GENOME WATCH

Brothers in arms

Nicholas Thomson, Matthew Holden and Julian Parkhill

This month we discuss the insights that genomics can provide into the evolution and speciation of virulent human pathogens, using two pertinent examples from the *Yersinia* and *Burkholderia* genera.

Yersinia pestis is renowned for being the agent of the Black Death and is one of the most potent bacterial pathogens that can infect man (FIG. 1). Fully virulent strains of this pathogen have been sequenced in the past, but the publication of the whole genome sequence of an avirulent isolate, *Y. pestis* strain 91001, by Song and colleagues[1] provides an intriguing comparison. Biochemically, *Y. pestis* strain 91001 is a member of biovar Mediaevalis, in common with the previously sequenced KIM 10+ strain (KIM)[2]. Taken with *Y. pestis* strain CO92 (biovar Orientalis)[3], the genomes of three *Y. pestis* strains have now been sequenced. *Y. pestis* strain 91001 was isolated from Brandt's vole (*Microtus brandti*), which is a small rodent species found in China. This strain is lethal in mice — the LD_{50} in mice is just 23 cells — but infection of rabbits, or even human volunteers, with challenge doses in excess of 10^7 cells failed to cause plague.

Song *et al.* identify differences between the human virulent and avirulent strains and go on to discuss whether *Y. pestis* strain 91001 could indeed be a member of a new *Y. pestis* biovar. *Y. pestis* strain 91001 contains the three essential virulence plasmids pCD1, pMT1 and pPCP1, which the authors note contain sequence modifications in genes such as *yopM* and *pla* that might explain the avirulent phenotype of this strain in humans and rabbits. Whole-genome comparisons of *Y. pestis* strain 91001 with *Y. pestis* strains CO92 and KIM revealed considerable intrachromosomal rearrangements, largely mediated by insertion sequence (IS) elements, and the presence of a large number of pseudogenes (141), some of which are intact in the virulent strains.

The chromosome of *Y. pestis* strain 91001 is slightly smaller than the previous two *Y. pestis* genomes (4.6 Mb), but all three genomes share 90% of their coding sequences, with between 3–56 unique genes in pairwise comparisons. *Y. pestis* strain 91001 also has several large unique regions including DFR4, which is only found in the *Y. pestis* strains isolated from *Microtus* species and in *Yersinia pseudotuberculosis*. Moreover Song *et al.* note that even though *Y. pestis* strains KIM and 91001 belong to the same biovar, there is more sequence similarity between the KIM and CO92 strains than between KIM and 91001, which is further evidence that strain 91001 belongs to a different, perhaps earlier, lineage of *Y. pestis*. Other distinguishing features include a mutation in the *napA* gene that differs from that found in other members of biovar Mediaevalis as well as several biochemical traits, which indicate that strain 91001 represents a novel clade, which the authors name biovar Microtus.

Y. pseudotuberculosis is a pathogen that causes diarrhoea, has a broad host range and is widely distributed in the environment. As the closest relative of *Y. pestis* — ~1,500–20,000 years since the divergence of these species — the genome sequence of *Y. pseudotuberculosis* strain IP32953 (REF. 4) is a useful aid to explaining how a potent pathogen seemingly evolved from a relatively benign species (reviewed in REF. 5). Chain *et al.* present a detailed comparison of the *Y. pseudotuberculosis* and *Y. pestis* chromosomes and plasmids. The genome of *Y. pseudotuberculosis* strain IP32953 comprises a 4.7-Mb chromosome, the well-characterized *Yersinia* general virulence plasmid pYV (68 kb) and a novel cryptic plasmid pYptb32953 (27 kb).

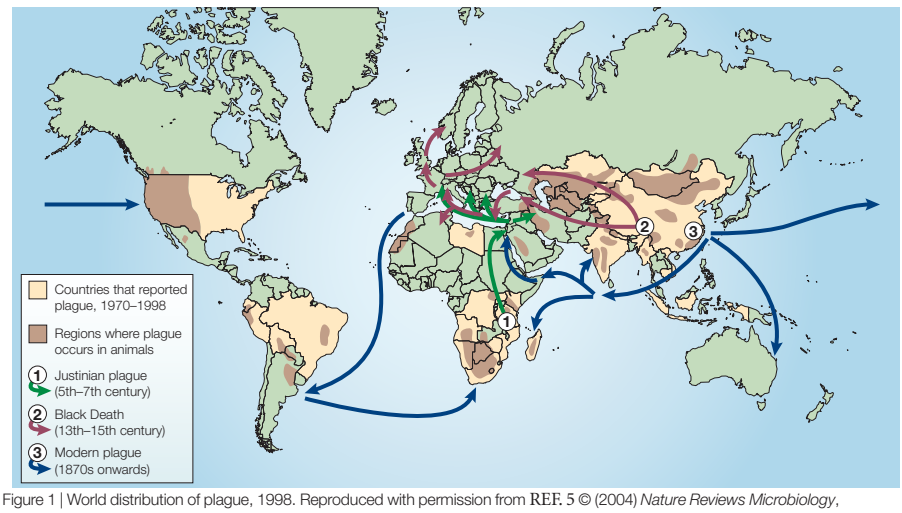

Figure 1 | World distribution of plague, 1998. Reproduced with permission from REF. 5 © (2004) *Nature Reviews Microbiology*, Macmillan Magazines Ltd.

© 2005 Nature Publishing Group

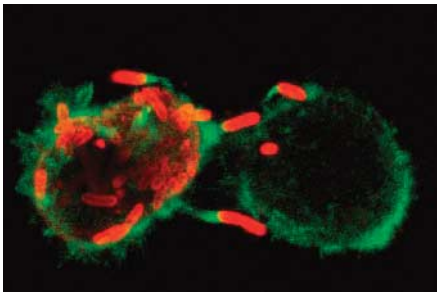

Figure 2 | Actin tail formation by *Burkholderia pseudomallei*. Macrophage filamentous actin was stained with FITC-conjugated phalloidin (green) and bacteria were stained with an antibody conjugate that detects *B. pseudomallei* LPS (red). Image courtesy of Mark P. Stevens, Institute for Animal Health, Compton, UK.

Of the predicted 3,974 genes on the chromosome, 75% have highly conserved (>97% amino acid identity) orthologues in *Y. pestis* (CO92 and KIM10). Like *Y. pestis* strains KIM, CO92 and 91001, pseudogenes are also present in the *Y. pseudotuberculosis* strain IP32953 genome — 62 have been identified, of which 43 are present in all the *Y. pestis* strains. With the benefit of a three-way comparison between *Y. pseudotuberculosis* strain IP32953 and *Y. pestis* strains KIM and CO92, Chain *et al.* identified 149 extra pseudogenes in *Y. pestis*, which, if functional, would encode surface-exposed, regulatory and virulence proteins, possibly significant in the pathogenesis of this bacterium.

Chain *et al.* used PCR to screen the *Y. pseudotuberculosis* strain IP32953-specific genes against a selection of *Y. pestis* and *Y. pseudotuberculosis* isolates and found 11 genes that are unique to *Y. pseudotuberculosis* species. This gene-set included genes that are involved in general metabolism, which might have been lost from *Y. pestis*, and glucan (osmoprotectant) biosynthetic genes. Using the same approach 32 *Y. pestis*-unique genes were found that could represent important targets for future study.

IS element expansion was one of the most striking observations made from the analysis of *Yersinia* genomes. *Y. pseudotuberculosis* contains 20 IS elements compared with 117, 138 and 109 in *Y. pestis* strains KIM, CO92 and 91001, respectively. It is evident that IS elements have driven genome fluidity in the yersiniae, which is responsible for some of the *Y. pseudotuberculosis*-unique regions, owing to IS-mediated deletion in *Y. pestis*, and the considerable amount of rearrangements that have occurred since *Y. pseudotuberculosis* and *Y. pestis* diverged, and more recently after the split of the *Y. pestis* biovars. Chain *et al.* were not able to include *Y. pestis* strain 91001 in their study, so it is with great anticipation that we await a new comparison of the non-plague and plague yersiniae genomes.

Further insight into the evolution and speciation of virulent human pathogens has been provided by the genomes of *Burkholderia pseudomallei*[6] and *Burkholderia mallei*[7] (FIG. 2). Both organisms are category B biothreat agents and cause clinically and pathologically similar illnesses. However, the natural reservoirs of the two organisms are markedly different; *B. pseudomallei* is a soil saprophyte and *B. mallei* survives only in equine hosts. Despite the differences in their ecology, the two organisms are closely related. Phylogenetic analysis using multilocus sequence typing has indicated that *B. mallei* is a clone of *B. pseudomallei*[8], thereby making these two genomes ideal for a comparative analysis to unravel the short-term evolutionary events that have shaped their niche adaptation.

B. pseudomallei is the causative agent of melioidosis, which is a complex disease that can involve many organs and tissues in the body and has high fatality rates. The disease is endemic in south-east Asia and northern Australia, and occurs after contamination of skin lesions or by inhalation after contact with water or soil. A striking feature of the disease is the incubation period, which can extend from two days to several years — the longest recorded period being 26 years in a Vietnam War veteran.

The versatility of *B. pseudomallei* is reflected in the 7.25-Mb bipartite genome, which encodes a large metabolic repertoire and many putative survival and virulence functions. The diversity and apparent redundancy of some of these functions indicates niche-specific functionality. For example, there are three complete type III secretion systems in *B. pseudomallei*. One of the systems is similar in sequence and organization to the Inv/Mxi-Spa type III systems of the animal and human pathogens *Salmonella* and *Shigella*, and the other two are similar to the type III (*hrp*) systems of the plant pathogens *Ralstonia solanacearum* and *Xanthomonas* species. The presence of three distinct systems raises interesting questions about diversity of the cell–cell interactions of which *B. pseudomallei* is capable as part of its versatile existence.

An insight into some of the short-term evolution events that have shaped the *B. pseudomallei* genome was gained by the identification of sixteen genomic islands (which comprise 6.1% of the genome) that encode a diverse array of functions. The differential distribution of these horizontally transferred regions in the wider population was confirmed by probing phylogenetically diverse clinical and environmental isolates of *B. pseudomallei* from Thailand using multiplex PCR.

Although acquisition of DNA seems to have been intrinsic to the evolution of *B. pseudomallei*, quite the opposite is true of *B. mallei*. *B. mallei* is the aetiological agent of glanders, which is a disease of horses, mules and donkeys and which can be transmitted to humans to cause a life-threatening illness. Unlike *B. pseudomallei*, it is not able to survive in the soil and is host-restricted. The bipartite genome of *B. mallei* is 1.41 Mb smaller than the *B. pseudomallei* genome. Pairwise comparisons revealed that a large proportion of the 'extra DNA' in *B. pseudomallei* is actually DNA that has been deleted from the *B. mallei* genome since it diverged from a common ancestor. Genomic islands also seem to be absent from *B. mallei*, with the exception of a single region that seems to contain the remnants of a genomic island.

In common with the *Y. pestis* genome, the *B. mallei* genome contains an increased number of IS elements, IS-mediated intrachromosomal rearrangements, and detectable pseudogenes in comparison with its more versatile relation. For both of these organisms the recent evolutionary transition from versatile to specialized pathogens has resulted in marked genomic changes that belie their relationships to their brothers in arms.

Nicholas Thomson, Matthew Holden and Julian Parkhill are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk
doi:10.1038/nrmicro1092
Published online 10th January 2005

1. Song, Y. *et al.* Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res.* **11**, 179–197 (2004).
2. Deng, W. *et al.* Genome sequence of *Yersinia pestis* KIM. *J. Bacteriol.* **184**, 4601–4611 (2002).
3. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
4. Chain, P. S. *et al.* Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* **101**, 13826–13831 (2004).
5. Wren, B. W. The yersiniae — a model genus to study the rapid evolution of bacterial pathogens. *Nature Rev. Microbiol.* **1**, 55–64 (2003).
6. Holden, M. T. G. *et al.* Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl Acad. Sci. USA* **101**, 14244–14255 (2004).
7. Nierman, W. C. *et al.* Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl Acad. Sci. USA* **101**, 14246–14251 (2004).
8. Godoy, D. *et al.* Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J. Clin. Microbiol.* **41**, 2068-2079 (2003).

Online links

DATABASES
The following terms in this article are linked online to:
Entrez: http://www.ncbi.nlm.nih.gov/Entrez/
Burkholderia mallei | *Burkholderia pseudomallei* | *Y. pestis* strain CO92 | *Y. pestis* strain 91001 | *Y. pestis* KIM 10 strain | *Y. pseudotuberculosis* strain IP32953

FURTHER INFORMATION
The Pathogen Sequencing Unit:
http://www.sanger.ac.uk/Projects/Pathogens/
Access to this links box is available online.

© 2005 Nature Publishing Group