps we are seeing spatial segregation and adaptation. Spatial structuring of *P. aeruginosa* and other pathogens has been shown in some CF lungs [11, 80], but our limited data from year 4 does not support an allopatric model; isolates from extensive pulmonary sampling associated with the lobectomy showed evidence of genotype mixing across pulmonary space and no observable spatial differentiation at a macro-scale (Fig 6). However, many newly emerged lineages went extinct during year 4, with only these two long-lived lineages persisting. The two evolutionary paths towards commensalism mostly occurred in the following 12 years and could involve spatial differentiation in the respiratory tract that occurred at some point since the lobectomy samples were collected. However, with functional mucociliary clearance mechanisms and aspiration, bacterial movement within the respiratory tract is likely, limiting the potential for spatial segregation.

An alternative to the neutral and spatial segregation hypotheses is a shared niche with differential selective pressures. For example, temporal shifts in host/microbe interactions may vary selective pressures in the same space, allowing for the maintenance of different lineages. Frequency dependent selection due to different host interactions and avoidance abilities can also result in balancing selection to maintain diversity in sympatry. In contrast to competition between lineages, cooperation may also result in balancing selection if both lineages possess different sets of competitive advantages that can be mutually exploited. For example, in nutrient poor environments, differential production of extracellular small metabolites such as siderophores can provide nutrients to conspecifics that would not otherwise be able to obtain them [88]. In a clonally reproducing population, the maintenance of both genotypes could be critical to survival as selective sweeps and background selection will frequently purge alleles

(including beneficial ones). Because we have isolates from each lineage, it should be possible to test hypotheses concerning adaptation to particular environmental variables (e.g., oxygen levels). Additionally, PCR analyses of future sputum samples will enable us to monitor the ratios of the different lineages over time.

B. pseudomallei in P314 is evolving convergently with *B. mallei*

Similar to this population, *B. mallei* is a strictly clonal pathogen derived from a *B. pseudomallei* ancestor. It is also host restricted (found primarily in equines), and its evolution is characterized by genome decay [65]. In addition to these striking general similarities, even similar genomic deletions are shared by the P314 isolates and *B. mallei* (Fig 5). *B. mallei* is also manifested by chronic respiratory infections and passed from animal to animal with only a transitory environmental phase. Because *B. mallei* is a clonal derivation from *B. pseudomallei*, it is postulated that an equine host contracted melioidosis, which then evolved into a chronic respiratory disease. Equine pulmonary melioidosis may have occurred many times, but only once did it result in all of the extant *B. mallei* isolates. For our chronic patient's isolates, we can only speculate on the likelihood of continued evolution, perhaps through recombination, that may occur if some members of this population were introduced into the environment (via expectoration). Such convergent evolutionary patterns of evolution likely represent both neutral and adaptive processes that can lead to populations with very different phenotypes and ecological strategies. None the less, this human chronic infection provides a single rare window into the evolutionary pathways of *B. mallei* and other potentially deadly pathogens or commensals.

Supporting information

S1 Fig. Core genome relationships among ST131 isolates. SNPs (17,359) from core genomic regions were identified and then used to construct a maximum parsimony tree of 11 ST131 isolates' genetic relationships. K96243 was used as an outgroup for rooting. The two strains (MSHR1043 and MSHR6686) isolated from patient sputum clustered tightly with the six environmental isolates from the patient's residence. The P314 reference genome is the closed genome of the environmental MSHR1435. The CI of this phylogenetic reconstruction was 0.96, which is indicative of a highly clonal mode of genetic inheritance. Over 16,000 SNPs separate the P314 cluster (clinical and associated environmental isolates) from the next nearest ST131 strains.
(PDF)

S2 Fig. Tree for branch labels. Maximum parsimony tree from Fig 3 with branch labels. S2 Table lists all polymorphisms and includes information on the branch where the mutation occurred.
(PDF)

S3 Fig. Morbidity of P314 isolates and P314-associated environmental isolates. A-D) BALB/c mice were intranasally inoculated with *B. pseudomallei* isolates (n = 5/group). Mice were weighed and clinically scored daily. A) The average change ± standard deviation in weight of the group over time is shown for the initial 6 day period. B) The average clinical score ± standard deviation of the group over time is shown for the initial 6 day period. An increase in clinical score indicate an accumulation of clinical symptoms or an increase in morbidity. C) Individual clinical scores at day 2 are shown by dots (n = 5/group) and the mean score is shown by the bar. The isolates are listed by the relatedness to the environmental isolate cluster (right to left).
(PDF)

S4 Fig. Distances from the phylogenetic root to each genome. Similar slopes for isolates phylogenetically located before the divergence of the two long-lived lineages and those of the long-lived lineage 2 suggests similar rates of evolution. A smaller slope for isolates from the long-lived lineage 1 suggests a lower evolutionary rate, however relatively large root-to-tip distances for the earliest long-lived lineage 1 isolates may indicate a temporary jump in evolutionary rate. Extrapolation of root to tip distances (excluding the long lived-lineage 1 isolates) suggests divergence from a common environmental isolate ~308 days before presenting at the hospital. This is a slightly longer estimate than provided by BEAST (Fig 9) which also utilizes invariant sites to estimate mutation rates.
(PDF)

S5 Fig. Subsampled skygrid. Downsampling of isolates collected at the time of the lobectomy show a similar estimation of effective population size compared to the full dataset.
(PDF)

S1 Table. Genome details.
(XLSX)

S2 Table. Phylogeny matrix.
(XLSX)

S3 Table. Details of large deletions.
(XLSX)

S4 Table. Generalized stepping stone analysis.
(XLSX)

S5 Table. *B. mallei* accession numbers.
(XLSX)

S6 Table. BSR.
(XLSX)

S7 Table. Small deletions.
(XLSX)

S8 Table. Phenotype data.
(XLSX)

S9 Table. Virulence gene screen.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Apichai Tuanyok, Bart J. Currie, Paul Keim.

Data curation: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Karthik Handady, Heidie Hornstra, Adam J. Vazquez, Erik Settles, Mark Mayo, Mirjam Kaestli, Charles H. D.

Williamson, Erin P. Price, Derek S. Sarovich, Jeffrey T. Foster, Kevin P. Drees, Timothy J. Kidd, Scott C. Bell, Bart J. Currie.

Formal analysis: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Karthik Handady, Heidie Hornstra, Adam J. Vazquez, Erik Settles, Charles H. D. Williamson, Erin P. Price, Derek S. Sarovich, Richard A. Bowen.

Funding acquisition: Talima Pearson, Apichai Tuanyok, Bart J. Currie, Paul Keim.

Investigation: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Erik Settles, Erin P. Price, Derek S. Sarovich, Richard A. Bowen, Apichai Tuanyok, Bart J. Currie, Paul Keim.

Methodology: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Erik Settles, Charles H. D. Williamson, Derek S. Sarovich, James M. Cook, Spenser R. Wolken, Richard A. Bowen, Apichai Tuanyok, Kevin P. Drees, Bart J. Currie, Paul Keim.

Project administration: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Heidie Hornstra, Erik Settles, Mark Mayo, Mirjam Kaestli, Apichai Tuanyok, Bart J. Currie, Paul Keim.

Resources: Talima Pearson, Timothy J. Kidd, Scott C. Bell, Bart J. Currie, Paul Keim.

Supervision: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Mark Mayo, Mirjam Kaestli, Charles H. D. Williamson, Jeffrey T. Foster, Bart J. Currie, Paul Keim.

Visualization: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Erik Settles.

Writing – original draft: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Erik Settles, Bart J. Currie, Paul Keim.

Writing – review & editing: Talima Pearson, Jason W. Sahl, Crystal M. Hepp, Karthik Handady, Heidie Hornstra, Adam J. Vazquez, Erik Settles, Mark Mayo, Mirjam Kaestli, Charles H. D. Williamson, Erin P. Price, Derek S. Sarovich, James M. Cook, Spenser R. Wolken, Richard A. Bowen, Apichai Tuanyok, Jeffrey T. Foster, Kevin P. Drees, Timothy J. Kidd, Scott C. Bell, Bart J. Currie, Paul Keim.

References

1. Finlay BB, McFadden G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. Cell. 2006; 124(4):767–82. Epub 2006/02/25. https://doi.org/10.1016/j.cell.2006.01.034 PMID: 16497587.

2. Cvijovic I, Good BH, Jerison ER, Desai MM. Fate of a mutation in a fluctuating environment. Proc Natl Acad Sci U S A. 2015; 112(36):E5021–8. https://doi.org/10.1073/pnas.1505406112 PMID: 26305937.

3. Chewapreecha C, Holden MT, Vehkala M, Valimaki N, Yang Z, Harris SR, et al. Global and regional dissemination and evolution of Burkholderia pseudomallei. Nat Microbiol. 2017; 2:16263. Epub 2017/01/24. https://doi.org/10.1038/nmicrobiol.2016.263 PMID: 28112723.

4. Currie BJ, Dance DA, Cheng AC. The global distribution of Burkholderia pseudomallei and melioidosis: an update. Trans R Soc Trop Med Hyg. 2008; 102 Suppl 1:S1–4. Epub 2009/01/16. https://doi.org/10.1016/S0035-9203(08)70002-6 PMID: 19121666.

5. Limmathurotsakul D, Golding N, Dance DA, Messina JP, Pigott DM, Moyes CL, et al. Predicted global distribution of Burkholderia pseudomallei and burden of melioidosis. Nat Microbiol. 2016; 1(1). Epub 2016/02/16. https://doi.org/10.1038/nmicrobiol.2015.8 PMID: 26877885.

6. Limmathurotsakul D, Peacock SJ. Melioidosis: a clinical overview. British medical bulletin. 2011; 99:125–39. https://doi.org/10.1093/bmb/ldr007 PMID: 21558159.

7. Wiersinga WJ, Currie BJ, Peacock SJ. Melioidosis. The New England journal of medicine. 2012; 367 (11):1035–44. https://doi.org/10.1056/NEJMra1204699 PMID: 22970946.

8. Currie BJ, Ward L, Cheng AC. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. PLoS Negl Trop Dis. 2010; 4(11):e900. Epub 2010/12/15. https://doi.org/10.1371/journal.pntd.0000900 PMID: 21152057.

9. Bartell JA, Sommer LM, Haagensen JAJ, Loch A, Espinosa R, Molin S, et al. Evolutionary highways to persistent bacterial infection. Nat Commun. 2019; 10(1):629. https://doi.org/10.1038/s41467-019-08504-7 PMID: 30733448.

10. Goss CH, Muhlebach MS. Review: Staphylococcus aureus and MRSA in cystic fibrosis. J Cyst Fibros. 2011; 10(5):298–306. https://doi.org/10.1016/j.jcf.2011.06.002 PMID: 21719362.

11. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, et al. Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. Cell Host Microbe. 2015; 18(3):307–19. https://doi.org/10.1016/j.chom.2015.07.006 PMID: 26299432.

12. Geake JB, Reid DW, Currie BJ, Bell SC, Melioid CFI, Bright-Thomas R, et al. An international, multicentre evaluation and description of Burkholderia pseudomallei infection in cystic fibrosis. BMC Pulm Med. 2015; 15:116. Epub 2015/10/11. https://doi.org/10.1186/s12890-015-0109-9 PMID: 26453341.

13. Price EP, Viberg LT, Kidd TJ, Bell SC, Currie BJ, Sarovich DS. Transcriptomic analysis of longitudinal Burkholderia pseudomallei infecting the cystic fibrosis lung. Microb Genom. 2018; 4(8). Epub 2018/07/11. https://doi.org/10.1099/mgen.0.000194 PMID: 29989529.

14. Viberg LT, Sarovich DS, Kidd TJ, Geake JB, Bell SC, Currie BJ, et al. Within-Host Evolution of Burkholderia pseudomallei during Chronic Infection of Seven Australasian Cystic Fibrosis Patients. MBio. 2017; 8(2). Epub 2017/04/13. https://doi.org/10.1128/mBio.00356-17 PMID: 28400528.

15. Meumann EM, Cheng AC, Ward L, Currie BJ. Clinical features and epidemiology of melioidosis pneumonia: results from a 21-year study and review of the literature. Clin Infect Dis. 2012; 54(3):362–9. Epub 2011/11/08. https://doi.org/10.1093/cid/cir808 PMID: 22057702.

16. Yeager JJ, Facemire P, Dabisch PA, Robinson CG, Nyakiti D, Beck K, et al. Natural history of inhalation melioidosis in rhesus macaques (Macaca mulatta) and African green monkeys (Chlorocebus aethiops). Infect Immun. 2012; 80(9):3332–40. Epub 2012/07/11. https://doi.org/10.1128/IAI.00675-12 PMID: 22778104.

17. Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, Burkholderia pseudomallei. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(39):14240–5. Epub 2004/09/21. https://doi.org/10.1073/pnas.0403302101 PMID: 15377794.

18. Spring-Pearson SM, Stone JK, Doyle A, Allender CJ, Okinaka RT, Mayo M, et al. Pangenome Analysis of Burkholderia pseudomallei: Genome Evolution Preserves Gene Order despite High Recombination Rates. PLoS One. 2015; 10(10):e0140274. Epub 2015/10/21. https://doi.org/10.1371/journal.pone.0140274 PMID: 26484663.

19. Tuanyok A, Leadem BR, Auerbach RK, Beckstrom-Sternberg SM, Beckstrom-Sternberg JS, Mayo M, et al. Genomic islands from five strains of Burkholderia pseudomallei. BMC Genomics. 2008; 9:566. Epub 2008/11/29. https://doi.org/10.1186/1471-2164-9-566 PMID: 19038032.

20. Dillon MM, Sung W, Lynch M, Cooper VS. The Rate and Molecular Spectrum of Spontaneous Mutations in the GC-Rich Multichromosome Genome of Burkholderia cenocepacia. Genetics. 2015; 200 (3):935–46. Epub 2015/05/15. https://doi.org/10.1534/genetics.115.176834 PMID: 25971664.

21. Kuo CH, Ochman H. Deletional bias across the three domains of life. Genome Biol Evol. 2009; 1:145–52. Epub 2009/01/01. https://doi.org/10.1093/gbe/evp016 PMID: 20333185.

22. Sela I, Wolf YI, Koonin EV. Theory of prokaryotic genome evolution. Proc Natl Acad Sci U S A. 2016; 113(41):11399–407. Epub 2016/10/06. https://doi.org/10.1073/pnas.1614083113 PMID: 27702904.

23. Stone JK, DeShazer D, Brett PJ, Burtnick MN. Melioidosis: molecular aspects of pathogenesis. Expert Rev Anti Infect Ther. 2014; 12(12):1487–99. Epub 2014/10/15. https://doi.org/10.1586/14787210.2014.970634 PMID: 25312349.

24. Sarovich DS, Price EP, Webb JR, Ward LM, Voutsinos MY, Tuanyok A, et al. Variable virulence factors in Burkholderia pseudomallei (melioidosis) associated with human disease. PLoS One. 2014; 9(3): e91682. Epub 2014/03/13. https://doi.org/10.1371/journal.pone.0091682 PMID: 24618705.

25. Price EP, Sarovich DS, Mayo M, Tuanyok A, Drees KP, Kaestli M, et al. Within-host evolution of Burkholderia pseudomallei over a twelve-year chronic carriage infection. MBio. 2013; 4(4). Epub 2013/07/19. https://doi.org/10.1128/mBio.00388-13 PMID: 23860767.

26. Pearson T, U'Ren JM, Schupp JM, Allan GJ, Foster PG, Mayo MJ, et al. VNTR analysis of selected outbreaks of Burkholderia pseudomallei in Australia. Infect Genet Evol. 2007; 7(4):416–23. Epub 2007/01/30. https://doi.org/10.1016/j.meegid.2006.12.002 PMID: 17258514.

27. Ashdown LR. An improved screening technique for isolation of Pseudomonas pseudomallei from clinical specimens. Pathology. 1979; 11(2):293–7. https://doi.org/10.3109/00313027909061954 PMID: 460953.

28. Novak RT, Glass MB, Gee JE, Gal D, Mayo MJ, Currie BJ, et al. Development and evaluation of a real-time PCR assay targeting the type III secretion system of Burkholderia pseudomallei. J Clin Microbiol. 2006; 44(1):85–90. Epub 2006/01/05. https://doi.org/10.1128/JCM.44.1.85-90.2006 PMID: 16390953.

29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology: a journal of computational molecular cell biology. 2012; 19(5):455–77. Epub 2012/04/18. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599.

30. Sahl JW, Mayo M, Price EP, Sarovich DS, Kaestli M, Pearson T, et al. Complete Genome Sequence of the Environmental Burkholderia pseudomallei Sequence Type 131 Isolate MSHR1435, Associated with a Chronic Melioidosis Infection. Genome Announc. 2018; 6(11). Epub 2018/03/17. https://doi.org/10.1128/genomeA.00072-18 PMID: 29545292.

31. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXivorg. 2013;(arXiv:1303.3997 [q-bio.GN]).

32. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. Epub 2011/04/12. https://doi.org/10.1038/ng.806 PMID: 21478889.

33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. Epub 2010/07/21. https://doi.org/10.1101/gr.107524.110 PMID: 20644199.

34. Delcher AL, Phillippy A, Carlton J, Salzberg SL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 2002; 30(11):2478–83. Epub 2002/05/30. https://doi.org/10.1093/nar/30.11.2478 PMID: 12034836.

35. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999; 27 (2):573–80. Epub 1998/12/24. https://doi.org/10.1093/nar/27.2.573 PMID: 9862982.

36. Sahl JW, Lemmer D, Travis J, Schupp JM, Gillece JD, Aziz M, et al. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microbial Genomics. 2016. https://doi.org/10.1099/mgen.0.000074 PMID: 28348869

37. Sahl JW, Beckstrom-Sternberg JS, Babic-Sternberg J, Gillece JD, Hepp CM, Auerbach RK, et al. The In Silico Genotyper (ISG): an open-source pipeline to rapidly identiy and annotate nucleotide variants for comparative genomics applications. bioRxiv. 2015. https://doi.org/10.1101/015578.

38. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012; 6(2):80–92. Epub 2012/06/26. https://doi.org/10.4161/fly.19695 PMID: 22728672.

39. Swofford DL. Phylogenetic Analysis Using Parsimony *. 2002.

40. Novacek MJ. Fossils, topologies, missing data, and the higher level phylogeny of Eutherian mammals. Systematic Biology. 1992; 41(1):58–73.

41. Simmons MP, Donovan Bailey C, Nixon KC. Phylogeny reconstruction using duplicate genes. Mol Biol Evol. 2000; 17(4):469–73. Epub 2000/03/31. https://doi.org/10.1093/oxfordjournals.molbev.a026327 PMID: 10742039.

42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. Epub 2009/06/10. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943.

43. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19(9):1639–45. Epub 2009/06/23. https://doi.org/10.1101/gr.092759.109 PMID: 19541911.

44. Sahl JW, Caporaso JG, Rasko DA, Keim P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2014; 2:e332. https://doi.org/10.7717/peerj.332 PMID: 24749011

45. Kent WJ. BLAT—the BLAST-like alignment tool. Genome Res. 2002; 12(4):656–64. Epub 2002/04/05. https://doi.org/10.1101/gr.229202 Article published online before March 2002. PMID: 11932250.

46. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012; 29(8):1969–73. Epub 2012/03/01. https://doi.org/10.1093/molbev/mss075 PMID: 22367748.

47. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33(7):1870–4. Epub 2016/03/24. https://doi.org/10.1093/molbev/msw054 PMID: 27004904.

48. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 1985; 22(2):160–74. Epub 1985/01/01. https://doi.org/10.1007/bf02101694 PMID: 3934395.

49. Baele G, Lemey P, Suchard MA. Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. Syst Biol. 2016; 65(2):250–64. Epub 2015/11/04. https://doi.org/10.1093/sysbio/syv083 PMID: 26526428.

50. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018; 67(5):901–4. Epub 2018/05/03. https://doi.org/10.1093/sysbio/syy032 PMID: 29718447.

51. Rambaut A. FigTree v. 1.4.0 2012. http://tree.bio.ed.ac.uk/software/figtree.

52. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005; 22(5):1208–22. Epub 2005/02/11. https://doi.org/10.1093/molbev/msi105 PMID: 15703242.

53. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986; 3(5):418–26. Epub 1986/09/01. https://doi.org/10.1093/oxfordjournals.molbev.a040410 PMID: 3444411.

54. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123(3):585–95. Epub 1989/11/01. PMID: 2513255.

55. Haiko J, Westerlund-Wikstrom B. The role of the bacterial flagellum in adhesion and virulence. Biology (Basel). 2013; 2(4):1242–67. Epub 2013/01/01. https://doi.org/10.3390/biology2041242 PMID: 24833223.

56. Shields P, Cathcart L. Motility test medium protocol. American society for microbiology, 2011.

57. Cuccui J, Milne TS, Harmer N, George AJ, Harding SV, Dean RE, et al. Characterization of the Burkholderia pseudomallei K96243 capsular polysaccharide I coding region. Infect Immun. 2012; 80(3):1209–21. Epub 2012/01/19. https://doi.org/10.1128/IAI.05805-11 PMID: 22252864.

58. Houghton RL, Reed DE, Hubbard MA, Dillon MJ, Chen H, Currie BJ, et al. Development of a prototype lateral flow immunoassay (LFI) for the rapid diagnosis of melioidosis. PLoS Negl Trop Dis. 2014; 8(3): e2727. Epub 2014/03/22. https://doi.org/10.1371/journal.pntd.0002727 PMID: 24651568.

59. AuCoin DP, Reed DE, Marlenee NL, Bowen RA, Thorkildson P, Judy BM, et al. Polysaccharide specific monoclonal antibodies provide passive protection against intranasal challenge with Burkholderia pseudomallei. PLoS One. 2012; 7(4):e35386. https://doi.org/10.1371/journal.pone.0035386 PMID: 22530013.

60. Stewart JD, Smith S, Binotto E, McBride WJ, Currie BJ, Hanson J. The epidemiology and clinical features of melioidosis in Far North Queensland: Implications for patient management. PLoS Negl Trop Dis. 2017; 11(3):e0005411. Epub 2017/03/07. https://doi.org/10.1371/journal.pntd.0005411 PMID: 28264029.

61. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. Proc Natl Acad Sci U S A. 2014; 111(25):9259–64. Epub 2014/06/14. https://doi.org/10.1073/pnas.1401734111 PMID: 24927540.

62. Vander Broek CW, Stevens JM. Type III Secretion in the Melioidosis Pathogen Burkholderia pseudomallei. Front Cell Infect Microbiol. 2017; 7:255. Epub 2017/07/01. https://doi.org/10.3389/fcimb.2017.00255 PMID: 28664152.

63. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011; 39(Web Server issue):W339–46. Epub 2011/06/16. https://doi.org/10.1093/nar/gkr466 PMID: 21672958.

64. Whitlock GC, Estes DM, Torres AG. Glanders: off to the races with Burkholderia mallei. FEMS Microbiol Lett. 2007; 277(2):115–22. https://doi.org/10.1111/j.1574-6968.2007.00949.x PMID: 18031330.

65. Losada L, Ronning CM, DeShazer D, Woods D, Fedorova N, Kim HS, et al. Continuing evolution of Burkholderia mallei through genome reduction and large-scale rearrangements. Genome Biol Evol. 2010; 2:102–16. Epub 2010/03/25. https://doi.org/10.1093/gbe/evq003 PMID: 20333227.

66. Tuanyok A, Stone JK, Mayo M, Kaestli M, Gruendike J, Georgia S, et al. The genetic and molecular basis of O-antigenic diversity in Burkholderia pseudomallei lipopolysaccharide. PLoS Negl Trop Dis. 2012; 6(1):e1453. Epub 2012/01/12. https://doi.org/10.1371/journal.pntd.0001453 PMID: 22235357.

67. Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, U'Ren JM, et al. Bacillus anthracis in China and its relationship to worldwide lineages. BMC Microbiol. 2009; 9:71. Epub 2009/04/17. https://doi.org/10.1186/1471-2180-9-71 PMID: 19368722.

68. Litvintseva AP, Marsden-Haug N, Hurst S, Hill H, Gade L, Driebe EM, et al. Valley fever: finding new places for an old disease: Coccidioides immitis found in Washington State soil associated with recent

human infection. Clin Infect Dis. 2015; 60(1):e1–3. Epub 2014/08/29. https://doi.org/10.1093/cid/ciu681 PMID: 25165087.

69. Fleshman A, Mullins K, Sahl J, Hepp C, Nieto N, Wiggins K, et al. Comparative pan-genomic analyses of Orientia tsutsugamushi reveal an exceptional model of bacterial evolution driving genomic diversity. Microb Genom. 2018; 4(9). https://doi.org/10.1099/mgen.0.000199 PMID: 30035711.

70. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, et al. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC biology. 2009; 7:78. https://doi.org/10.1186/1741-7007-7-78 PMID: 19922616.

71. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, et al. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proc Natl Acad Sci U S A. 2004; 101(37):13536–41. Epub 2004/09/07. https://doi.org/10.1073/pnas.0403844101 PMID: 15347815.

72. Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, et al. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biol. 2009; 7:78. Epub 2009/11/20. https://doi.org/10.1186/1741-7007-7-78 PMID: 19922616.

73. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, et al. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:13536–41. https://doi.org/10.1073/pnas.0403844101 PMID: 15347815.

74. Pearson T, Hornstra HM, Sahl JW, Schaack S, Schupp JM, Beckstrom-Sternberg SM, et al. When outgroups fail; phylogenomics of rooting the emerging pathogen, Coxiella burnetii. Systematic biology. 2013; 62:752–62. https://doi.org/10.1093/sysbio/syt038 PMID: 23736103.

75. Pearson T, Okinaka RT, Foster JT, Keim P. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infection, Genetics and Evolution. 2009; 9:1010–9. https://doi.org/10.1016/j.meegid.2009.05.014 PMID: 19477301.

76. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution. 1985; 39(4):783. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x PMID: 28561359

77. Archie JW. Measures of homoplasy. In: Sanderson MJ, Hufford L, editors. Homoplasy: The recurrence of similarity in evolution. San Diego: Academic Press; 1996. p. 153–88.

78. Limmathurotsakul D, Wuthiekanun V, Chantratita N, Wongsuvan G, Thanwisai A, Biaklang M, et al. Simultaneous infection with more than one strain of Burkholderia pseudomallei is uncommon in human melioidosis. Journal of Clinical Microbiology. 2007; 45(11):3830–2. https://doi.org/10.1128/JCM.01297-07 PMID: 17804651

79. Price EP, Sarovich DS, Viberg L, Mayo M, Kaestli M, Tuanyok A, et al. Whole-genome sequencing of Burkholderia pseudomallei isolates from an unusual melioidosis case identifies a polyclonal infection with the same multilocus sequence type. J Clin Microbiol. 2015; 53(1):282–6. Epub 2014/10/24. https://doi.org/10.1128/JCM.02560-14 PMID: 25339397.

80. Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR Jr., et al. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet. 2011; 43 (12):1275–80. https://doi.org/10.1038/ng.997 PMID: 22081229.

81. Sahl JW, Allender CJ, Colman RE, Califf KJ, Schupp JM, Currie BJ, et al. Genomic characterization of Burkholderia pseudomallei isolates selected for medical countermeasures testing: comparative genomics associated with differential virulence. PLoS One. 2015; 10(3):e0121052. Epub 2015/03/25. https://doi.org/10.1371/journal.pone.0121052 PMID: 25803742.

82. Silva IN, Santos PM, Santos MR, Zlosnik JE, Speert DP, Buskirk SW, et al. Long-Term Evolution of Burkholderia multivorans during a Chronic Cystic Fibrosis Infection Reveals Shifting Forces of Selection. mSystems. 2016; 1(3). https://doi.org/10.1128/mSystems.00029-16 PMID: 27822534.

83. McRobb E, Kaestli M, Price EP, Sarovich DS, Mayo M, Warner J, et al. Distribution of Burkholderia pseudomallei in northern Australia, a land of diversity. Appl Environ Microbiol. 2014; 80(11):3463–8. Epub 2014/03/25. https://doi.org/10.1128/AEM.00128-14 PMID: 24657869.

84. Jansen G, Crummenerl LL, Gilbert F, Mohr T, Pfefferkorn R, Thanert R, et al. Evolutionary Transition from Pathogenicity to Commensalism: Global Regulator Mutations Mediate Fitness Gains through Virulence Attenuation. Mol Biol Evol. 2015; 32(11):2883–96. Epub 2015/07/23. https://doi.org/10.1093/molbev/msv160 PMID: 26199376.

85. Lebeer S, Vanderleyden J, De Keersmaecker SCJ. Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. Nature Reviews Microbiology. 2010; 8:171. https://doi.org/10.1038/nrmicro2297 PMID: 20157338

86. Kahl BC, Becker K, Loffler B. Clinical Significance and Pathogenesis of Staphylococcal Small Colony Variants in Persistent Infections. Clin Microbiol Rev. 2016; 29(2):401–27. Epub 2016/03/11. https://doi.org/10.1128/CMR.00069-15 PMID: 26960941.

87. Haussler S, Ziegler I, Lottel A, von Gotz F, Rohde M, Wehmhohner D, et al. Highly adherent small-colony variants of Pseudomonas aeruginosa in cystic fibrosis lung infection. J Med Microbiol. 2003; 52(Pt 4):295–301. Epub 2003/04/05. https://doi.org/10.1099/jmm.0.05069-0 PMID: 12676867.

88. Buckling A, Harrison F, Vos M, Brockhurst MA, Gardner A, West SA, et al. Siderophore-mediated cooperation and virulence in Pseudomonas aeruginosa. FEMS Microbiol Ecol. 2007; 62(2):135–41. https://doi.org/10.1111/j.1574-6941.2007.00388.x PMID: 17919300.