Burkholderia pseudomallei Sequencing Identifies Genomic Clades with Distinct Recombination, Accessory and Epigenetic Profiles

Tannistha Nandi[1], Mathew T. G. Holden[2ξ], Xavier Didelot[3], Kurosh Mehershahi[4], Justin A. Boddey[5†], Ifor Beacham[5], Ian Peak[5], John Harting[6], Primo Baybayan[6], Yan Guo[6], Susana Wang[6], Lee Chee How[6], Bernice Sim[1], Angela Essex-Lopresti[7], Mitali Sarkar-Tyson[7], Michelle Nelson[7], Sophie Smither[7], Catherine Ong[8], Lay Tin Aw[8], Chua Hui Hoon[1], Stephen Michell[9], David J. Studholme[9], Richard Titball[9,10], Swaine L. Chen[1,4], Julian Parkhill[2], Patrick Tan[1,11,12]*

[1]Genome Institute of Singapore, Singapore, Republic of Singapore
[2]The Wellcome Trust Sanger Institute, Cambridge, UK
[3]Department of Infectious Disease Epidemiology, Imperial College London, United Kingdom
[4]Department of Medicine, National University of Singapore, Singapore, Republic of Singapore
[5]Institute for Glycomics, Griffith University (Gold Coast Campus), Southport, Queensland, Australia
[6]Pacific Biosciences, 1380 Willow Road, Menlo Park, CA, United States
[7]Defence Science and Technology Laboratory, Porton Down, Salisbury, United Kingdom
[8]Defense Medical and Environmental Research Institute, DSO National Laboratories, Singapore, Republic of Singapore
[9]Biosciences, University of Exeter, Exeter, United Kingdom
[10]Faculty of Infectious and Tropical Diseases, Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, United Kingdom
[11]Duke-NUS Graduate Medical School Singapore, Singapore, Republic of Singapore
[12]Cancer Science Institute of Singapore, National University of Singapore, Republic of Singapore

[ξ]Present address: School of Medicine, University of St Andrews, St Andrews, UK
[†]Present address: Division of Infection and Immunity, The Walter and Eliza Hall Institute of Medical Research, Parkville, 3052, Victoria, Australia

*Address correspondence to tanbop@gis.a-star.edu.sg

Running title: Genomic landscape of *Burkholderia pseudomallei*

Key words: whole genome sequencing, bacterial haplotypes, recombination

ABSTRACT

Burkholderia pseudomallei (Bp) is the causative agent of the infectious disease melioidosis. To investigate population diversity, recombination, and horizontal gene transfer in closely-related Bp isolates, we performed whole-genome sequencing (WGS) on 106 clinical, animal, and environmental strains from a restricted Asian locale. Whole-genome phylogenies resolved multiple genomic clades of Bp, largely congruent with multi-locus sequence typing (MLST). We discovered widespread recombination in the Bp core genome, involving hundreds of regions associated with multiple haplotypes. Highly recombinant regions exhibited functional enrichments which may contribute to virulence. We observed clade-specific patterns of recombination and accessory gene exchange, and provide evidence that this is likely due to ongoing recombination between clade members. Reciprocally, inter-clade exchanges were rarely observed, suggesting mechanisms restricting gene flow between clades. Interrogation of accessory elements revealed that each clade harbored a distinct complement of restriction-modification (RM) systems, predicted to cause clade-specific patterns of DNA methylation. Using methylome sequencing, we confirmed that representative strains from separate clades indeed exhibit distinct methylation profiles. Finally, using an *E. coli* system, we demonstrate that Bp RM systems can inhibit uptake of non-self DNA. Our data suggests that RM systems borne on mobile elements, besides preventing foreign DNA invasion, may also contribute to limit exchanges of genetic material between individuals of the same species. Genomic clades may thus represent functional units of genetic isolation in Bp, modulating intra-species genetic diversity.

INTRODUCTION

Burkholderia pseudomallei (Bp) is the causative agent of melioidosis, a serious infectious disease of humans and animals and a leading cause of community acquired sepsis and pneumonia in endemic regions (Currie et al. 2010). Initially thought to be confined to South East Asia and Northern Australia, the prevalence of Bp appears to be spreading(Wiersinga et al. 2012), and Bp has been designated a biothreat select agent in the USA. Bp can persist in extreme environmental conditions and can infect several animal and plant hosts, including birds, dolphins and humans(Wuthiekanun et al. 1995; Howard and Inglis 2003; Sprague and Neubauer 2004; Larsen et al. 2013). Treatment of clinical melioidosis is challenging as the bacterium is inherently resistant to many antibiotics, and Bp infections can persist in humans for over a decade(Hayden et al. 2012; Wiersinga et al. 2012).

The Bp genome comprises one of the largest and most complex bacterial genomes sequenced to date. Consisting of two large circular replicons (chromosomes) with a combined 7.2Mb genome size(Holden et al. 2004), it contains a rich arsenal of genes related to virulence (e.g. Type III and Type VI secretion systems, polysaccharide biosynthesis clusters), metabolic pathways, and environmental adaptation(Wiersinga et al. 2012). Besides conserved regions, accessory genes on mobile elements and genomic islands may also contribute to phenotypic and clinical differences in microbial behavior (Currie et al. 2000; Sim et al. 2008). Analysis of the Bp genome has revealed previously unknown toxins and mechanisms of antibiotic resistance (Chantratita et al. 2011; Cruz-Migoni et al. 2011).

Most large-scale studies of Bp genetic diversity to date have analyzed strains using Multi-Locus Sequence Typing (MLST). These studies have suggested a high degree of genetic variability between Bp strains and related *Burkholderia* species(Cheng et al. 2008), and shown that Bp strains belonging to different sequence types (STs) can often co-exist in the same locale and sometimes even within the same sample (Pitt et al. 2007; Wuthiekanun et al. 2009). However, due to the limited number of genes analyzed by MLST, these studies cannot comment on the global proportion of genetic material shared between strains of different STs, nor on the relative contribution of recombination, mutation, and horizontal gene transfer on intra-species genetic diversity. Moreover, while previous studies have applied whole genome sequencing (WGS) to study global patterns of Bp genetic heterogeneity and evolution, earlier Bp WGS reports have been confined to a limited number of isolates (10-12) derived from diverse geographical regions (Nandi et al. 2010), where geophysical barriers likely limit the propensity of the analyzed strains to exchange genetic material. To achieve a comprehensive understanding of genetic variation among closely-related Bp strains, WGS analysis of much larger strain panels, ideally performed on strains isolated from a common region and belonging to the same (or closely related) ST groups, are required.

In this study, we attempted to fill this important knowledge gap by performing WGS on 106 Bp strains drawn from a restricted Asian locale (Singapore and Malaysia). The WGS data, exceeding previous Bp WGS studies by 10-fold, enabled us to identify specific genomic clades of Bp, molecular features of Bp recombination at the whole-genome level, and accessory genome features contributing to recombination and horizontal gene transfer. We found a consistent pattern of genetic separation correlating with MLST, recombination haplotypes, shared accessory genes, and restriction modification (RM) systems. We provide evidence that restriction modification, beyond its role as defense against foreign DNA invasion, may have also partitioned the Bp species by restricting gene flow, resulting in the other observed correlations. Because RM-systems are widely dispersed through the bacterial kingdom, it is possible that similar principles may apply to other bacterial species, implicating a potential role for epigenetic barriers as a driver of early incipient speciation.

RESULTS

Bp Genome Sequencing

We analyzed 106 Bp strains, including 97 strains from Singapore and Malaysia (87/10) and 9 strains from Thailand (Supplemental Table S1). The Singaporean and Malaysian strains were isolated from various clinical, animal, and environmental sources over a 10 year period (1996 - 2005) (see Methods). MLST classified the strains into 22 sequence types (ST). Supporting their close phylogenetic relationships, 20 STs belonged to clonal complex CC48 (Supplemental Figs. S1A, S1B). The majority of strains were of ST51 (43 strains) and ST423 (16 strains).

Due to the high GC-bias of the Bp genome, we initially found that conventional Illumina sequencing protocols resulted in uneven genome coverage and suboptimal assemblies (median N_{50}: 2907 bp) (Supplemental Table S2). We overcame this problem by applying a PCR amplification-free strategy (Kozarewa et al. 2009), resulting in markedly improved genome coverage and assemblies (median depth 100X, median N_{50}: 102577bp). In total, we predicted 84846 high quality SNPs in the WGS panel compared to the K96243 reference (Chr I: 43829 and Chr II: 41017). We validated the technical accuracy of the WGS data by Sanger sequencing of 50 randomly-selected SNPs. Of the predicted SNPs, all 50 were confirmed by Sanger sequencing.

Whole-Genome Phylogenetic Analysis Resolves Genomic Clades

We excluded SNPs associated with regions of recombination as previously described (Croucher et al. 2011), resulting in a set of 10314 SNPs representing mutations inherited by vertical descent along different lineages ("lineage SNPs" (L-SNPs)). Maximum likelihood phylogenies using the L-SNPs identified three major clades (“genomic clades”) containing all the Singapore and Malaysian strains, clustering apart from Thailand strains (Fig. 1A). Strains of the same ST grouped together within the same genomic clade, indicating strong similarities between phylogenies based on WGS and MLST. However, compared to MLST, the WGS phylogenies provided increased resolving power. For example, while MLST indicated a high degree of relatedness between ST50 and ST414 strains, WGS revealed that ST50 is more related to ST46 (Genomic Clade C), with ST414 being a more distant group (Fig. 1A). WGS also subdivided the ST51 strains into two distinct subclades - ST51a (39 strains) and ST51b (4 strains) (Fig. 1B), distinguished by ~342 distinct L-SNPs (Supplemental Table S3). Notably, all three clades comprised a heterogeneous intermingling of strains from different isolation sources (e.g. clinical, animal, and environmental), arguing against the existence of a genetically distinct Bp subpopulation preferentially associated with human disease. The three clades also contained strains isolated during similar time periods (1996-2005), suggesting that genetically distinct Bp strains from different clades can co-exist in the same region over many years.

L-SNPs occurred at a ~1.2 fold higher frequency on Chr II compared to Chr I (6.1×10^{-3} SNPs per site for Chr 1 vs 7.5×10^{-3} for Chr II, p=3.08e-14, χ^2 =57.68, chi-squared test) (Supplemental Table S4a), suggesting a preferential accumulation of genetic mutations on Chr II during evolution. The majority of L-SNPs corresponded to C/G → T/A transitions (Fig. 1C; for Clade ST51a), in the context of CG dinucleotides ($p = 3 \times 10^{-10}$, binomial test, Fig. 1D), likely reflecting the tendency of methylated cytosines to form thymines (Kahramanoglou et al. 2012). For both chromosomes, L-SNPs preferentially localized to intergenic regions (Chr I: p<2.2e-16, χ^2 =101.42; Chr II: p=4.196e-14, χ^2 = 57.04, chi-square test), and one-third of L-SNPs occurring within genes were non-synonymous (Supplemental Table S4b). The d_N/d_S ratio (proportion of rate of nonsynonymous substitutions per site to rate of synonymous substitutions per site) for the major STs (e.g. ST51a, ST84) ranged between 0.17 and 0.64 per genome. Similar results were obtained when we analyzed a more restricted subset of 8035 L-SNPs associated with the Bp "core" genome (regions common to all Bp strains, estimated size 5.64 Mb) (Supplemental Table S5).

Widespread Recombination Among Bp Isolates

Bp strains that are genetically distinct may interact within environmental reservoirs such as soil or water (Chantratita et al. 2008; Mayo et al. 2011) or in co-infected animals and human hosts (Pitt, Trakulsomboon and Dance 2007), thereby providing opportunities for recombination. Hence, it is vital to understand pathways and processes that facilitate or constrain gene flow between strains. To identify genomic characteristics of Bp recombination, we proceeded to analyse SNPs associated with recombination (R-SNPs). From 74532 R-SNPs, we identified 2373 recombination events across the three genomic clades, with recombination tract lengths ranging from 3bp to 71kb (median ~5kb). We computed recombination/mutation (r/m) values, corresponding to the ratio of rates at which substitutions are introduced by recombination and mutation, across the entire population. The overall per site r/m ratio was 7.2. Based upon this data, we estimate that at least 78% of the BpK96243 reference genome (~5.67Mb) has undergone recombination, a level comparable to *S. pneumoniae*, a highly recombinogenic species (74K/85K R-SNPs for Bp; 50K/57K R-SNPs for *S. pneumoniae*) (Croucher et al. 2011). Similar to L-SNPs, higher recombination levels were observed for Chr II than Chr I ($p < 2.2e-16$, Mann-Whitney *U* test),

Besides estimating whole-population metrics, we also computed clade-specific recombination and mutation rates for the three major Bp genomic clades, using a previously-described Bayesian approach (ClonalFrame (Didelot and Falush 2007)). To minimize mapping artefacts, we excluded mobile genetic elements (eg phages, transposons, and genomic islands) and based our analysis on a reduced core genome of 5.6 Mb (see Methods). For all three clades, the ratio of mutation rate (theta) to recombination rate (rho) was close to one, suggesting that recombination and mutation both happen at approximately the same rates. Recombination was also found to introduce more substitutions than mutation (r/m=4.5 in clade A, r/m=8.5 in clade B and r/m=6 in clade C) with the highest impact observed in Clade B (Supplemental Table S6). These values are in general agreement with the values obtained from the total population.

The high levels of recombination in the Bp clades motivated us to also analyze potential sources of recombination imports. We used previously established methods to assess intra and inter-clade recombination flux (Didelot et al. 2009; Didelot et al. 2011). Briefly, recombined fragments were compared with homologous sequences from other Bp genomes across the three clades, and a “match” was found if the sequence was identical or contained a single nucleotide difference. If a match was found to members of a single clade, the origin of the recombination event was attributed to this clade (matches to strains from multiple clades were categorized as ambiguous). If no matches were found, the origin was categorized as unknown. To estimate their relative impact on genomic diversification, the flux of genomic content between clades was summarized as the proportion of each genome originating from different origins. Of 2481, 821, and 334 recombination events detected within genomic clade A, B, and C respectively, we could assign sources (“matches”) for ~60% of recombination events (1112 matches to single clades and 1059 matches to multiple clades). On average, approximately 5% of each genome from a given clade was found to have originated from another clade and approximately another 7% from a source not present in our data set (Supplemental Table S7). Several of the inter-clade recombination events were found on recent branches of the clonal genealogy, suggesting that the isolation is not complete between the clades.

Genome-wide median recombination frequencies (RFs) were computed to identify genomic regions exhibiting elevated recombination rates and multiple recombination events (Fig. 2A). We identified 1630 protein coding genes (Chr1: 897 genes; Chr II: 733) associated with regions of high recombination (RF > RFmedian + 3MAD, median absolute deviation). Genes experiencing high recombination frequencies were significantly enriched in intracellular trafficking and secretion pathways (corrected p=0.0006, binomial test) while genes involved in protein translation were underrepresented (corrected p=0.012, binomial test) (Supplemental Table S8). Examples of genomic regions exhibiting elevated recombination included a Type III secretion cluster (*TTSS3*; *BPSS1520–BPSS1537*) previously linked to mammalian virulence(Stevens et al. 2002), and a Type IVB pilus cluster (*TFP8*, Chr II: *BPSS2185-BPSS2198*) (Fig. 2B). Type IV pili (*TFP*), including those of the subtype IVB, encode surface-associated protein complexes involved in multiple cellular processes(Craig et al. 2004). To evaluate if *TFP8* might modulate Bp virulence, we generated an isogenic Bp deletion strain lacking approximately 12.9 kb of the *TFP8* locus, and assessed the virulence of the *TFP8* mutant in a BALB/c mouse intranasal infection assay. The *TFP8* deletion mutant exhibited significantly reduced virulence compared to parental Bp K96243 wild-type controls (p=0.026, Mantel-Haenszel log-rank test, Fig. 2C). These results support a role for Type IVB pili in Bp murine virulence, and more generally that a subset of recombination hotspots in Bp may influence mammalian virulence.

Genome-Wide Recombination Haplotype Map of Bp

Extended genomic stretches with high recombination rates often displayed specific combinations of local independent recombination events in individual strains, resulting in the creation of distinct haplotypes. Using the *TFP8* gene cluster as an example, some strains displayed recombination events R2, R7, and R8 (Haplotype 1 (H1)), while other strains displayed events R3, R7, and R8 (Haplotype 2 (H2)). In total, we identified 5 haplotypes (H1-H5) in the *TFP8* gene cluster (Supplemental Table S9). We found that these five *TFP8* haplotypes were tightly associated with specific clades – for example, haplotype H1 was associated with ST51 strains, while haplotype H4 (corresponding to recombination event R4) was associated with ST84 strains (Fig. 2D). To evaluate this association at the whole-genome level, we generated a whole-genome haplotype map of Bp, identifying 85 genomic regions exhibiting multiple (>=5) haplotypes (Supplemental Table S10). Similar to *TFP8*, the vast majority of haplotypes occurred in a genomic-clade specific pattern (Supplemental Fig. S2). Many of the multi-haplotype genomic regions were involved in specialized functions such as iron and cofactor metabolism, detoxification, and virulence (Supplemental Table S10). Almost half (48%) of the multiple-haplotype genomic regions exhibited at least one haplotype with an excess of non-synonymous to synonymous SNPs, consistent with these regions having altered phenotypic properties. For instance, one haplotype over-represented in non-synonymous SNPs occurred in the virulence-associated *TFP1* locus (*BPSL0782-BPSL0783*), within the *pilA* gene in ST51a strains. Notably, *pilA* plays a role in virulence yet its role in adherence and microcolony formation varies considerably in different Bp strains(Essex-Lopresti et al. 2005; Boddey et al. 2006). These findings suggest that haplotype variation may contribute to differences in Bp pathogenicity or survival in different strains.

Bp Accessory Genome Elements Exhibit Clade Restriction

The availability of WGS data for a large Bp panel also provided the opportunity to quantitatively assess the Bp accessory genome. Using the Velvet and Nucmer algorithms, we generated *de novo* sequence assemblies of genomic sequences not found in the K96243 reference genome. On average, ~183 kb of novel accessory regions (N_{AE}) were identified for each Bp strain (minimum region length 1 kb). We found that the Bp genome is “open”, with at least 2897 new non-K96243 genes associated with the accessory regions (Fig. 3A). The Bp pan genome (Bp core + Bp accessory) is thus at least 8802 genes, which is 2x the size of the Bp core genome. Accessory genes were characterized by a lower %GC content (median value: ~59% ± 5.6) than the core genome (~68%), consistent with their horizontally-acquired nature. Accessory genes were also significantly enriched in pathways related to defense mechanisms (corrected $p < 0.0005$ relative to Bp core genes) (Fig. 3B).

Using pair-wise similarity metrics, we evaluated the extent to which accessory elements found in one Bp strain might be shared with other strains. Similar to the recombination haplotypes, strains belonging to the same genomic clade had a tendency to share many of the same accessory elements (Fig. 3C). For example, strains in genomic clade A shared a 15 kb gene cluster of metabolic genes including biotin carboxylase, NAD-dependent malic enzymes, mandelate racemase, and 5-enolpyruvylshikimate-3-phosphate synthase(Priestman et al. 2005; Tang et al. 2005; Li et al. 2009). Similarly, strains from genomic clade B (ST423/ST84/ST289) shared accessory genes such as filamentous hemagglutinin (Fha), *fhaC*, which plays a crucial role in mediating adherence to eukaryotic cells(Relman et al. 1989).

Evidence of Ongoing Recombination and Gene Exchange Within Clades

The analyses described above revealed a strong correlation between Bp clades, core genome recombination haplotypes, and complements of accessory elements. We hypothesized that these correlations could be explained by two alternative models – "Ongoing Recombination" or "Vertical Descent" (Fig. 4A). In the first model, active recombination is ongoing in Bp, but preferentially restricted to exchange of DNA within a clade. To test for ongoing recombination, we computed within-clade nucleotide divergence levels in DNA sequences predicted to have undergone recombination, and compared these to divergence levels within regions of non-recombined DNA in strains exhibiting the recombination event. If recombination is ongoing among strains within a clade, then this would serve to homogenize the recombining sequences across strains, while non-recombining regions would accumulate mutations independently in different strains (Fig. 4A, "Ongoing Recombination"). Thus, the within-clade nucleotide divergence of recombined regions would be predicted to be lower relative to non-recombined regions. In the alternative model, recombination is not commonly taking place. Here, recombined regions would have entered a clade in its founder, and then would be found throughout the clade due mostly to strict vertical descent (Fig. 4A, "Vertical Descent"). In this case, recombined regions would be predicted to accumulate mutations at the same rate as non-recombined regions. For each recombined region in a clade, we calculated the average sequence divergence level in that region, using strains exhibiting the recombination event (see Supplemental Fig. S3 and Supplemental Text for a detailed description of this analysis). To obtain a conservative set of non-recombined sequences for comparison, we then took only those sections of the Bp genome predicted not to have undergone any recombination in any of the Bp strains, and for the same strains we calculated the sequence divergence levels in these non-recombined sequences. We found that recombined regions in the core genome had uniformly lower sequence divergence than non-recombined regions (Fig. 4B), suggesting that recombination is active and ongoing within clades. Extending this analysis to the gene level, we found that recombination has opposite effects on within- and between-clade divergence; genes in recombined regions had higher between-clade diversity compared with genes in non-recombined regions, as expected, but much lower within-clade diversity (Supplemental Fig. S4). To rule out the possibility that the lower sequence divergence in recombined regions might be due to recombined regions possessing different sequence features or gene functions than non-recombined regions (despite both regions being part of the same core genome), we also compared GC content, effective codon number, sequence complexity, and COG functions between the recombined regions, non-recombined regions, and accessory elements. The latter was included as accessory elements are known to be distinct in gene function and sequence characteristics from the core genome (Kung et al. 2010). We found recombined and non-recombined regions to be highly similar and distinct from accessory elements (Fig. 4C). For example, nucleotide frequencies between recombined and non-recombined regions were similar (Chr1: χ^2= 0.0012, p-value=1; Chr2: χ^2= 2e-04, p-value=1, Chi-square test) (Supplemental Table S11), and COG analysis of all genes associated with recombination regions failed to reveal any significantly enriched biological pathways compared to the whole genome (Supplemental Table S12), indicating that general baseline recombination in Bp is not functionally selected.

Besides recombination in the core genome, the correlation between Bp clades and complements of accessory elements suggested a further test for whether recombination is ongoing in Bp. Similar to the logic for the core recombined sequences above, accessory elements that are pervasive throughout a given clade could be undergoing active, ongoing exchange (which would homogenize their sequence within the clade) or be inherited through vertical descent (and thus accumulate mutations similar to adjacent non-recombining regions) (Fig. 4A). Both of these possibilities are consistent with clade-specific complements of accessory elements (Fig. 3C). We found that within each clade, accessory elements also showed lower sequence divergence levels compared to non-accessory elements in the same strains (Fig. 4D). Thus, these results suggest that in both the core and accessory genome, there is a strong signal for ongoing, active recombination within Bp clades.

Identification of Clade-Specific RM Systems

The clade-specific pattern of haplotypes and accessory elements in Bp, coupled with evidence of ongoing gene flow within strains of the same clade, suggests that reciprocal barriers to gene exchange may exist between strains belonging to different clades. We hypothesized that these barriers might be due, at least in part, to the use of distinct restriction-modification (RM) systems in each clade. RM systems comprise different combinations of endonuclease, methylase, and DNA specificity domains that employ specific methylation patterns to label endogenous 'self' genomic DNA, while unmodified exogenous DNA is recognized as 'non-self' and subsequently cleaved and destroyed (Ershova et al. 2012; Makarova et al. 2013). Studies have proposed that RM systems can act as barriers to horizontal gene transfer(Waldron and Lindsay 2006; Hoskisson and Smith 2007; Dwivedi et al. 2013). However a role for RM systems in restricting intra-species recombination is less well described (Waldron and Lindsay 2006).

By interrogating genes in the Bp accessory genome and mobile genetic elements, we identified four different Bp RM systems (I, II, III, and IV) (Roberts et al. 2007). Notably, specific sets of RM systems were found in association with genomic clades bearing distinct haplotypes and accessory genome features. For example, while clearly related by lineage (Fig. 1), clades ST51 and ST422 exhibit different recombination patterns and sets of accessory elements – we found that ST51 strains contained RM Type IIGC genes, while ST422 strains harbored both RM Type IIGC and RM Type IC systems. Similarly, strains from genomic clade B (ST423/ST84/ST289) were largely dominated by RM Type IC and Type IBC systems, with type III RM genes additionally present in ST84 strains (Fig. 3C). The presence of these clade and ST-specific RM systems, which are predicted to result in clade-specific patterns of DNA methylation, may provide a molecular barrier to inter-clade gene sharing.

Methylome Sequencing Reveals Clade-Specific Epigenetic Profiles

To provide direct experimental data that Bp strains from different clades have distinct methylation profiles, we subjected one representative strain from Clade A (Bp35) and one strain from Clade B (Bp33) to whole-genome methylome analysis using SMRT® sequencing technology(Murray et al. 2012). Because SMRT sequencing has the ability to measure DNA polymerase activity in real time, base modifications such as methylation can be detected as a change in the kinetics of base pair incorporation (Flusberg et al. 2010; Schadt et al. 2010).

SMRT sequencing followed by *de novo* assembly for Bp33 and Bp35 was performed to obtain two circular contigs of 4.0 and 3.1 Mb (average GC content is 68%) with 240X and 147X post-filter base coverage, 21X and 22X preassembled read coverage respectively (Chin et al. 2013). We identified a 12.4 kb plasmid in Bp35 called pBp35 that to our knowledge represents the first plasmid described for Bp (Supplemental Fig. S5, Supplemental data: plasmid sequence and annotation in Genbank format). We analyzed the local sequence contexts for all the methylated bases in both the strains and identified sequence motifs associated with these methylated bases. Both Bp strains showed methylation throughout their entire genomes. In total six unique methylated motifs were identified in the two Bp strains (Table 1). Of these, one motif (5' CACAG 3') was shared by the two strains, while the other five were strains or clade-specific. For example, the type II motif (5' GTAW**A**C 3') is unique to Bp35 (Clade A representative) while the type I motif (5' GTC**A**TN$_5$TGG 3') is present only in Bp33 (Clade B representative).

We proceeded to match the different motifs to specific RM systems found in the two genomes. Reassuringly, the shared CACAG motif was found to be associated with a conserved Type III RM system found in both strains. However, Bp 35 exhibited three strain-specific methylated motifs, and these could be associated with Type I and II systems found specifically in the Bp35 clade. Similarly, Bp 33 exhibited two strain-specific methylated motifs, and these could be associated with Type I RM systems found specifically in the Bp33 clade. For both strains, the fraction of strain-specific methylated motifs was close to 100%, consistent with their predicted methylation and restriction functions operating at high efficiency (Table 1). The demonstration that strains belonging to different Bp clades indeed have distinct methylation patterns is consistent with our hypothesis that clade-specific RM activity may represent a barrier to Bp inter-clade recombination and accessory element transfer.

Bp RM Systems Impede Foreign DNA Uptake in *E. Coli*

To functionally test if clade-specific RM systems of Bp can impede the transfer of non-self DNA, we cloned and tested the Type I RM system associated with Genomic Clade A (Bp33) in *E. coli* (Janscak et al. 1999; Kasarjian et al. 2003). We engineered one plasmid to carry the Type I 'restriction' endonuclease (R$^+$), and a second separate plasmid to carry the 'specificity' and 'methylase' proteins (M$^+$) (Fig. 5A). M$^+$R$^+$ and M$^+$R$^-$ *E. coli* strains were then secondarily transformed with reporter plasmids carrying 0, 1, and 2 copies of the RM recognition site predicted from SMRT sequencing (5'GTC**A**TN_5TGG 3'; see Table 1), and efficiencies of transformation (EOTs) were calculated (Fig. 5B). We found that when transformed into M^+R^+ strains, unmethylated reporter plasmids carrying 1 or 2 recognition sites exhibited a >100-fold decrease in EOT compared to reporter plasmids with no recognition sites ($p<0.01$; Student's t-test) (Fig. 5C). Importantly, the Type I restriction endonuclease is required for this decrease, as no EOT differences were observed when the plasmids (0, 1, 2 sites) were transformed into M^+R^- strains which only express the methylase (Fig. 5C). This result indicates that the restriction endonuclease of the Clade A-specific Type I RM system is indeed active and capable of impeding the uptake of non-self DNA harboring an unmethylated Type I recognition site.

Next, we isolated plasmids with methylated recognition sites by passaging them through M^+R^- *E.coli* strains and transformed them into M^+R^+ strains. In contrast to the results using unmethylated plasmids, all three plasmids (0, 1, or 2 recognition sites) exhibited no significant EOT differences (Fig. 5D). This result indicates that, at least for one Bp clade-specific RM system, methylation of the recognition sites by RM methylases is sufficient to facilitate uptake of non-self DNA, even in the presence of its cognate restriction endonuclease.

DISCUSSION

In this study, we performed WGS on a panel of Bp strains drawn from a restricted geographic locale, to explore the contribution of gene mutation, recombination, and horizontal gene transfer to the molecular diversity of closely-related Bp isolates. We found that Bp strains can be partitioned into distinct genomic clades and that a major proportion of the Bp core genome variation is strongly influenced by both mutation and recombination. Bp diversity is further enhanced by an accessory genome component that is at least double the Bp core genome. Moreover, using diverse approaches including a) sequence diversity comparisons in both recombination and accessory regions supporting active gene flow within but not across clades; b) genome-wide methylome sequencing demonstrating clade-specific epigenetic profiles associated with distinct RM-systems, and c) experimental demonstration in a *E. coli* system that Bp RM systems are functionally active and sufficient to mediate the methylation and restriction of non-self DNA, our results point towards a model where Bp RM systems may function as a barrier to gene exchange between different Bp clades.

Phylogenetic analysis of the Bp clades revealed that they comprised mixtures of Bp isolates from animal, clinical, and environmental sources, arguing against the existence of a genetically distinct population of Bp capable of infecting humans. Supporting this model, in a separate analysis, we were unable to confidently identify a consistent set of signature genetic changes in strains associated with human disease (TN, unpublished observations). The genetic similarity between clinical, animal, and environmental Bp strains raises the possibility that additional genetic changes may not be required for an environmental Bp strain to successfully cause human disease. This model is consistent with previous proposals that Bp is an "accidental pathogen", where adaptations incurred by Bp to survive in its natural reservoir (soil and potentially single-celled organisms located therein e.g. Amoebae) must have indirectly contributed to its ability to colonize a mammalian host (Casadevall and Pirofski 2007; Nandi et al. 2010). This "accidental virulence" hypothesis is further supported by epidemiological data where patients with clinical melioidosis often possess pre-infection morbidities such as diabetes, which may contribute to a weakened host immune response(Currie et al. 2010).

Our data revealed several genome-wide features of the Bp recombination landscape. We found that recombination in Bp is pervasive, approaching levels previously reported for *S. pneumonia* (Croucher et al. 2011), and frequently involved defined sets of haplotypes.

Importantly, analysis of genes in regions associated with high recombination suggests that haplotypes and recombination hot-spots in Bp are not randomly distributed, but biased towards genomic regions associated with niche adaption, survival and virulence. This included a *TTSS* cluster involved in intracellular survival of Bp and virulence (Stevens et al. 2002) and a type IVB pilus cluster (*TFP8*) that we have validated in this study as required for maximal virulence in murine infection assays. It is thus possible that the other regions of high recombination identified in this study may contain additional genes involved in Bp adaptation, survival, and virulence.

Previous studies have shown that strains of different STs can be frequently co-isolated in the wild(Wuthiekanun et al. 2009). However, it is not known if such co-isolated strains are able to engage in an unrestricted exchange and transfer of genetic material. We found that strains associated with different Bp genomic clades tended to exhibit distinct sets of recombination haplotypes and accessory elements. These findings suggest that Bp clade/ST subgroups may represent a functional and potentially limiting unit of within-species genetic diversity. However, it is important to note that while our findings suggest a general scenario of recombination events being largely clade-specific, exceptions do exist. One example of a possible recombination across separate clades involved a heme/porphyrin locus (*BPSS1245-BPSS1247*) where a haplotype present in strain P171/04 from ST84 (but not other ST84 strains) was highly similar to the haplotypes of ST51 strains (Supplemental Fig. S6). One explanation for these exceptions is that while barriers to inter-clade gene exchange do exist, they may be incomplete or perhaps only recently established.

Two broad models of Bp evolution could explain this clade-restriction pattern. First, clade restriction might result from effective physical or niche separation, where strains from different clades are adapted or restricted to distinct niches in the environment, and therefore do not share DNA. However, as mentioned above strains of different STs can be co-isolated, and while this does not rule out the presence of micro-scale niche differences between strains, such "micro-niches" remain to be experimentally proven. Alternatively, the discovery of clade-specific RM systems provides an epigenetic explanation for Bp clade-restriction. In this model, acquisition of a diversity of RM systems to combat invading DNA may thus have occurred as a primary event, and the resulting epigenetic differences may in turn have established barriers to intra-clade DNA exchange. A synthesis of these two models is also possible, where early subspeciation is initially driven by epigenetic barriers and followed subsequently by traditional niche selection. Consistent with this model, we observed that amongst ST84 strains, the two Malaysian Bp strains EY2 and EY5 clustered separately from the remaining seven Singapore strains, being separated by 13 L-SNPs mapping to 12 protein coding genes. This scenario could be explained by initial RM-driven epigenetic isolation of the ST84 clade, followed by geographical separation between the Singapore and Malaysian ST84 strains. The presence of localized concentrations of nonsynonymous changes suggests the possible existence of specific selective pressures driving further divergence, and it is also possible that the driver for speciation in Bp is currently shifting from divergence due to genetic/epigenetic separation to divergence due to selection in different niches. As such, our results provide a potential snapshot of early incipient speciation in a microorganism associated with diverse genomes and habitats.

METHODS

Ethics Statement

This research was approved by the GIS Institutional Review Board. Animal studies were performed in accordance with the UK Scientific Procedures Act (Animals) 1986 and UK Codes of Practice for the Housing and Care of Animals Used in Scientific Procedures, 1989.

Bacterial Strains, Plasmids and Primers

Strains used were obtained from DMERI, DSO National Laboratories. These include: a) 56 clinical isolates from melioidosis patients between 1996 and 2004, b) 34 animal isolates from various species (eg monkeys, pigs, birds, and dogs) diagnosed with melioidosis between 1996 and 2005, and c) 16 soil isolates from 1996 to 2000 (Supplemental Table S1). The isolates were sampled from a diversity of locations and not a single site (Aw Lay Tin personal communication). For *E. coli* experiments, plasmids bearing predicted recognition sequences for the Clade A-specific Type I RM system (5' GTC**A**TN$_5$TGG 3') were generated by PCR mediated insertion (see Supplemental Methods). Gene sequences encoding the Bp33 Type I Restriction Modification system proteins (*'specificity'*, *'methylase'* and *'restriction endonuclease'*) were cloned into expression vectors driven by a T5 inducible promoter (DNA2.0, Singapore). All plasmids were propagated in *E. coli* strain K12 MG1655. Bacterial strains, plasmids and primers are listed in Supplemental Table S13.

Genomic DNA Extraction and Multiplex Sequencing

Live bacteria were grown in a BioSafety Level 3 facility in DSO National Laboratories. Genomic DNA was extracted using the Qiagen Genomic Tip 500/G kit (Qiagen, Valencia, CA). Unique index-tagged libraries for each sample were created, and up to 33 separate libraries were sequenced per lane on an Illumina HiSeq instrument with 100 base paired-end reads. Libraries were constructed using an amplification-free method (Kozarewa et al. 2009). Raw Illumina data were split to generate paired-end reads, and assembled using a *de novo* genome-assembly program, Velvet v0.7.03(Zerbino and Birney 2008), to generate a multi-contig draft genome for each Bp isolate.

Gene Annotation, SNP, and Phylogenetic Analysis

Paired-end reads were mapped against the chromosomes of *B. pseudomallei* K96243 (accession numbers BX571965 and BX571965) (Holden et al. 2004). Bp genes were predicted using FGENESB (www.softberry.com). Gene orthologs were determined using OrthoMCL (Chen et al. 2006). RM systems were inferred based on the specificity sequences of homologs in REBASE(Roberts et al. 2007) and categorized into subtypes—IC, IBC, IIG, IIGC, IIP, and IIB, on the basis of their genetic organization, mode of action, recognition sites and cleavage loci (Roberts et al. 2003) .SNPs predicted to have arisen by homologous recombination were identified using Gubbins and excluded from phylogenetic reconstruction (Croucher et al. 2011). Indels where identified using Dindel (Albers et al. 2011). Maximum likelihood phylogenies were constructed using RAxML v0.7.4(Stamatakis et al. 2005). SNPs Ancestral and derived alleles (polarization) were determined according to the outgroup reference strain sequence.

Recombination Analysis

The general time-reversible model with gamma correction was used for among-site rate variation for ten initial random trees. To measure clade-specific recombination rates, ClonalFrame (Didelot and Falush 2007) was applied separately to each Bp clade. To reduce mapping artefacts, we focused on the 5.6 Mb portion of the core genome that excludes mobile genetic elements and other potentially biased regions such as surface polysaccharides, secretion systems, and tandem repeats. Recombination events were extracted from ClonalFrame as genomic fragments where the probability of recombination for a given branch of the tree was consistently above 50% and reached 95% in at least one location. Potential origins of recombination imports were investigated as previously described (Didelot et al. 2009; Didelot et al. 2011; Sheppard et al. 2013). To determine haplotypes, SNP alleles at the recombination loci were concatenated to give a single haplotype string for each strain. The aligned strings were then subjected to hierarchical clustering as implemented in R package ‘hclust’. The resulting dendrogram was used to assign strains to distinct haplotype groups using the ‘cutree’ function in R. Within-clade sequence diversity comparisons between recombined and non-recombined regions were performed as described in Supplemental Fig. S3 and the Supplementary Text. Potential differences in sequence composition between recombined and non-recombined regions were assessed using Artemis release 13.0 (Carver et al. 2012), or the K2 algorithm in CLC Main workbench 6.5 (www.clcbio.com) (Wootton and Federhen 1993).

Bp Mutagenesis and Mouse Virulence Studies

Isogenic Bp mutants carrying a 12.9 kb deletion *TFP8* were generated in a two-step process as previously described (Essex-Lopresti et al. 2005; Boddey et al. 2006) (see Supplemental Methods). Virulence of wild-type and mutant Bp strains were assessed using an intranasal BALB/c mouse model (Essex-Lopresti et al. 2005). Briefly, groups of six age-matched BALB/c female mice were anesthetized and infected intra nasally with 10-fold dilutions (10^1–10^6) of either wild-type Bp K96243 or *TFP8* deletion strains grown overnight at 37 °C with shaking. Mice were recovered and survival was recorded for up to 51 days. Analysis was performed using the Mantel-Haenszel log rank test in GraphPad Prism 4 or by Regression with Life Data in MiniTAB v13.0, using a significance threshold of P= 0.05.

Accessory genome analysis

Nucmer (Kurtz et al. 2004) was used to generate alignments of Velvet contigs against the reference strain Bp K96243 to identify novel accessory regions (N_{AE}). N_{AE} values for individual Bp strains were defined as blocks with a minimal 1000 bp length that was absent in Bp K96243 (median N_{AE} per strain = 183482 bp). Sequence diversity comparisons between accessory and non-accessory regions utilized accessory regions >1,000 bp, and performed using MUMmer 3.20 under DNAdiff default settings (Kurtz et al. 2004).

SMRT Sequencing and Data Analysis

20 ug of gDNA was processed to create SMRTbell sequencing templates greater than 10 kb (average insert size 17 kb) and sequenced using a PacBio RSII System where polymerase-MagBead bound templates were loaded at an on-plate concentration of 150 pM. Templates were subsequently sequenced using DNA Sequencing Kit 2.0, with data collection of 180 mins (Pacific Biosciences, Menlo Park, CA. USA). Genomes were assembled using HGAP (Chin et al. 2013) with default parameters in SMRT Analysis Suite version 2.1 (Pacific Biosciences, Menlo Park, CA, USA). Additional manual assembly of contigs was carried out in cases of unique overlapping sequence. Consensus sequence polishing was done using the Quiver

algorithm in Genomic Consensus version 0.7.0. Base modification analysis was performed by mapping SMRT sequencing reads to the respective assemblies using the BLASR mapper (Chaisson and Tesler 2012) and SMRT Analysis Suite version 2.1 using standard mapping protocols. Clustering of sequence motifs was performed using Motif Finder (https://github.com/PacificBiosciences/DevNet/wiki/Motiffinder). See Supplemental Methods for further details.

Restriction-modification assay

Plasmids containing methylated and unmehylated Type I restriction sites were transformed into *E. coli* strains engineered to express all three proteins of the Type I RM system, or only the specificity and methylase units. Efficiency of transformation (EOT) values were computed by comparing bacterial titres (colony forming units per ml, cfu/ml) on antibiotic selection plates divided by the corresponding titres from LB plates. EOT values were log transformed and plotted for analysis of RM system restriction activity. EOT values from triplicate experiments were compared using a 2-tailed Student's t-test. See Supplemental Methods for further details.

Statistical Analysis

All statistical analyses were performed using R-2.15.1(Ihaka 1996).

DATA ACCESS

The data from this study have been submitted to the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena) under accession number ERP000251. Methylation data has been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE55168.

ACKNOWLEDGEMENTS

This study was supported by a core grant from the GIS, an A-STAR research institute, to PT. The sequencing of the *Burkholderia pseudomallei* strains was supported by Wellcome Trust grant 098051 to JP.

DISCLOSURE DECLARATION

The authors declare no competing financial interests associated with this study.

FIGURE LEGENDS

Figure 1: Whole-genome phylogeny and sequence variation of Bp strains.

A. Global phylogeny of Bp strains. The maximum likelihood tree was constructed using SNPs not associated with recombination events (see Results). Tip labels are colored according to the geographic locations of isolation (Red: Singapore, Green: Malaysia, Blue: Thailand, Black: Unknown, Pink: Imported to UK). Inset bars at right indicate the MLST scheme (Blue: ST51, Cyan: ST422, Red: ST423, Dark Green: ST84, Pink: ST289, Light green: ST46). Three major genomic clades are identified: Clade A (ST51, ST422, ST414, ST169, nST4), Clade B (ST423, ST84, ST289, nST5) and Clade C (ST46, ST50).
B. Intra-ST subgroups resolved by WGS. ST51 strains cluster into two groups: ST51a and ST51b. Genomic locations of 342 L-SNPs (including both intra- and intergenic SNPs) distinguishing ST51a and ST51b are shown. The top and the bottom panel with four rows shows SNPs exclusively present in the two groups ST51a$^+$ST51b$^-$ and ST51a$^+$ST51b$^-$ respectively.
C. Mutation spectra of ST51a: Relative rates of six possible mutation categories. The most common mutations are C/G → T/A transitions.
D. Fraction of the three classes of cytosine mutations occurring at CG dinucleotides in the Bp genome, compared with the expected fraction based on the average of 100 simulated genomes of the same size and composition (grey).

Figure 2: Recombination Landscape of Bp

A. Recombination hotspots in Bp. Circles: (outside) genome coordinates; (middle) compositionally biased regions identified by Alien hunter (Vernikos and Parkhill 2006) (green) and Bp core genome (violet); (innermost) regions of elevated recombination (height of red bars). Note that recombination levels are higher on Chr II than Chr I. Location of the *TFP8* and *TTSS3* clusters are indicated.

B. Local recombination events in the Type III secretion system and Type IVB pilus cluster. (top) Genomic coordinates and location of protein coding genes. (dark blue) Predicted recombination events (R1 to Rn, n=number of recombination events) observed in Bp strains belonging to genomic clades (ST group 46, 51, 84, 289, 422 and 423). The recombination boundaries are indicated by the dark blue circles and the boundaries that fall beyond the depicted locus are shown as open ended.

C. Relative virulence of *TFP8* deletion mutant. Graphs show survival curves of BALB/c mice following intranasal challenge with varying dosages of Bp (left – K96243 wild-type, right – *TFP8* deletion mutant, units are colony forming units, CFU). See Methods for infection assay details. The *TFP8* deletion mutant is significantly less virulent compared to Bp K96243 parental controls (p = 0.026, Mantel-Haenszel log rank test).

D. Distinct haplotypes at the *TFP8* genomic locus: Each row represents an individual Bp strain arranged according to genomic clade/ST (shown on left, color bars indicating ST51 (blue), ST289 (pink), ST422 (cyan), ST423 (red), ST84 (dark green), and ST46 (light green)). Across each row (strain), SNP positions are ordered by genomic coordinate (top numbers, Bp Chr II, genomic locus 2935860 – 2976718), and color-coded according to nucleotide identity (A→ green; T→ blue; C→ orange; and G→ red). The right y-axis "Haplotypes" refers to the specific linear combination of SNPs exhibited by individual strains. In some cases, haplotypes can be composed of a specific combination of smaller recombination regions (R). For example, Haplotype H1 is composed of recombination regions R2, R7, and R8. Haplotype alignments were generated using clustalx (Larkin et al. 2007).

Figure 3: Accessory Genome Landscape of Bp

A. Accumulation curves for Bp novel accessory genes (blue). Vertical bars represent standard deviation values based upon one hundred randomized input orders of different Bp STs. The total number of accessory genes is indicated by the red dotted line.

B. Functional enrichment of Bp accessory genes. COG functional categories are indicated on the y-axis, and the percentage of genes in each COG category is shown on the x-axis. Dark blue columns represent novel accessory genes, and light blue columns indicate all Bp core genes with COG annotations. COG categories exhibiting a significant enrichment among the Bp accessory genes are highlighted by asterisks (*p < 0.0005, binomial test; after Bonferroni correction). The COG category "DNA replication, recombination and repair" was excluded as it was represented by mainly mobility genes particularly transposases and integrases.

C. Distribution of accessory elements across Bp clades. The heatmap represents an all-pairwise strain comparison showing the degree of accessory element overlap between pairs of strains. Strains are arranged on the x- and y-axis according to their genomic clades and sequence types (ST51 (blue), ST289 (pink), ST422 (cyan), ST423 (red), ST84 (dark green) and ST46 (light green)). The color scale bar at the bottom indicates the degree of accessory element sharing (more blue = increased sharing). The right-hand chart depicts the different types of restriction-modification (RM) systems associated with different clades. In each column the RM systems are color-coded based on their encoded protein coding sequences. In the first column the bars in green and blue refer to two distinct sets of RM genes that belong to the Type IC RM systems. Strain specific RM systems are in black.

Figure 4: Distinguishing between Ongoing Recombination and Vertical Descent and in Bp.

A. Alternative models for clade-specific recombination haplotypes. (Left) In the *Ongoing Recombination* model, an imported fragment sweeps through the population via recombination, resulting in homogenization of the recombining fragment across strains. The recombining fragment should show lower levels of sequence diversity compared to non-imported regions. (Right) In the *Vertical Descent* model, an ancestral strain acquires a genomic fragment (yellow) from an external strain and subsequently transmits that fragment to all daughter strains in a clonal fashion. In this model, the imported fragment should accumulate new point mutations (green bars) at a similar rate to non-imported regions.

B. Within-clade sequence diversity of recombined regions compared to non-recombined regions. Scatter plots comparing within-clade sequence diversity values of individual recombined regions (x-axis) to non-recombined regions (y-axis) for the same strains in a given clade. Sequence diversity decreases in the direction of the red arrows (to right and upwards). (*) Data points highlighted by the red bar correspond to recombined regions exhibiting 100% sequence identity. To visualize these points in a manner that captures both their density and extremely low sequence diversity, these were plotted within the x-axis range 6.9 -7.6 on a negative log scale. Sequence diversity is defined as the number of SNPs per kb.

C. Sequence features of non-recombined regions (NR), recombined regions (R), and accessory elements (AE). (Top) Bp K96243 genomic tracks of Chr1 and Chr2. Row 1: Genomic locations of recombined regions (red). Row 2: Genomic locations of 16 known Bp genomic islands (gray). (Bottom) Sequence feature comparison of genes in non-recombined (white; NR), recombined (red; R), and accessory elements (gray; AE): i) GC content (Puigbo et al. 2008) ii) effective codon number (Puigbo et al. 2008) and iii) sequence complexity (Pietrokovski et al. 1990). Each hourglass plot spans the 25th to 75th percentile (interquartile range, IQR) of all genes in that category, with the bottleneck at the median. Horizontal tick marks show data ranges within 1.5 X IQR of 25th and 75th percentiles. Open circles represent outliers outside this range. The width of the bottleneck (i.e., the length of the V-shaped notch) depicts the 95% confidence interval for the median.

D. Within-clade sequence diversity of accessory elements compared to non-accessory elements. Accessory elements are defined as regions not present in the BpK96243 reference strain (see Methods). Scatter plots compare average sequence diversity values for individual accessory elements (x-axis) to corresponding non accessory elements (y-axis) for the same strain pairs in a given clade. Sequence diversity is defined as the number of SNPs per kb.

Figure 5 : Restriction of Non-self DNA by Clade-Specific Bp RM Systems

A. Molecular cloning of a Type I RM system specific to Bp genomic clade A. The RM system comprises three genes - 'restriction' (R), 'methylase' (M), 'specificity' (S). Genes S (yellow) and M (blue) were cloned in plasmid pSLC-279 with kanamycin resistance (Km^R) to give the M^+ plasmid. Gene R (red) was cloned in plasmid pSLC-280 with ampicillin resistance (Ap^R) to give the R^+ plasmid. Resistance genes are depicted in cyan. Green represents the T5 promoter used to induce expression of the cloned genes. Plasmids are not drawn to scale.

B. Efficiency of Transformation (EOT) assay. Reporter plasmids p0, p1, and p2 harbor 0, 1, and 2 copies of the predicted Type I recognition site (5'-GTC**A**TN_5TGG-3'; indicated by green triangles). Plasmid p0 should not show any EOT changes as it does not contain Type I recognition sequences. Unmethylated plasmids p1 and p2, when transformed into M^+R^+ strains, should be recognized via their Type I sites and cleaved by the Type I restriction enzyme (registered as a drop in number of transformants). However, when transformed into M^+R^- strains that express the methyltransferase alone, no EOT differences should be observed. In contrast, methylated p1 and p2 plasmids (obtained by passage through M^+R^- strains; methylated sites indicated by red stars and superscript ^{m}), when transformed into M^+R^+ strains should be recognized as 'self' DNA by the Type I system and resist cleavage, resulting in minimal EOT changes.

C. EOT Assay results using unmethylated plasmids. Host strains are: MG1655 (M^-R^-, no RM system, cyan); SLC-623 (M^+R^+, complete RM system, red); and SLC-621 (M^+R^-, methyltransferase only, green). Reporter plasmids are: pACYC184 (p0, control plasmid); pSLC-277 (p1, 1 recognition site) and pSLC-278 (p2, 2 recognition sites). Significant differences in EOT are observed between control plasmid p0 and plasmids p1 and p2 when transformed into M^+R^+ strains ($p<0.01$) but not in host *E. coli* or M^+R^- strains. EOT in this study is the normalized number of Cm^R transformants obtained per unit amount of plasmid DNA.

D. EOT assay using methylated plasmids. Reporter plasmids were passaged through M^+R^- strains prior to transformation, which is predicted to cause recognition site methylation. No significant EOT differences are observed across the strains. All experiments were performed in triplicate, and data are presented as means and standard deviations. Data are presented as $\log_{10}$ values of EOT. Student's t-test was used to test for significant differences.

Table 1: DNA Methylation Sequence Motifs in Bp35 (Clade A) and Bp33 (Clade B)

Type of RM system	Methyltransferase activity[a]	Type of methylation	Total number of sites[b]	Number of methylated sites	% sites methylated	Assignment	Locus	Reference
				Bp35 Strain				
type I	5' G**A**TCN$_5$GATG 3' 3' CTAGN$_5$CT**A**C 5'	^{m6}A	3086	3082	99.87	--	--	this study
type II	5' GTAW**A**C 3' 3' C**A**TWTG 5'	^{m6}A	1152	1141	99.05	--	--	this study
	5' C**A**GN$_6$CTG 3' 3' GTCN$_6$G**A**C 5'	^{m6}A	5214	5197	99.67	--	--	this study
type III	5' CAC**A**G 3'	^{m6}A	4584	4572	99.74	BceJI	BURCENBC7_AP5195	REBASE
				Bp33 Strain				
type I	5' CC**A**TN$_7$CTTC 3' 3' GGTAN$_7$G**A**AG 5'	^{m6}A	86	86	100	--	--	this study
	5' GTC**A**TN$_5$TGG 3' 3' CAGTAN$_5$**A**CC 5'	^{m6}A	211	210	99.53	--	--	this study
type III	5' CAC**A**G 3'	^{m6}A	5214	5197	99.67	BceJI	BURCENBC7_AP5195	REBASE

[a] The methylated position within the motif is highlighted in bold. Pairs of reverse-complementary motifs belonging to the same recognition sequence were grouped together.

[b] The total number includes motifs occurring on both the '+'- and '–'-strands.

REFERENCES

Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R. 2011. Dindel: accurate indel calls from short-read data. *Genome Res* **21**(6): 961-973.

Boddey JA, Flegg CP, Day CJ, Beacham IR, Peak IR. 2006. Temperature-regulated microcolony formation by *Burkholderia pseudomallei* requires pilA and enhances association with cultured human cells. *Infect Immun* **74**(9): 5374-5381.

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**(4): 464-469.

Casadevall A, Pirofski LA. 2007. Accidental virulence, cryptic pathogenesis, martians, lost hosts, and the pathogenicity of environmental microbes. *Eukaryot Cell* **6**(12): 2169-2174.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**: 238.

Chantratita N, Rholl DA, Sim B, Wuthiekanun V, Limmathurotsakul D, Amornchai P, Thanwisai A, Chua HH, Ooi WF, Holden MT et al. 2011. Antimicrobial resistance to ceftazidime involving loss of penicillin-binding protein 3 in *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **108**(41): 17165-17170.

Chantratita N, Wuthiekanun V, Limmathurotsakul D, Vesaratchavest M, Thanwisai A, Amornchai P, Tumapa S, Feil EJ, Day NP, Peacock SJ. 2008. Genetic diversity and microevolution of *Burkholderia pseudomallei* in the environment. *PLoS Negl Trop Dis* **2**(2): e182.

Chen F, Mackey AJ, Stoeckert CJ, Jr., Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* **34**(Database issue): D363-368.

Cheng AC, Ward L, Godoy D, Norton R, Mayo M, Gal D, Spratt BG, Currie BJ. 2008. Genetic diversity of *Burkholderia pseudomallei* isolates in Australia. *J Clin Microbiol* **46**(1): 249-254.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**(6): 563-569.

Craig L, Pique ME, Tainer JA. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**(5): 363-378.

Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS et al. 2011. Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**(6016): 430-434.

Cruz-Migoni A, Hautbergue GM, Artymiuk PJ, Baker PJ, Bokori-Brown M, Chang CT, Dickman MJ, Essex-Lopresti A, Harding SV, Mahadi NM et al. 2011. A *Burkholderia pseudomallei* toxin inhibits helicase activity of translation factor eIF4A. *Science* **334**(6057): 821-824.

Currie BJ, Fisher DA, Howard DM, Burrow JN. 2000. Neurological melioidosis. *Acta Trop* **74**(2-3): 145-151.

Currie BJ, Ward L, Cheng AC. 2010. The epidemiology and clinical spectrum of melioidosis: 540 cases from the 20 year Darwin prospective study. *PLoS Negl Trop Dis* **4**(11): e900.

Didelot X, Barker M, Falush D, Priest FG. 2009. Evolution of pathogenicity in the *Bacillus cereus* group. *Syst Appl Microbiol* **32**(2): 81-90.

Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D et al. 2011. Recombination and population structure in *Salmonella enterica*. *PLoS Genet* **7**(7): e1002191.

Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**(3): 1251-1266.

Dwivedi GR, Sharma E, Rao DN. 2013. *Helicobacter pylori* DprA alleviates restriction barrier for incoming DNA. *Nucleic Acids Res* **41**(5): 3274-3288.

Ershova AS, Karyagina AS, Vasiliev MO, Lyashchuk AM, Lunin VG, Spirin SA, Alexeevski AV. 2012. Solitary restriction endonucleases in prokaryotic genomes. *Nucleic Acids Res* **40**(20): 10107-10115.

Essex-Lopresti AE, Boddey JA, Thomas R, Smith MP, Hartley MG, Atkins T, Brown NF, Tsang CH, Peak IR, Hill J et al. 2005. A type IV pilin, PilA, Contributes To Adherence of *Burkholderia pseudomallei* and virulence in vivo. *Infect Immun* **73**(2): 1260-1264.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7**(6): 461-465.

Hayden HS, Lim R, Brittnacher MJ, Sims EH, Ramage ER, Fong C, Wu Z, Crist E, Chang J, Zhou Y et al. 2012. Evolution of *Burkholderia pseudomallei* in recurrent melioidosis. *PLoS One* **7**(5): e36507.

Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD et al. 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**(39): 14240-14245.

Hoskisson PA, Smith MC. 2007. Hypervariation and phase variation in the bacteriophage 'resistome'. *Curr Opin Microbiol* **10**(4): 396-400.

Howard K, Inglis TJ. 2003. Novel selective medium for isolation of *Burkholderia pseudomallei*. *J Clin Microbiol* **41**(7): 3312-3316.

Ihaka RG, Robert. 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics (American Statistical Association)* **5**(3): 299–314.

Janscak P, MacWilliams MP, Sandmeier U, Nagaraja V, Bickle TA. 1999. DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J* **18**(9): 2638-2647.

Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS. 2012. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nat Commun* **3**: 886.

Kasarjian JK, Iida M, Ryu J. 2003. New restriction enzymes discovered from *Escherichia coli* clinical strains using a plasmid transformation method. *Nucleic Acids Res* **31**(5): e22.

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* **6**(4): 291-295.

Kung VL, Ozer EA, Hauser AR. 2010. The accessory genome of *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev* **74**(4): 621-641.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5**(2): R12.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948.

Larsen E, Smith JJ, Norton R, Corkeron M. 2013. Survival, sublethal injury, and recovery of environmental *Burkholderia pseudomallei* in soil subjected to desiccation. *Appl Environ Microbiol* **79**(7): 2424-2427.

Li L, Lu W, Han Y, Ping S, Zhang W, Chen M, Zhao Z, Yan Y, Jiang Y, Lin M. 2009. A novel RPMXR motif among class II 5-enolpyruvylshikimate-3-phosphate synthases is required for enzymatic activity and glyphosate resistance. *J Biotechnol* **144**(4): 330-336.

Makarova KS, Wolf YI, Koonin EV. 2013. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* **41**(8): 4360-4377.

Mayo M, Kaesti M, Harrington G, Cheng AC, Ward L, Karp D, Jolly P, Godoy D, Spratt BG, Currie BJ. 2011. *Burkholderia pseudomallei* in unchlorinated domestic bore water, Tropical Northern Australia. *Emerg Infect Dis* **17**(7): 1283-1285.

Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. 2012. The methylomes of six bacteria. *Nucleic Acids Res* **40**(22): 11450-11462.

Nandi T, Ong C, Singh AP, Boddey J, Atkins T, Sarkar-Tyson M, Essex-Lopresti AE, Chua HH, Pearson T, Kreisberg JF et al. 2010. A genomic survey of positive selection in *Burkholderia pseudomallei* provides insights into the evolution of accidental virulence. *PLoS Pathog* **6**(4): e1000845.

Pietrokovski S, Hirshon J, Trifonov EN. 1990. Linguistic measure of taxonomic and functional relatedness of nucleotide sequences. *J Biomol Struct Dyn* **7**(6): 1251-1268.

Pitt TL, Trakulsomboon S, Dance DA. 2007. Recurrent melioidosis: possible role of infection with multiple strains of *Burkholderia pseudomallei*. *J Clin Microbiol* **45**(2): 680-681.

Priestman MA, Funke T, Singh IM, Crupper SS, Schonbrunn E. 2005. 5-Enolpyruvylshikimate-3-phosphate synthase from *Staphylococcus aureus* is insensitive to glyphosate. *FEBS Lett* **579**(3): 728-732.

Puigbo P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* **3**: 38.

Relman DA, Domenighini M, Tuomanen E, Rappuoli R, Falkow S. 1989. Filamentous hemagglutinin of *Bordetella pertussis*: nucleotide sequence and crucial role in adherence. *Proc Natl Acad Sci U S A* **86**(8): 2637-2641.

Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev S, Dryden DT, Dybvig K et al. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* **31**(7): 1805-1812.

Roberts RJ, Vincze T, Posfai J, Macelis D. 2007. REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* **35**(Database issue): D269-270.

Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Hum Mol Genet* **19**(R2): R227-240.

Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, Kelly DJ, Cody A, Colles FM, Strachan NJ et al. 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Mol Ecol* **22**(4): 1051-1064.

Sim SH, Yu Y, Lin CH, Karuturi RK, Wuthiekanun V, Tuanyok A, Chua HH, Ong C, Paramalingam SS, Tan G et al. 2008. The core and accessory genomes of *Burkholderia pseudomallei*: implications for human melioidosis. *PLoS Pathog* **4**(10): e1000178.

Sprague LD, Neubauer H. 2004. Melioidosis in animals: a review on epizootiology, diagnosis and clinical presentation. *J Vet Med B Infect Dis Vet Public Health* **51**(7): 305-320.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**(4): 456-463.

Stevens MP, Wood MW, Taylor LA, Monaghan P, Hawes P, Jones PW, Wallis TS, Galyov EE. 2002. An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol Microbiol* **46**(3): 649-659.

Tang DJ, He YQ, Feng JX, He BR, Jiang BL, Lu GT, Chen B, Tang JL. 2005. *Xanthomonas campestris pv. campestris* possesses a single gluconeogenic pathway that is required for virulence. *J Bacteriol* **187**(17): 6231-6237.

Vernikos GS, Parkhill J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics* **22**(18): 2196-2203.

Waldron DE, Lindsay JA. 2006. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *J Bacteriol* **188**(15): 5578-5585.

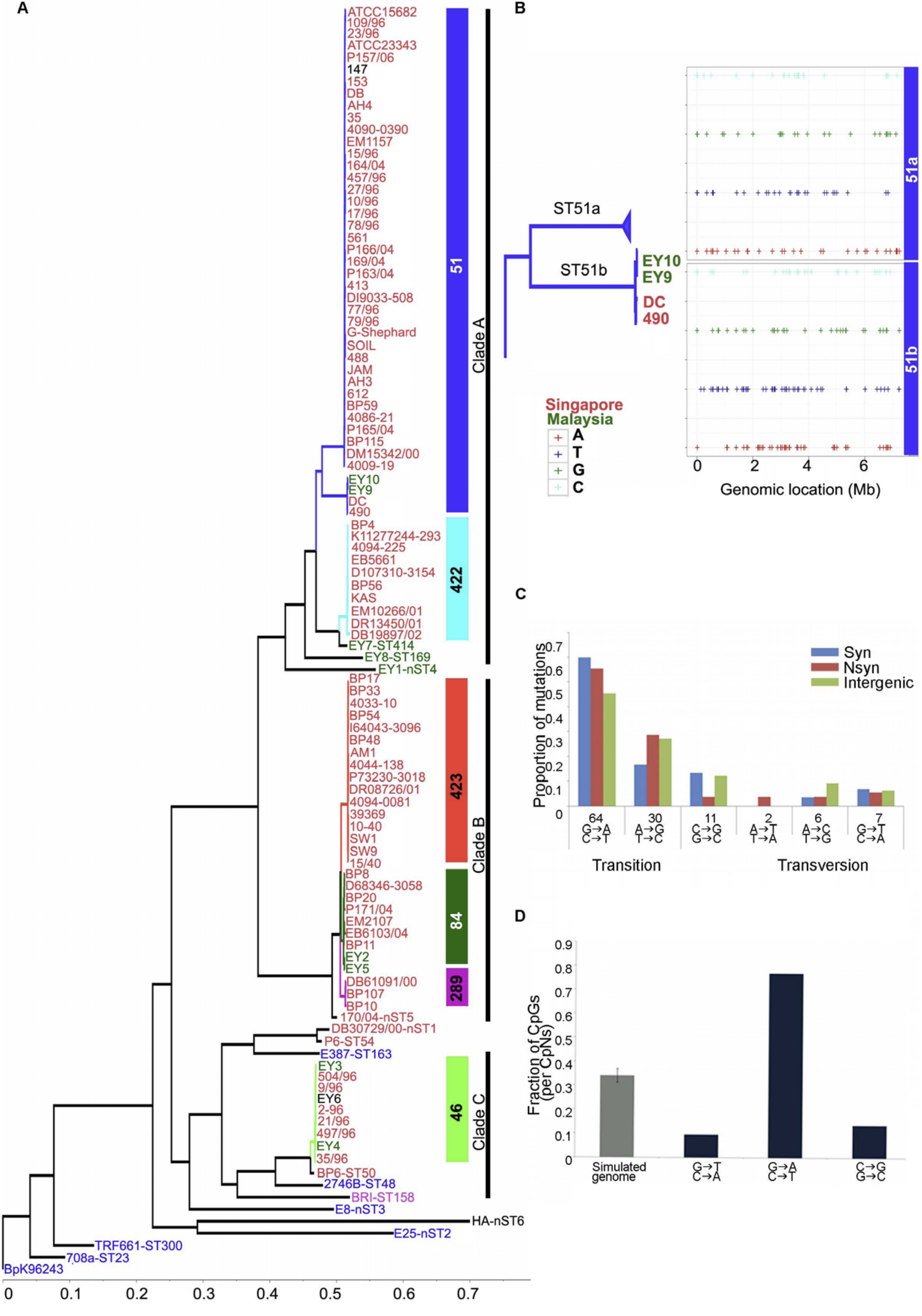
Wiersinga WJ, Currie BJ, Peacock SJ. 2012. Melioidosis. *N Engl J Med* **367**(11): 1035-1044.

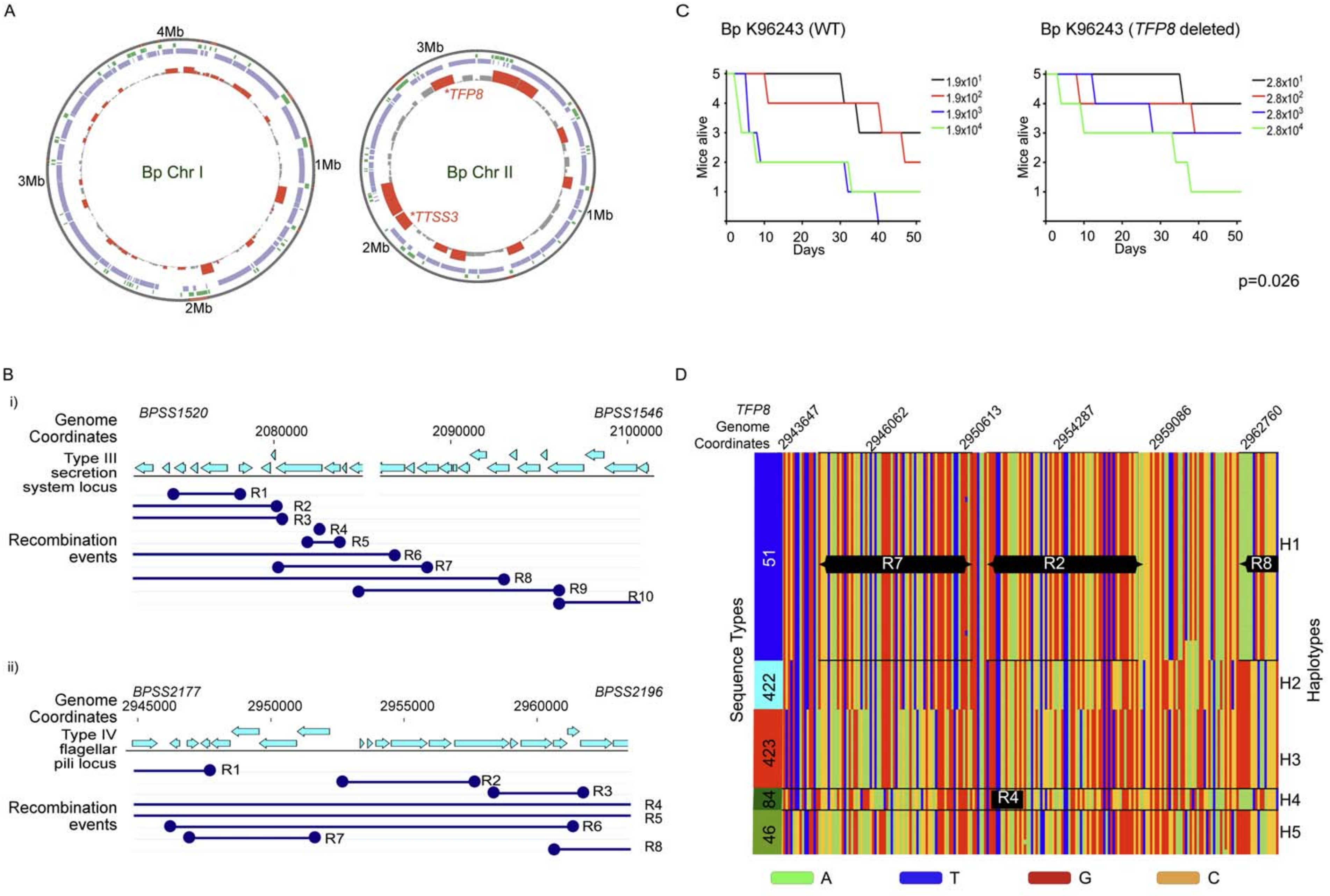
Wootton JC, Federhen S. 1993. Statistics of Local Complexity in Amino Acid Sequences and Sequence Databases. *Comput Chem* **17**: 149-163.

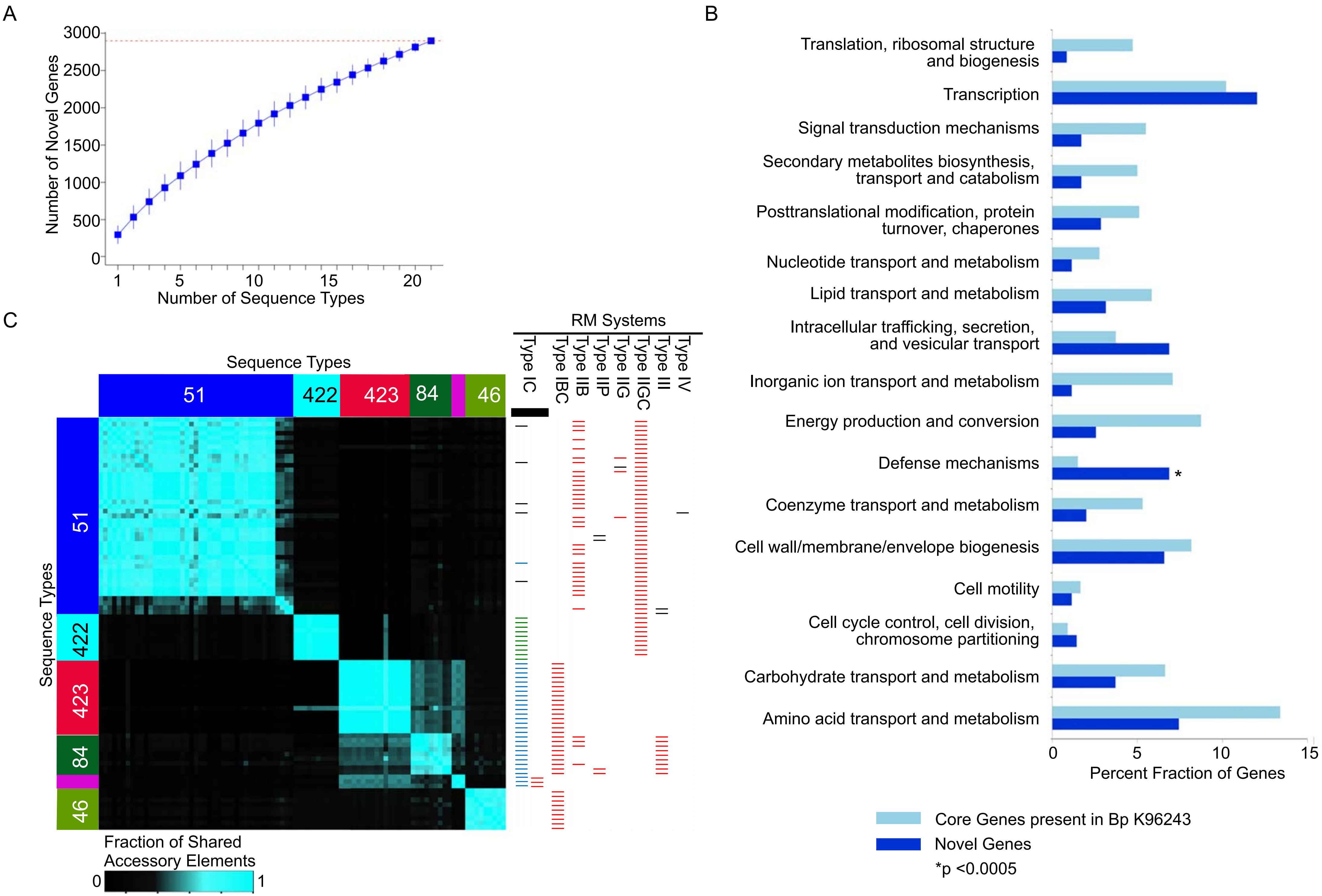
Wuthiekanun V, Limmathurotsakul D, Chantratita N, Feil EJ, Day NP, Peacock SJ. 2009. *Burkholderia pseudomallei* is genetically diverse in agricultural land in Northeast Thailand. *PLoS Negl Trop Dis* **3**(8): e496.

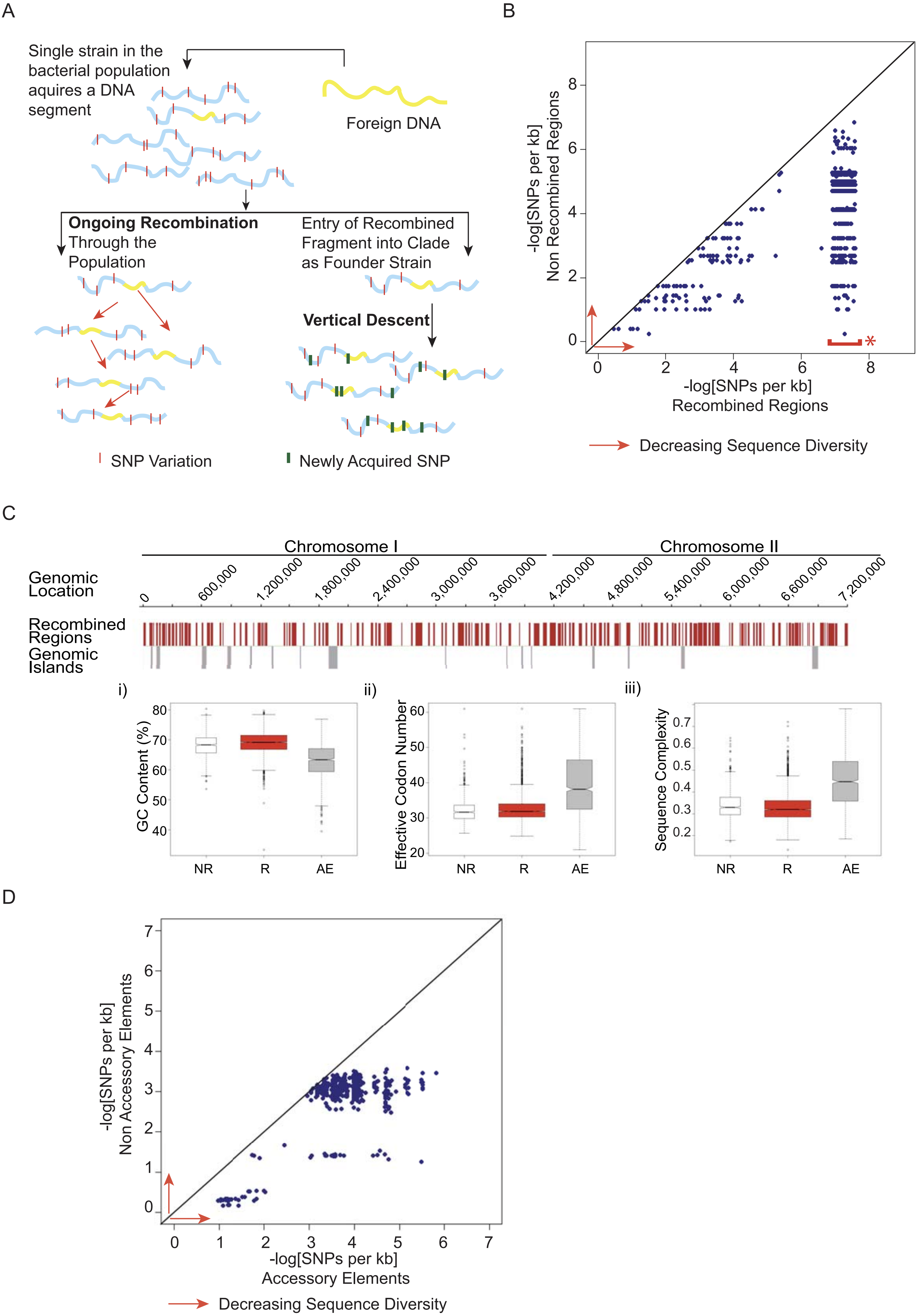
Wuthiekanun V, Smith MD, Dance DA, White NJ. 1995. Isolation of *Pseudomonas pseudomallei* from soil in north-eastern Thailand. *Trans R Soc Trop Med Hyg* **89**(1): 41-43.

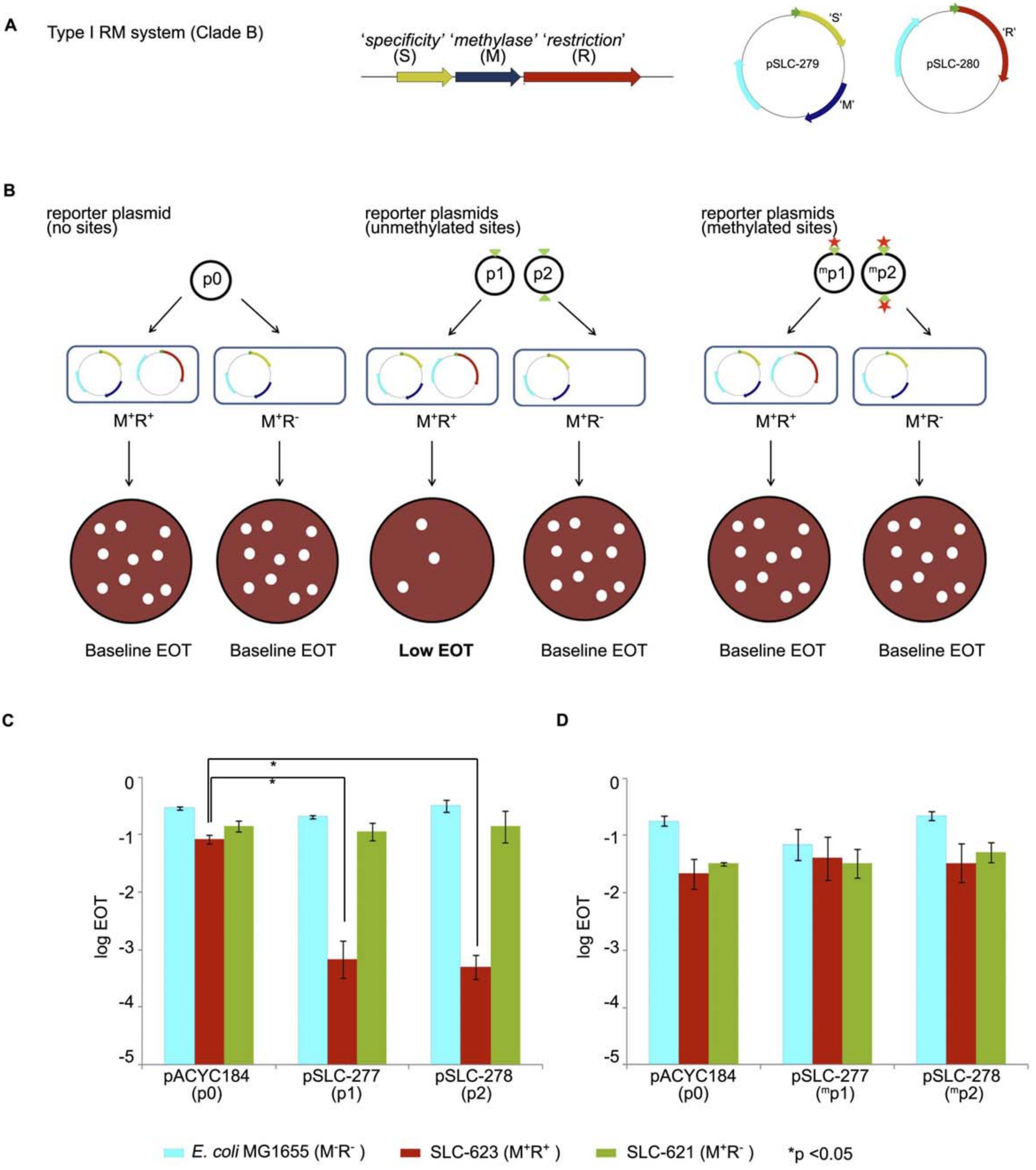
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**(5): 821-829.


A
ATCC15682
109/96
23/96
ATCC23343
P157/06
147
153
DB
AH4
35
4090-0390
EM1157
15/96
164/04
457/96
27/96
10/96
17/96
78/96
561
P166/04
169/04
P163/04
413
DI9033-508
77/96
79/96
G-Shephard
SOIL
488
JAM
AH3
612
BP59
4086-21
P165/04
BP115
DM15342/00
4009-19
EY10
EY9
DC
490
51
BP4
K11277244-293
4094-225
EB5661
D107310-3154
BP56
KAS
EM10266/01
DR13450/01
DB19897/02
422
EY7-ST414
EY8-ST169
EY1-nST4
Clade A
BP17
BP33
4033-10
BP54
I64043-3096
BP48
AM1
4044-138
P73230-3018
DR08726/01
4094-0081
39369
10-40
SW1
SW9
15/40
423
BP8
D68346-3058
BP20
P171/04
EM2107
EB6103/04
BP11
EY2
EY5
84
DB61091/00
BP107
BP10
289
170/04-nST5
Clade B
DB30729/00-nST1
P6-ST54
E387-ST163
EY3
504/96
9/96
EY6
2-96
21/96
497/96
EY4
35/96
BP6-ST50
46
2746B-ST48
BRI-ST158
E8-nST3
Clade C
HA-nST6
E25-nST2
TRF661-ST300
708a-ST23
BpK96243
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
B
ST51a
ST51b
EY10
EY9
DC
490
51a
51b
Singapore
Malaysia
A
T
G
C
0
2
4
6
Genomic location (Mb)
C
Proportion of mutations
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Syn
Nsyn
Intergenic
64
G→A
C→T
30
A→G
T→C
11
C→G
G→C
2
A→T
T→A
6
A→C
T→G
7
G→T
C→A
Transition
Transversion
D
Fraction of CpGs
(per CpNs)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Simulated
genome
G→T
C→A
G→A
C→T
C→G
G→C


A
4Mb
3Mb
Bp Chr I
1Mb
2Mb
3Mb
*TFP8
Bp Chr II
*TTSS3
1Mb
2Mb
C
Bp K96243 (WT)
Bp K96243 (TFP8 deleted)
Mice alive
Days
0 10 20 30 40 50
1 2 3 4 5
1.9x10^1
1.9x10^2
1.9x10^3
1.9x10^4
2.8x10^1
2.8x10^2
2.8x10^3
2.8x10^4
p=0.026
B
i)
Genome Coordinates
BPSS1520
2080000
2090000
BPSS1546
2100000
Type III secretion system locus
Recombination events
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
ii)
Genome Coordinates
BPSS2177
2945000
2950000
2955000
2960000
BPSS2196
Type IV flagellar pili locus
Recombination events
R1
R2
R3
R4
R5
R6
R7
R8
D
TFP8 Genome Coordinates
2943647
2946062
2950613
2954287
2959086
2962760
Sequence Types
51
422
423
84
46
R7
R2
R8
R4
Haplotypes
H1
H2
H3
H4
H5
A
T
G
C


A
Number of Novel Genes
3000
2500
2000
1500
1000
500
0
1
5
10
15
20
Number of Sequence Types
B
Translation, ribosomal structure and biogenesis
Transcription
Signal transduction mechanisms
Secondary metabolites biosynthesis, transport and catabolism
Posttranslational modification, protein turnover, chaperones
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking, secretion, and vesicular transport
Inorganic ion transport and metabolism
Energy production and conversion
Defense mechanisms
*
Coenzyme transport and metabolism
Cell wall/membrane/envelope biogenesis
Cell motility
Cell cycle control, cell division, chromosome partitioning
Carbohydrate transport and metabolism
Amino acid transport and metabolism
0
5
10
15
Percent Fraction of Genes
Core Genes present in Bp K96243
Novel Genes
*p <0.0005
C
Sequence Types
51
422
423
84
46
Sequence Types
51
422
423
84
46
RM Systems
Type IC
Type IBC
Type IIB
Type IIP
Type IIG
Type IIGC
Type III
Type IV
Fraction of Shared Accessory Elements
0
1


A
Single strain in the bacterial population aquires a DNA segment
Foreign DNA
Ongoing Recombination Through the Population
Entry of Recombined Fragment into Clade as Founder Strain
Vertical Descent
SNP Variation
Newly Acquired SNP
B
-log[SNPs per kb] Non Recombined Regions
0
2
4
6
8
-log[SNPs per kb] Recombined Regions
*
Decreasing Sequence Diversity
C
Chromosome I
Chromosome II
Genomic Location
0
600,000
1,200,000
1,800,000
2,400,000
3,000,000
3,600,000
4,200,000
4,800,000
5,400,000
6,000,000
6,600,000
7,200,000
Recombined Regions
Genomic Islands
i)
GC Content (%)
80
70
60
50
40
NR
R
AE
ii)
Effective Codon Number
60
50
40
30
20
NR
R
AE
iii)
Sequence Complexity
0.7
0.6
0.5
0.4
0.3
0.2
NR
R
AE
D
-log[SNPs per kb] Non Accessory Elements
7
6
5
4
3
2
1
0
-log[SNPs per kb] Accessory Elements
Decreasing Sequence Diversity


A
Type I RM system (Clade B)
'specificity' 'methylase' 'restriction'
(S)
(M)
(R)
'S'
pSLC-279
'M'
'R'
pSLC-280
B
reporter plasmid (no sites)
reporter plasmids (unmethylated sites)
reporter plasmids (methylated sites)
p0
p1
p2
mp1
mp2
M+R+
M+R-
Baseline EOT
Baseline EOT
Low EOT
Baseline EOT
Baseline EOT
Baseline EOT
C
D
log EOT
0
-1
-2
-3
-4
-5
*
*
pACYC184 (p0)
pSLC-277 (p1)
pSLC-278 (p2)
pACYC184 (p0)
pSLC-277 (mp1)
pSLC-278 (mp2)
E. coli MG1655 (M-R-)
SLC-623 (M+R+)
SLC-621 (M+R-)
*p <0.05