Comparative genomics for the identification of virulence factors in *Burkholderia cepacia* complex

Helena M.B. Seth-Smith[1*], Alison Dennis[1*], Matthew Holden[1], Lisa C. Crossman[1], Eshwar Mahenthiralingam[2], Julian Parkhill[1]

[1]Pathogen Sequencing Unit, The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, U.K. [2]Cardiff School of Biosciences, Cardiff University, Cardiff, Wales, U.K.

Introduction

- The *Burkholderia cepacia* complex (Bcc) of bacteria comprises ten species
- Bcc have a wide range of phenotypes: human pathogens to environmental isolates with biopesticidal capabilities
- Species *B. cenocepacia* and *B. multivorans* are significant human pathogens, particularly infecting cystic fibrosis and immuno-compromised patients
- Bcc genetics have not been well characterised; more information is required on pathogenesis and virulence factors
- *B. cenocepacia* strain J2315 has been sequenced at The Sanger Institute – annotation is ongoing
- Comparative hybridisation[1] will be used to determine differences between species without full scale sequencing

- **Strains investigated:**
  - *B. cenocepacia* strain J2315[2]
  -sequenced strain[3] (human pathogen)
  - *B. multivorans* strain C1576[4]
  -caused outbreak in CF patients
  - *B. ambifaria* strain AMMD[5] (LMG19182)
  -protects crops from fungal attack

Results

Analysis of *B. cenocepacia* sequence

The genome of strain J2315

- 8.056 Mb in three replicons of 3.870, 3.217 and 0.876 Mb
- Plasmid of 92.7 kb
- G+C content of approximately 66.9%

Comparing the genome with related organisms indicates many of the functions of chr 1 (largest replicon) as primarily housekeeping, chr 3 (smallest) as primarily accessory and chromosome 2 as both (Fig.1)

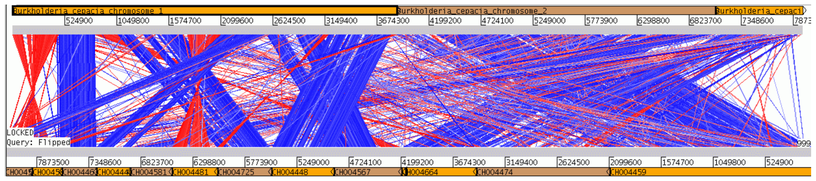

Figure 1: ACT[6] comparison of strain J2315 genome against that of *Burkholderia* sp. strain SAR-1 sampled from Sargasso Sea[7]. Top line is a linear representation of the three J2315 chromosomes (1-3, left to right). Bottom line is a linear representation of the SAR-1 contigs. Blocks of synteny between the two genomes are shown in blue and red. Chromosome 3 of J2315 displays less identity with SAR-1 than the two larger chromosomes.

Chromosome 3 (Fig. 2)

- Most variable chromosome
- Carries BcepMu phage and low G+C island with many ISs
- Many regulators
- Many membrane associated proteins
- Putative antibiotic resistance

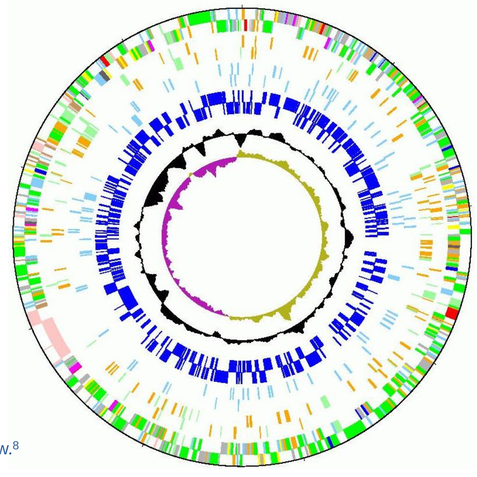

Figure 2: Circular representation of strain J2315 chromosome 3. The circles represent the following genes, numbering from the outside in: 1,2, all genes (transcribed clockwise and anti-clockwise); 3,4, hypothetical and conserved hypothetical genes; 5,6 regulators; 7,8 genes shared with *B. pseudomallei*; 9, G+C content (plotted using a 10-kb window); 10, GC deviation ((G - C)/(G + C) plotted using a 10-kb window.[8]

Comparative hybridisation between Bcc species

Aim: To discover novel/ variable genes between Bcc species

- High-density colony blots of 1kb-pUC library of unsequenced strain (*B. ambifaria* strain AMMD) (Fig. 3)
- Identify and sequence clones present only in unsequenced strain (filter 1)
- Investigate and annotate reads <70% identical to strain J2315

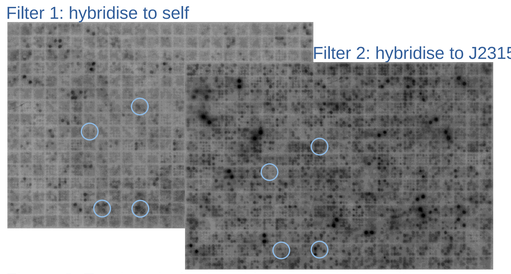

Figure 3: Two identical colony blots probed with self/ sequenced strain. Blue circles illustrate clones with differential hybridisation patterns.

Preliminary results

Regions present in AMMD, and not J2315

- Several matches to the *prn* operon from *Pseudomonas fluorescens*, for pyrrolnitrin (anti-fungal) biosynthesis (Fig. 4)
- Components of type I and III secretion systems
- Regulators
- Environmental-related genes
- Non-ribosomal peptide synthetases
- Homologues of haemagglutinins and RTX proteins
- Plasmid-related functions (Fig. 5)

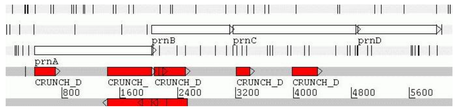

Figure 4: *prn* operon from *P. fluorescens*. The white bars indicate *prn* genes. The red bars indicate regions of homology with AMMD clones.

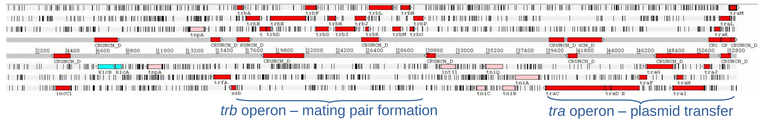

Figure 5: Plasmid pR751 – IncPβ plasmid from *Enterobacter aerogenes*[9]. The red "CRUNCH" bars on the internal grey bars indicate regions of homology with AMMD clones.

This method will be used to screen a 5x library of the unsequenced strain. Subsequent probing of BAC libraries will determine the extent of potential islands. Further strains can then be screened for islands. This approach will also be used to compare *B. multivorans* strain C1576 to strain J2315.

Summary

- *B. cenocepacia* strain J2315 has been sequenced and is being annotated
- Comparative hybridisations between related strains can provide useful information without the need for complete genome sequencing

References

1 **Liang, X., Pham, X.-Q.T., Olson, M.V. and Lory, S.** (2001) Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. *J. Bacteriol.* **183:**843-853
2 **Govan, J.R.W. et al.** (1993) Evidence for transmission of *Pseudomonas cepacia* by social contact in cystic fibrosis. *The Lancet* **342:**15-19
3 http://www.sanger.ac.uk/Projects/B_cenocepacia/
4 **Whiteford, M.L. et al.** (1995) Outcome of *Burkholderia* (*Pseudomonas*) *cepacia* colonisation in children with cystic fibrosis following a hospital outbreak. *Thorax* **50:**1194-1198
5 **Coenye, T. et al.** (2001) *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *Int. J. Syst. Evol. Microbiol.* **51:**1481-1490
6 Artemis Comparison tool (ACT). http://www.sanger.ac.uk/Software/ACT/
7 **Venter, J.C. et al.** (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:**66-74
8 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 ISEs; gray, miscellaneous.
9 **Thorsted, P.B. et al.** (1998) Complete sequence of the IncPβ plasmid R751: Implications for evolution and organisation of the IncP backbone. *J. Mol. Biol.* **282:**969-990

Acknowledgements

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