Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*

Matthew T. G. Holden[a], Richard W. Titball[b,c], Sharon J. Peacock[d,e], Ana M. Cerdeño-Tárraga[a], Timothy Atkins[b], Lisa C. Crossman[a], Tyrone Pitt[f], Carol Churcher[a], Karen Mungall[a], Stephen D. Bentley[a], Mohammed Sebaihia[a], Nicholas R. Thomson[a], Nathalie Bason[a], Ifor R. Beacham[g], Karen Brooks[a], Katherine A. Brown[h], Nat F. Brown[g], Greg L. Challis[i], Inna Cherevach[a], Tracy Chillingworth[a], Ann Cronin[a], Ben Crossett[h], Paul Davis[a], David DeShazer[j], Theresa Feltwell[a], Audrey Fraser[a], Zahra Hance[a], Heidi Hauser[a], Simon Holroyd[a], Kay Jagels[a], Karen E. Keith[h], Mark Maddison[a], Sharon Moule[a], Claire Price[a], Michael A. Quail[a], Ester Rabbinowitsch[a], Kim Rutherford[a], Mandy Sanders[a], Mark Simmonds[a], Sirirurg Songsivilai[k], Kim Stevens[a], Sarinna Tumapa[e], Monkgol Vesaratchavest[e], Sally Whitehead[a], Corin Yeats[a], Bart G. Barrell[a], Petra C. F. Oyston[b], and Julian Parkhill[a,l]

[a]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; [b]Defence Science and Technology Laboratory, Porton Down, Salisbury SP4 0JQ, United Kingdom; [c]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; [d]Nuffield Department of Clinical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, United Kingdom; [e]Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand; [f]Laboratory of Hospital Infection, Division of Nosocomial Infection Prevention and Control, Central Public Health Laboratory, London NW9 5HT, United Kingdom; [g]School of Health Science, Griffith University, Gold Coast, Queensland 9726, Australia; [h]Department of Biological Sciences, Centre for Molecular Microbiology and Infection, Flowers Building, Imperial College, London SW7 2AZ, United Kingdom; [i]Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom; [j]U.S. Army Medical Research Institute for Infectious Diseases, 1425 Porter Street, Fort Detrick, MD 21702-5011; and [k]Department of Immunology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand

Edited by E. Peter Greenberg, University of Iowa, Iowa City, IA, and approved August 13, 2004 (received for review May 10, 2004)

Burkholderia pseudomallei **is a recognized biothreat agent and the causative agent of melioidosis. This Gram-negative bacterium exists as a soil saprophyte in melioidosis-endemic areas of the world and accounts for 20% of community-acquired septicaemias in northeastern Thailand where half of those affected die. Here we report the complete genome of *B. pseudomallei*, which is composed of two chromosomes of 4.07 megabase pairs and 3.17 megabase pairs, showing significant functional partitioning of genes between them. The large chromosome encodes many of the core functions associated with central metabolism and cell growth, whereas the small chromosome carries more accessory functions associated with adaptation and survival in different niches. Genomic comparisons with closely and more distantly related bacteria revealed a greater level of gene order conservation and a greater number of orthologous genes on the large chromosome, suggesting that the two replicons have distinct evolutionary origins. A striking feature of the genome was the presence of 16 genomic islands (GIs) that together made up 6.1% of the genome. Further analysis revealed these islands to be variably present in a collection of invasive and soil isolates but entirely absent from the clonally related organism *B. mallei*. We propose that variable horizontal gene acquisition by *B. pseudomallei* is an important feature of recent genetic evolution and that this has resulted in a genetically diverse pathogenic species.**

Melioidosis is a bacterial infection caused by *Burkholderia pseudomallei*, an environmental Gram-negative saprophyte present in wet soil and rice paddies in endemic areas (1–3). The majority of infections are reported from east Asia and northern Australia, the highest documented rate being in northeastern Thailand, where melioidosis accounts for 20% of all community-acquired septicaemias (4). Disease occurs after bacterial contamination of breaks in the skin or by inhalation after contact with water or soil. A pneumonic form of the disease can also result from the inhalation of contaminated dusts and was reported in U.S. helicopter pilots during the Vietnam War. The potential for the bacterium to cause disease after inhalation has also resulted in the inclusion of this pathogen on the Centers for Disease Control list of potential biothreat agents as a Category B agent (5). The most frequent clinical picture is a septicaemic illness associated with bacterial dissemination to distant sites, such that metastatic pneumonia and hepatic and splenic abcesses are common. However, clinical manifestations are protean and have led to the infection being termed "the great mimicker" (6). Of the cases in Thailand, one-fifth occur in children under the age of 14 years, for whom the overall mortality of infected individuals is 51% (3). Death usually occurs within the first 48 h as a result of septic shock and in a setting where optimal antimicrobial chemotherapy is given. Of equal concern, there is evidence that the bacterium does not cause overt disease in all individuals exposed to the bacterium but is able to persist at unknown sites in the body to become reactivated later in life. Possibly the best documented examples of this are in Vietnam veterans who developed disease, in one case 26 years later, after returning to the U.S. (7–9).

Here we describe the 7.25-megabase pair (Mb) genome of *B. pseudomallei* strain K96243, isolated from a case of human melioidosis. Comparative analysis highlights the role that horizontal gene acquisition has played in the evolution of the genome and clarifies the genetic relationship of *B. pseudomallei* with *Burkholderia mallei*, the genome of which is described in an accompanying manuscript (10). The work presented here also provides insights into the molecular basis of environmental survival, virulence, and antimicrobial resistance.

Materials and Methods

Bacterial Strain, Growth, and DNA Isolation. *B. pseudomallei* strain K96243 was isolated in 1996 from a 34-year-old female diabetic patient in Khon Kaen hospital in Thailand. K96243 is sensitive to imipenem, ceftazidime, chloramphenicol, ciprofloxacin, and augmentin and resistant to minocycline, gentamicin, cotrimoxazole, and streptomycin. The API 20NE profile of the bacterium was 1156576. The median lethal dose in Porton strain mice by the intraperitoneal route was 262 colony-forming units. Bacteria were cultured in L broth at 37°C for

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: GI, genomic island; CDS, coding sequence; Mb, megabase pair; IS, insertion sequence.

Data deposition: The sequences reported in this paper have been deposited in the European Molecular Biology Laboratory database (accession nos. BX571965 and BX571966).

[l]To whom correspondence should be addressed. E-mail: parkhill@sanger.ac.uk.

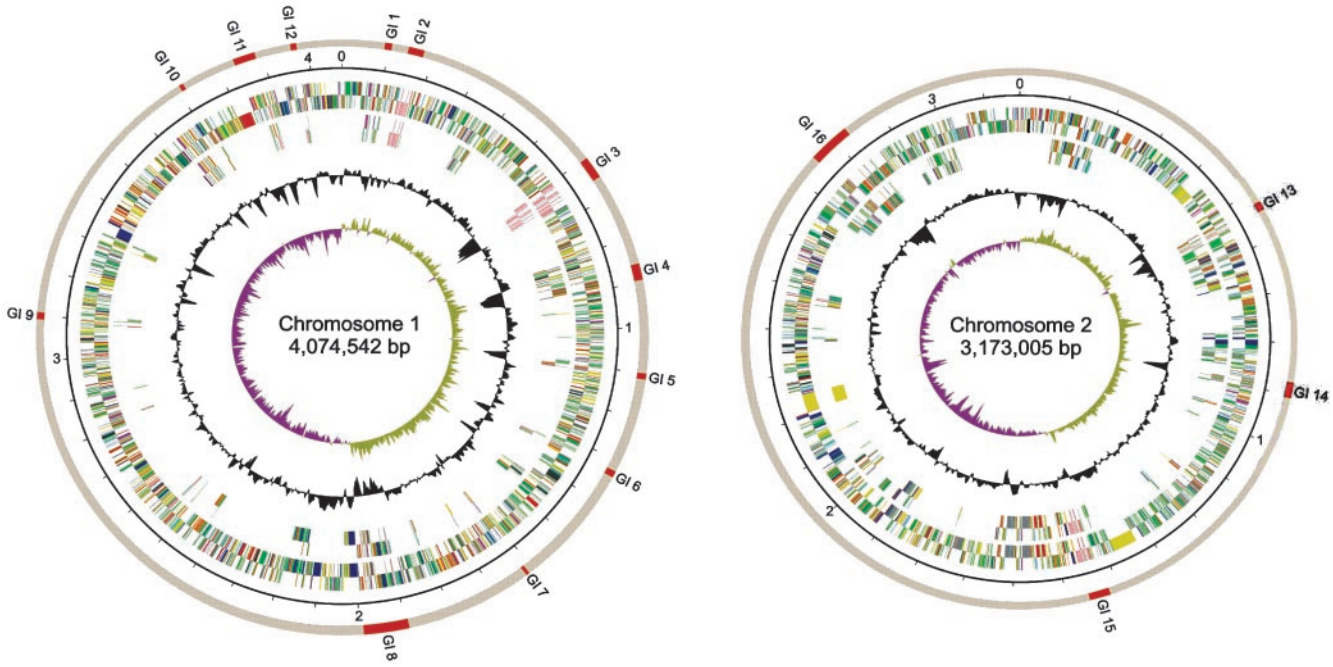

Fig. 1. Schematic circular diagrams of the large and small chromosomes of the *B. pseudomallei* genome. Where appropriate, categories are shown as pairs of concentric circles representing both coding strands. Rings from outside to inside: GIs represented by red segments; scale (in Mb); annotated CDSs colored according to predicted function; additional CDSs compared to the sequenced *B. mallei* strain ATCC 2944; the percentage of G + C content plot; (G − C)/(G + C) deviation plot (>0%, olive; <0%, purple). Color coding for CDSs: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, surface-associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central/intermediary metabolism; pale green, unknown; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage plus IS elements; gray, miscellaneous.

18 h and pelleted at 10,000 × *g*. The cells were resuspended in 30 ml of lysis solution (10 mM NaCl/20 mM Tris·HCl, pH 8.0/1 mM EDTA/0.5% SDS) and incubated at 50°C overnight. Three milliliters of 5 M sodium perchlorate was added and incubated for 1 h at ambient temperature. After phenol chloroform extraction, the DNA was precipitated with ethanol, spooled into deionized water, and stored at −20°C.

Whole-Genome Sequencing. The sequence was assembled, finished, and annotated as described in ref. 11 by using ARTEMIS to collate data and facilitate annotation (12); detailed information is available in *Supporting Text* and Tables 3–6, which are published as supporting information on the PNAS web site.

Comparative Genomics. Comparison of the genome sequences was facilitated by using the ACT (Artemis Comparison Tool) suite of programs (K.R., unpublished data; see also www.sanger.ac.uk/Software/ACT), which enabled the visualization of BLASTN and TBLASTX comparisons (13) between the genomes. Orthologous proteins were identified as reciprocal best matches by using FASTA with subsequent manual curation. Pseudogenes had one or more mutations that would prevent correct translation; each of the inactivating mutations was subsequently checked against the original sequencing data.

Multiplex PCR. The presence or absence of 11 GIs was defined for 40 *B. pseudomallei* isolates by using PCR. Twenty isolates were recovered from patients with melioidosis admitted to Sappasitiprasong Hospital (Ubon Ratchatani, Thailand) during 2001, and 20 isolates were recovered from the soil samples from the surrounding area. Target genes within the GIs were amplified by two multiplex reactions according to standard methodology (14, 15) after optimization by using *B. pseudomallei* K96243 genomic DNA. Detailed information, including the primer sequences and PCR conditions, is available in *Supporting Text*.

Results and Discussion

The complete genome of *B. pseudomallei* strain K96243 consists of two circular replicons (European Molecular Biology Laboratory accession nos. BX571965 and BX571966) of 4.07 Mb and 3.17 Mb (Fig. 1) each that have been designated chromosome 1 and chromosome 2 and encode 3,460 and 2,395 coding sequences (CDSs), respectively (for a summary of the features of the chromosomes, see Table 7, which is published as supporting information on the PNAS web site). A skew in strand-specific G/C content was seen for both chromosomes, which enabled the prediction of the origins of replication (Fig. 1). Chromosome 1 shares components near to the origin that are common to other prokaryotic genomes, such as *dnaA* and *dnaN* homologues. Chromosome 2 has a weaker G + C deviation pattern, and the predicted origin of replication contains features similar to those associated with plasmid replication, such as *parA* and *parB* homologues. Identification of CDSs on chromosome 2 that are involved in central metabolism and essential functions has led us to designate this component of the genome as a chromosome rather than a megaplasmid. A direct example of the distribution of essential features can be found in the tRNA genes on both chromosomes: 53 are encoded on chromosome 1, and 8 are encoded on chromosome 2. One of the eight tRNAs on chromosome 2 is unique to this replicon (Ser^{GGA}); however, genes encoding Ser by using this codon are found in equal proportions on both chromosomes (12% of Ser codons). In addition to the presence of tRNA genes, chromosome 2 also contains an rRNA gene cluster, components of the replication (*polA* and *dnaG*) and transcription machinery (*rpoD*), and components of central

MICROBIOLOGY

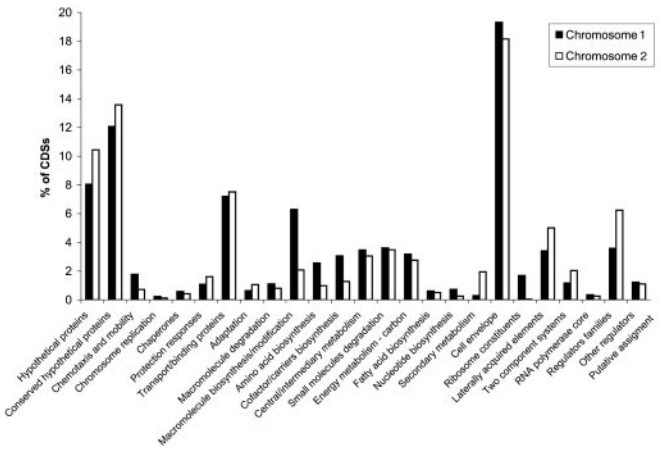

Fig. 2. Distribution of the CDSs belonging to different functional classes on the two chromosomes of *B. pseudomallei*. Figures for functional classes on each chromosome are expressed as a percentage of the total number of CDSs on that replicon.

metabolism involved in amino acid biosynthesis and energy metabolism that are absent from chromosome 1.

Analysis of the predicted functions of the CDSs on the two chromosomes reveals distinct partitioning of core and accessory functions (Fig. 2). Chromosome 1 contains a higher proportion of CDSs involved in core functions, such as macromolecule biosynthesis, amino acid metabolism, cofactor and carrier synthesis, nucleotide and protein biosynthesis, chemotaxis, and mobility. Chromosome 2, by contrast, contains a greater proportion of CDSs encoding accessory functions: adaptation to atypical conditions, osmotic protection and iron acquisition, secondary metabolism, regulation, and laterally acquired DNA. In addition, chromosome 2 contains a greater proportion of CDSs with matches to hypothetical proteins or proteins that have no database matches at all. Comparison of the two chromosomes reveals that there is very little similarity, except in the regions of the rRNA clusters. This partitioning of core and accessory functions is reminiscent of the arrangement in the actinomycete soil-dwelling bacterium, *Streptomyces coelicolor* strain A3 (2), where the 8-Mb single linear chromosome is broadly described as having a central portion that encodes the majority of essential and housekeeping functions flanked by two arms encoding accessory functions (16).

Analysis of *B. pseudomallei* proteins with orthologue matches to proteins encoded in other previously sequenced bacterial genomes revealed the highest number of matches with *Ralstonia solanacearum*, another member of the Burkholderiaceae (2,535 of the 5,855 *B. pseudomallei* proteins had matches in *R. solanacearum*; see Table 8, which is published as supporting information on the PNAS web site) (17). Notably, this organism also has a bipartite genome structure, containing a 3.7-Mb chromosome and a 2.0-Mb megaplasmid. Orthologous matches were identified on both chromosomes; 57% of CDSs on chromosome 1 and 25% of CDSs on chromosome 2 have matches. Comparison to the more distant pseudomonads, *Pseudomonas aeruginosa*, *Pseudomonas putida*, and *Pseudomonas syringae,* and *Xanthomonas campestris*, all of which have a single chromosome, produced fewer total orthologue matches (Table 8) (18–21). Chromosome 2 appears to be the more divergent of the *B. pseudomallei* replicons, containing a smaller percentage of CDSs with matches for each comparison.

Analysis of the levels of gene order conservation in comparison with *R. solanacearum* also highlights the relative genetic diversity of the two *B. pseudomallei* replicons. The large replicons contain discrete regions of conserved gene order, but there has been a large amount of recombination reciprocal about the origin of replication (see Fig. 4*A*, which is published as supporting information on the PNAS web site), as is common with other intergenomic comparisons between related species (22). When the small replicons were compared, no conserved gene order was detected, suggesting that extensive recombination may have shuffled the gene order beyond recognition. Equally, because the number of orthologous matches is relatively low (579 CDSs), it may be that the two secondary replicons do not share a common ancestry. Shared genes may have been transferred independently to secondary replicons since the two organisms diverged from a common ancestor. This model is also suggested by a lack of similarity of genes in the regions around the origins of replication. Similar genome comparisons with other organisms, including the pseudomonads, failed to reveal any significant levels of gene order conservation for either *B. pseudomallei* chromosome.

At 7.3 Mb, the *B. pseudomallei* genome is large in comparison with the typical prokaryotic genome. Horizontal acquisition of DNA appears to have been intrinsic to the evolution of this organism. Many regions within both chromosomes showed some of the characteristics of GIs acquired through very recent lateral transfer, such as anomalies in the percentage of G + C content or dinucleotide frequency signature of the DNA in these regions. Additionally, these regions often contained CDSs with similarities to genes associated with mobile genetic elements, such as insertion sequence (IS) elements, bacteriophages, and plasmids. Twelve putative GIs have been identified on chromosome 1, and four have been identified on chromosome 2, each comprising ≈7.6% and ≈4.2% of the DNA of these replicons, respectively (Table 1). In addition, there are also several other regions of the genome that may have arisen by means of gene acquisition, although the evidence is less clear.

At least three of the GIs appear to be prophages (GIs 2, 3, and 15). It is unclear from sequence analysis alone whether all of these prophages are capable of lysogeny, although we have shown that the K96243 strain produces at least one lysogenic phage after UV induction, ϕK96243 (see Fig. 5, which is published as supporting information on the PNAS web site). This phage can be classified as a member of the order *Caudovirales* and the family *Myoviridae* (23) and can also infect *B. mallei*. Notably, *B. mallei* strains that do not produce lipopolysaccharide O-antigen were not infected with ϕK96243, suggesting that this surface molecule is the phage receptor. Sequencing of DNA fragments from the ϕK96243 genome reveals that this bacteriophage corresponds to GI 2. Analysis of the integration site of this prophage identified a 45-bp repeat corresponding to the 3′ end of the tRNA-Phe as the likely attachment site.

Several of the other GIs encode large numbers of hypothetical proteins with no database matches plus a few phage-like proteins. It is unclear whether these are prophage-like entities, such as defective or satellite phages. In addition to GIs whose transfer may be bacteriophage mediated, some of the islands may be transferred by other mechanisms. One of the small GIs identified, GI 11, contains CDSs with similarities to plasmid conjugal transfer proteins, suggesting that it may be an integrated conjugative element (24). The other GIs identified do not share similarity with any characterized mobile elements. Some of these elements are located next to tRNA genes and are flanked by small repeats (GIs 4, 5, 7, 9, 10, and 13). Many of the GI regions contain multiple phage-like integrases (e.g., GIs 3 and 4), suggesting that they may be composite in nature, having arisen from multiple insertion events over a period of time.

B. mallei is the etiological agent of glanders, a disease of horses that can be transmitted to humans to cause a life-threatening

Table 1. Genomic islands of *B. pseudomallei*

Island	Size, kb	CDS Coordinates	Integrases	GC (%)	D	K	*Bp*08	Functional note
Chromosome 1								
GI 1	12.6	BPSL0081–BPSL0092	1	– (61.5)	–	+	–	Miscellaneous island; contains lipoprotein and hypothetical proteins
GI 2	36.2	BPSL0129–BPSL0176	1	– (65.4)	–	–	–	Prophage ϕK96243
GI 3	51.3	BPSL0548–BPSL0589	3	+ (56.6)	–	–	–	Putative prophage; contains dienelactone hydrolase family protein
GI 4	39.7	BPSL0745–BPSL0772	3 (1)	+ (56.8)	+	–	–	Miscellaneous island; contains putative RNA 2′-phosphotransferase and putative helicases
GI 5	11.7	BPSL0944–BPSL0953	1	+ (58.7)	–	–	–	Contains putative type I restriction system and plasmid replication protein
GI 6	15.0	BPSL1137–BPSL1157	1	+ (58.8)	–	–	–	Prophage-like
GI 7	5.9	BPSL1384–BPSL1393	(1)	+ (58.6)	–	–	+	Prophage-like; contains exported avidin family protein
GI 8	92.3	BPSL1637–BPSL1709	—	– (62.0)	–	+	+	Miscellaneous island; contains transport protein, hemolysin-related protein, amine catabolism, various regulators, and YadA-like exported protein
GI 9	9.8	BPSL2568–BPSL2586	1	– (64.0)	–	–	–	Prophage-like
GI 10	6.6	BPSL3114–BPSL3118	—	+ (54.5)	–	–	+	Restriction and modification system
GI 11	15.3	BPSL3257–BPSL3270	0	+ (55.9)	–	–	+	Putative integrated plasmid or integrated conjugative element; contains recombinase, conjugal plasmid transfer, and replication proteins
GI 12	11.5	BPSL3342–BPSL3353	1	+ (57.3)	+	–	–	Prophage-like
Chromosome 2								
GI 13	19.0	BPSS0378–BPSS0391A	1	+ (58.4)	–	+	–	Prophage-like, contains hypothetical proteins
GI 14	18.6	BPSS0652A–BPSS0666	—	+ (55.1)	+	+	+	Miscellaneous island; contains putative collagenase
GI 15	34.6	BPSS1047–BPSS1089	0	– (65.2)	+	–	–	Putative prophage; contains hypothetical proteins and ParA homologue
GI 16	61.8	BPSS2051–BPSS2090	0	+ (59.3)	+	+	+	Metabolic island; contains various ABC transporters, large cell surface protein, L-asparaginase, and fatty aldehyde dehydrogenase

The size and CDS coordinates of the putative genomic islands are indicated. The number of putative integrases contained within islands are given; —, not applicable. Partial genes and pseudogenes are indicated in parentheses. Whether anomalies are present in the properties of DNA within islands is indicated in the GC, D, and K columns, representing the percentage of G + C content, G + C deviation, and Karlin dinucleotide frequency signature, respectively: +, anomolies are present; –, anomolies are not present. The *Bp*08 column indicates regions of the K96243 genome that are identical to islands previously identified by using representative difference analysis of *B. pseudomallei* strain 08 against *B. thailandensis*: +, match with *B. thailandensis*; –, no match with *B. thailandensis*. GI 7 is the only GI found in the *B. mallei* ATCC 23344 genome. ABC, ATP-binding cassette.

illness (10). Compared with *B. pseudomallei*, the distribution and host range of *B. mallei* is more narrow. This organism does not survive in the environment, and its natural reservoir is thought to be equines (25). Before the sequencing of both genomes, DNA–DNA hybridization suggested that these two species were closely related (26). A recent multilocus sequence typing study of *B. pseudomallei* and *B. mallei* isolates from around the globe has revealed that *B. mallei* can be considered to be a clone of *B. pseudomallei* (27).

Although *B. mallei* strains appear to have recently evolved from a *B. pseudomallei* ancestor, there are large differences in the sizes of the genomes; the *B. pseudomallei* genome is 1.31 Mb larger than that of *B. mallei*. Comparative analysis of chromosomes reveals that 16% of chromosome 1 and 31% of chromosome 2 of *B. pseudomallei* are unique with respect to the equivalent chromosomes of *B. mallei* (Fig. 1). Conversely, >1% of chromosome 1 and 4% of chromosome 2 of *B. mallei* are unique with respect to the equivalent chromosomes of *B. pseudomallei*. Notably, the majority of unique regions of the *B. mallei* chromosome 2 consist of DNA regions of >2 kb. The single largest region of difference of *B. mallei* chromosome 2 is a 41-kb fragment, which is found on chromosome 1 of *B. pseudomallei* (CDS coordinates BPSL3040–BPSL3078), resulting from an interchromosomal transposition. Comparison of the two genomes also reveals that there has been a significant amount of intrachromosomal rearrangements in *B. mallei*, re-

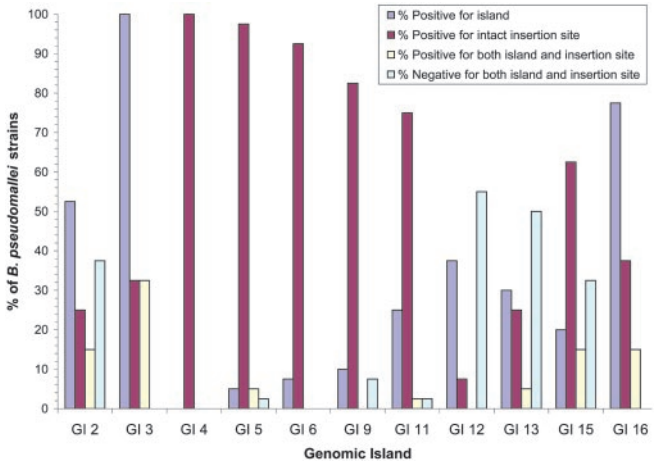

Fig. 3. Prevalence of GIs in environmental and invasive clinical isolates of *B. psuedomallei*. Distribution of GIs in environmental and invasive clinical *B. pseudomallei* isolates ($n = 40$) as determined by multiplex PCR.

sulting from recombination between the large number of repeated IS elements in its genome (see Fig. 4*B*) (10).

The unique regions in the *B. pseudomallei* genome with respect to *B. mallei* have been summarized in Table 9, which is published as supporting information on the PNAS web site. One of the main sources of additional DNA are GIs; all but one of the islands identified in Table 1 (GI 7) are absent in the *B. mallei* genome. In the case of GI 7, fragments of this island were found in the *B. mallei* genome, suggesting that a similar island may have been present in an ancestral strain that subsequently decayed in *B. mallei* ATCC 23344. Although many of the *B. pseudomallei* islands contain DNA with properties that suggest that they have recently been acquired, it is very unlikely that all of these islands have been acquired by K96243 in the time since *B. mallei* diverged from a *B. pseudomallei* ancestor. Rather, these islands can be seen as representative of the diversity within *B. pseudomallei* and specifically as being variably present between K96243 and the *B. pseudomallei* strain that was the immediate ancestor of *B. mallei*. This result suggests that the distribution of these islands is highly variable between different *B. pseudomallei* strains and that the islands are therefore actively transmissible. This interpretation is supported by probing phylogenetically diverse clinical and environmental isolates of *B. pseudomallei* from Thailand with multiplex PCR of target genes within a selection of the K96243 GIs (Fig. 3). The screen identified considerable variation in the distribution of the GIs: GI 3 was ubiquitous; GIs 2, 11, 12, and 16 were detected in 10–70% of strains; GIs 5, 6, and 9 were detectable in 10% or less of strains; and GI 4 was unique to K96243.

Furthermore, subtractive hybridisation analysis between *B. pseudomallei* strain 08 and *Burkholderia thailandensis*, a nonpathogenic relative of *B. pseudomallei*, indicates that GIs are key determinants of genome plasticity (28). Of 92 sequence fragments found to be present in *B. pseudomallei* strain 08 and absent from *B. thailandensis*, only 53 (58%) can be mapped to the K96243 genome. Of these fragments, 27 fell in GIs 7, 8, 10, 11, 14, and 16 (Table 1), confirming these GIs as likely to have been recently acquired by *B. pseudomallei* K96243. As a corollary to this finding, 42% of the sequences found to be unique to strain 08 were not present in the K96243 genome, confirming the large degree of variability in the carriage of these islands among *B. pseudomallei* strains.

The genomic diversity observed in the closely related but host-restricted *B. mallei* genome sheds light on differences in lifestyle and niche adaptation. Comparison of *B. mallei* and *B. pseudomallei* genomes reveals that both have features of genetic plasticity, but the observed variation points toward different modes of change. The absence of GIs in the *B. mallei* genome suggests that gene acquisition does not play as significant a role in the genetic variation of this species as it does in *B. pseudomallei*. Conversely, gene loss appears to have been an important source of genetic variation in the recent evolution of *B. mallei*. Several of the unique regions in the *B. pseudomallei* genome contained CDSs at their 5′ and 3′ ends that are truncated and adjacent to one another in the *B. mallei* genome, suggesting that these regions have been deleted in *B. mallei* ATCC 23344 since its divergence from *B. pseudomallei* (see Table 9). This observation is consistent with other host-restricted bacteria that have evolved from versatile ancestors and have undergone gene loss (29, 30). *B. mallei* also contains a much larger number of IS elements and detectable pseudogenes in comparison with *B. pseudomallei* (K96243 contains 42 IS elements and 26 pseudogenes), which points toward recent host adaptation and a reduction in selective pressure on parts of the *B. mallei* genome (10, 30, 31). In *B. pseudomallei* IS elements and pseudogenes do not appear to contribute to the genetic variation of this species to the same extent. Both species' genomes contain large numbers of small sequence repeats. In *B. mallei*, these small sequence repeats have been shown to play a role in the mutation of some of the detectable pseudogenes and a source of proteome variation (10). Comparison of small sequence repeats in the two species identified variation in the size of some of the repeats; however, the extent to which they contribute to variation in *B. pseudomallei* strains has yet to be determined.

The *B. pseudomallei* genome contains many genes that promote survival in diverse and challenging environments and a large selection of genes that modulate pathogenicity and host–cell interaction (Table 2). Comparative analysis of *B. pseudomallei* and *B. mallei* has identified many CDSs that may contribute to the phenotypic differences between the two species (Table 2). These phenotypes include known virulence determinants, such as flagella (32) and a type III protein secretion system (33); potential virulence determinants, such as surface polysaccharides, exoproteins, fimbriae, pili, and putative adhesins; drug resistance determinants; and potential environmental survival functions, including various secondary metabolite pathways, numerous catabolic pathways, transport systems, and stress-response proteins. A detailed description of these features and the role they play in the biology of *B. pseudomallei* is presented in *Supporting Text*.

The dual existence of *B. pseudomallei* as soil colonizer and effective human pathogen requires flexibility, a fact reflected in the genome. At the most rudimentary level, the size of the genome endows the organism with an expansive inventory of CDSs encoding diverse functions that promote survival and success in different environments. This capacity is evident from the large metabolic repertoire encoded and in the redundancy seen for some of the virulence determinants, such as the type III protein secretion systems and fimbriae (Table 2). The bipartite structure of the genome and distinct partitioning of encoded functions also suggest genetic suppleness. The reduced orthology and lack of conservation of gene order of chromosome 2 with other genomes suggest that this replicon may have its origins in the carriage of accessory functions that promote survival in diverse niches.

The content of the genome is supplemented by horizontally acquired DNA, an important feature in the recent evolution of *B. pseudomallei*. The diversity of the GIs also suggests that transfer occured by a variety of means. Subsequent analysis of the distribution of these islands in environmental and clinical strains of *B. pseudomallei* has detected variability in the complement of islands, suggesting a large pool of islands throughout the population. The presence of such genetic fluidity has important implications for the future study of disease pathogenesis and in the development of vaccination

Table 2. Survival and virulence functions encoded in the *B. pseudomallei* genome

Function	Notes and examples
Survival	
Secondary metabolism	Fourteen clusters encoding possible antibiotic, surfactant, and siderophore biosynthesis pathways, including putative hydroxamate (BPSL1774–BPSL1779) and pyochelin (BPSS0581–BPSS0588*) siderphores
Drug resistance	Seven Ambler class A, B, and D β-lactamases, including cephalosporinase (BPSS0946) and oxacillinase (BPSS1997*); six multidrug efflux systems, including aminoglycoside-*, macrolide-*, and polymyxin B-specific systems; aminoglycoside acetyltransferase (BPSS0262)
Intracellular stress	Superoxide detoxification, superoxide dismutases (BPSL0880 and BPSL1001), and catalases (BPSS0993 and BPSS2214*); nitric oxide detoxification; flavohaemoglobin (BPSL2840)
Motility and chemotaxis	Five gene clusters encode the components of a single flagella system; 38 chemotaxis-associated proteins*
Virulence	
Secretion	Type I,* type II, type III*, and type V protein secretion systems, including three type III systems (TTS1*, BPSS1390–BPSS1408; TTS2, BPSS1613–BPSS1629; and TTS3, BPSS1543–BPSS1552) direct the secretion of proteins outside the bacterium
Lipopolysaccharide and capsule	Capsular polysaccharide synthesis and export cluster (BPSL2787–BPSL2810*); lipopolysaccharides biosynthetic cluster (BPSL2672–BPSL2688); two other potential surface polysaccharides biosynthetic clusters (BPSS1825–BPSS1834* and BPSS0417–BPSS0429*).
Exoproteins	The genome contains many secreted exoenzymes that can break down host tissues, including phospholipases C (BPSL0338*, BPSL2403, and BPSS0067), metalloprotease A (BPSS1993*), and a homologue of the *P. aeruginosa* MucD Ser protease (BPSL0808) and a putative collogenase (BPSS0666*)
Adhesins	Several surface proteins potentially modulate host–cell interactions, these including seven Hep_Hag repeat family proteins (BPSL1631, BPSL1705*, BPSL2063, BPSS0796, BPSS0908, BPSS1434*, and BPSS1439).
Fimbriae and pili	Thirteen clusters encoding components of type I fimbriae*, type IV pili, and *tad*-type pili*

*Gene or genes associated with this function are absent in the *B. mallei* ATCC 23344 genome (see *Supporting Text*).

strategies. Adaptability in the human host may underlie the protean disease manifestations and the ability of this organism to cause chronic disease and to recrudesce many years after apparently effective treatment. There is a growing recognition that the global burden of melioidosis is much greater than current estimates, a result of limitations in diagnostic facilities (3). Coupled with the status of *B. pseudomallei* as a biothreat agent, there is therefore a pressing need to gain a better understand of the role that horizontal gene transfer plays in the genomic plasticity of this potent pathogen.

We thank William Nierman and the staff at The Institute for Genomic Research for sharing data before publication. We also acknowledge the support of the Wellcome Trust Sanger Institute core sequencing and informatics groups. This work was supported by the Wellcome Trust through its Beowulf Genomics initiative. S.J.P. holds a Wellcome Trust Career Development Award in Clinical Tropical Medicine.

1. Dance, D. A. B. (1991) *Clin. Microbiol. Rev.* **4,** 52–60.
2. Dance, D. A. B. (2002) *Curr. Opin. Infect. Dis.* **15,** 127–132.
3. White, N. J. (2003) *Lancet* **361,** 1715–1722.
4. Chaowagul, W., White, N. J., Dance, D. A. B., Wattanagoon, Y., Naigowit, P., Davis, T. M. E., Looareesuwan, S. & Pitakwatchara, N. (1989) *J. Infect. Dis.* **159,** 890–899.
5. Rotz, L. D., Khan, A. S., Lillibridge, S. R., Ostroff, S. M. & Hughes, J. M. (2002) *Emerg. Infect. Dis.* **8,** 225–230.
6. Yee, K. C., Lee, M. K., Chua, C. T. & Puthucheary, S. D. (1988) *J. Tropic. Med. Hyg.* **91,** 249–254.
7. Mays, E. E. & Ricketts, E. A. (1975) *Chest* **68,** 261–263.
8. Koponen, M. A., Zlock, D., Palmer, D. L. & Merlin, T. L. (1991) *Arch. Intern. Med. (Moscow)* **151,** 605–608.
9. Sanford, J. P. & Moore, W. L. (1971) *Am. Rev. Respir. Dis.* **104,** 452–453.
10. Nierman, W. C., DeShazer, D., Kim, H., Tettelin, H., Nelson, K. E., Feldblyum, T. V., Ulrich, R. L., Ronning, C. M., Brinkac, L. M., Daugherty, S. C., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101,** 14246–14251.
11. Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., *et al.* (2000) *Nature* **404,** 502–506.
12. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000) *Bioinformatics* **16,** 944–945.
13. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
14. Elnifro, E. M., Ashshi, A. M., Cooper, R. J. & Klapper, P. E. (2000) *Clin. Microbiol. Rev.* **13,** 559–570.
15. Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H. & Vogt, P. H. (1997) *BioTechniques* **23,** 504–511.
16. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., *et al.* (2002) *Nature* **417,** 141–147.
17. Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J. C., Cattolico, L., *et al.* (2002) *Nature* **415,** 497–502.
18. Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S. L., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., *et al.* (2000) *Nature* **406,** 959–964.
19. Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., dos Santos, V., Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., *et al.* (2002) *Environ. Microbiol.* **4,** 799–808.
20. Buell, C. R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I. T., Gwinn, M. L., Dodson, R. J., Deboy, R. T., Durkin, A. S., Kolonay, J. F., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10181–10186.
21. da Silva, A. C. R., Ferro, J. A., Reinach, F. C., Farah, C. S., Furlan, L. R., Quaggio, R. B., Monteiro-Vitorello, C. B., Van Sluys, M. A., Almeida, N. F., Alves, L. M. C., *et al.* (2002) *Nature* **417,** 459–463.
22. Eisen, J. A., Heidelberg, J. F., White, O. & Salzberg, S. L. (2000) *Genome Biol.* **1,** research0011.1–0011.9.
23. Ackermann, H. W. (2003) *Res. Microbiol.* **154,** 245–251.
24. Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. (2002) *Plasmid* **48,** 77–97.
25. Howe, C. (1950) in *The Oxford Medicine*, ed. Christian, H. A. (Oxford Univ. Press, New York), pp. 185–202.
26. Rogul, M., Brendle, J. J., Haapala, D. K. & Alexander, A. D. (1970) *J. Bacteriol.* **101,** 827–835.
27. Godoy, D., Randle, G., Simpson, A. J., Aanensen, D. M., Pitt, T. L., Kinoshita, R. & Spratt, B. G. (2003) *J. Clin. Microbiol.* **41,** 2068–2079.
28. Brown, N. F. & Beacham, I. R. (2000) *J. Med. Microbiol.* **49,** 993–1001.
29. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., *et al.* (2001) *Nature* **409,** 1007–1011.
30. Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N., Harris, D. E., Holden, M. T. G., Churcher, C. M., Bentley, S. D., Mungall, K. L., *et al.* (2003) *Nat. Genet.* **35,** 32–40.
31. Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T. G., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., *et al.* (2001) *Nature* **413,** 523–527.
32. Chua, K. L., Chan, Y. Y. & Gan, Y. H. (2003) *Infect. Immun.* **71,** 1622–1629.
33. Stevens, M., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., Wallis, T. S. & Galyov, E. E. (2002) *Mol. Microb.* **46,** 649–659.