Mobile Elements within the Burkholderiaceae

Helena M.B. Seth-Smith, Matthew T.G. Holden & Julian Parkhill

Pathogen Sequencing Unit, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, U.K.

Abstract

The genus *Burkholderia*, formerly categorised within *Pseudomonas*, contains in excess of 30 species, which occupy a wide range of ecological niches. They have been reported as soil organisms, plant root colonisers, biocontrol agents, bioremediation agents and pathogens of plants, animals and humans. *Burkholderia mallei* and *Burkholderia pseudomallei* are category B biothreat agents, causing glanders and meliodosis, respectively, and some members of the *Burkholderia cepacia* complex are considered to be opportunistic human pathogens. Many *Burkholderia* genomes, each of which comprise multiple replicons, are in the process of being sequenced. Five of these have been completed and released, covering four different species. In addition, the deep shotgun sequencing of microbes from the Sargasso Sea contained high sequence coverage of a member of the Burkholderiaceae. A wealth of data now exists, from which information on mobile DNA elements can be extracted.

Burkholderia cenocepacia strain J2315 annotation

The genome of *B. cenocepacia* strain J2315 is currently being annotated[1] (Figure 1). Particular attention has been paid to the annotation of IS elements. Several new elements have been described, and submitted to IS Finder[2] (ISBcen).

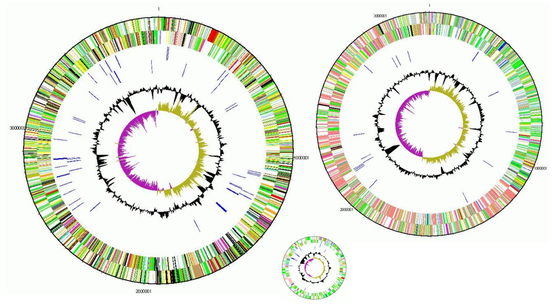

Figure 1: Circular representation of strain J2315 chromosomes 1, 2 and 3 (clockwise from left). The concentric circles on each chromosome represent the following genes, numbering from the outside in: 1,2, all genes (transcribed clockwise and anti-clockwise); 3,4, putative IS elements; 5, G+C content (plotted using a 10-kb window); 6, GC deviation ((G - C)/(G + C)) plotted using a 10-kb window[3].

Multiple insertion events

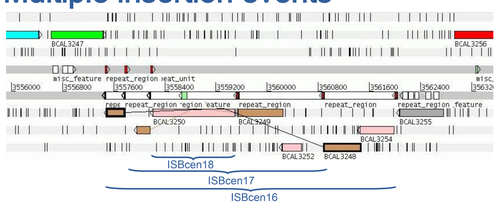

Figure 2: Artemis[4] view: In some cases, insertion sequences have inserted inside insertion sequences, creating a "graveyard" of non-functional insertion sequences.

Investigating ISs among several *Burkholderia* species

The IS elements of annotated genomes were studied, and IS elements from unpublished genomes were also investigated.

Methodology

Annotated IS elements within J2315 (Pfam, FASTA,inverted repeats (IRs), surrounding (pseudo)genes).
Made DNA database of known *Burkholderia* IS elements and searched genomes (BLASTN, 95% cut-off).
Searched all genomes (translated in 6 frames) for all transposases and related proteins, using HMMPfam and appropriate Pfam models.
Checked potential transposases against self genome to determine multiple insertions.
Defined these as genuine and determined extent of match and ends.
Made database of new elements and checked against all genomes to find any more matches.

General genome characteristics

Burkholderia pseudomallei K96243[5] (BPs) and 1710b[6] (1710b): two strains with similar IS composition
Burkholderia mallei[7] (BM): clone of *pseudomallei* which has undergone enormous IS expansion, now host restricted
Burkholderia sp. SAR1[8] (SAR1 - unfinished and unannotated): relatively few ISs: two novel
Burkholderia cepacia complex:
Burkholderia sp. strain 383[9] (383): only one multicopy IS (from criteria in methodology)
Burkholderia cenocepacia strain HI2424[10] (HI2424 - unfinished and unannotated): more than 9 novel ISs
Burkholderia cenocepacia strain J2315[1] (J2315): CF-associated, highly-transmissible epidemic strain. 10 novel ISs. More ISs than other *cepacia* complex strains.

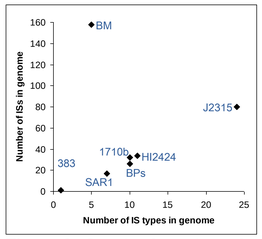

Figure 3: Graph of number of IS types per genome against number of ISs.

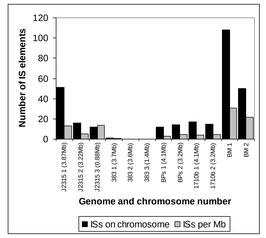

Figure 4: Graph of number of ISs per chromosome and per Mb.

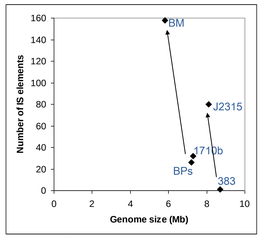

Figure 5: Graph of genome size against number of ISs.

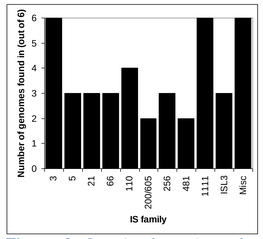

Figure 6: Graph of number of IS families found in *Burkholderia* genomes.

B. mallei has undergone enormous IS expansion (Figure 3).

In J2315, the smallest chromosome, comprising mainly accessory genes, contains more ISs per Mb than the other chromosomes (Figure 4).

A comparison between *B. pseudomallei* and *B. mallei* shows a large increase in the number of ISs. This pattern appears to be mirrored when comparing 383 with J2315 (Figure 5).

Some IS families are found spread more widely among *Burkholderia* species than others (Figure 6).

Gene inactivation

J2315 has a higher proportion (34%) of ISs disrupting coding sequences (CDSs) than any of the other strains, including *B. mallei* (28%). An example is shown in Figure 7. *B. mallei* carries more of these disruptions on on smaller chromosome 2, whereas J2315 has the higher proportions on the larger two chromosomes (Figure 8).
B. mallei is proposed to have gone through an evolutionary bottleneck, becoming host restricted. Has J2315, an epidemic strain, gone through the same process, and begun to adapt from a versatile organism to a specialised pathogen?

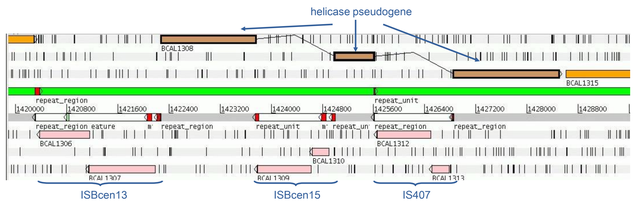

Figure 7: This helicase has been inactivated by multiple IS insertion events.

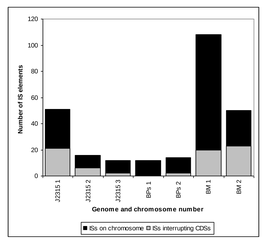

Figure 8: Graph of proportion of ISs per chromosome which interrupt CDSs.

ISs can mediate genome rearrangements

Comparing *B. pseudomallei* gene order along chromosome 1, to that of *B. mallei* chromosome 1, it is clear that the rearrangements have been mediated by recombination between IS elements (Figure 9).
Comparing gene order between *B. cepacia* complex strains J2315 and 383, there has been little rearrangement, and that which exists is not due to insertion sequences (Figure 10).

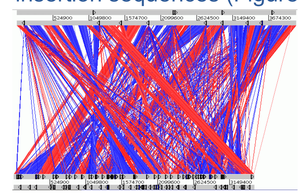

Figure 9: ACT[11]: A linear representation of BPs chromosome 1 at the top, compared with BM chromosome 1 at the bottom. The red and blue bars show matching regions. Positions of IS elements are shown as arrow heads in the forward and reverse strands.

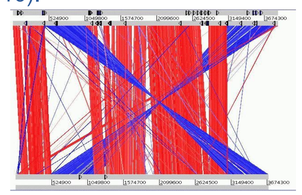

Figure 10: ACT[11]: A linear representation of strain J2315 chromosome 1 at the top, compared with strain 383 chromosome 1 at the bottom. The red and blue bars show matching regions. Positions of IS elements are shown as arrow heads in the forward and reverse strands.

Introns within the Burkholderiaceae

Group I introns

Each of the strains investigated appears to contain a group I intron. This seems to border a low GC region, possibly an island, in some strains.

Group II introns

There appear to be two types of Group II intron in the *Burkholderia* studied. In strain J2315, one of these contains ISBcen3 (Figure 11).

Figure 11: ISBcen3 inserted within a putative Group II intron.

References

1 http://www.sanger.ac.uk/Projects/B_cenocepacia/
2 http://www-is.biotoul.fr/is.html
3 Genes are colour-coded as follows: 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 and ISs; gray, miscellaneous; dark pink, unannotated.
4 **Rutherford, K. et al.** (2000) Artemis: sequence visualisation and annotation. *Bioinformatics* **16**:944-945
5 **Holden, M.T.G. et al** (2004) Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei. PNAS* **101**:14240
6 http://www.tigr.org/msc/b_pseudomallei/index.shtml
7 **Nierman, W.C. et al** (2004) Structural flexibility in the *Burkholderia mallei* genome. *PNAS* **101**:14246
8 **Venter, J.C. et al.** (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66-74
9 Joint Genome Institute, *Burkholderia* sp. 383. http://genome.jgi-psf.org/finished_microbes/bur94/bur94.home.html
10 Joint Genome Institute, *B. cenocepacia* sp. HI2424. http://genome.jgi-psf.org/draft_microbes/burce/burce.home.html
11 **Carver, T. et al.** (2005) ACT: the Artemis Comparison tool. *Bioinformatics* **21**:3422-3423. http://www.sanger.ac.uk/Software/ACT/

Acknowledgement

This project is funded by the Wellcome Trust through its core funding for the Sanger Institute