GENOME WATCH

Cooking with GAS

Matthew Holden

This month's Genome Watch discusses another fascinating example of the latest application of second-generation sequencing techniques to the study of bacterial populations.

The landscape of microbial genomics is changing — or, to be more precise, the view on the horizon is becoming clearer. Technological advances have allowed genomics to develop into a tool that can be used to examine populations rather than just individuals. A corollary of this is that genomics can be applied to questions about bacterial evolution in unprecedented ways, including its use for investigating the dynamic populations that constitute epidemics. In many cases, epidemiological studies are limited by the absence of detailed information that can unambiguously identify epidemic isolates. As highlighted in last month's column, second-generation sequencing can be used to generate high-resolution phylogenies that reveal the diversity of bacterial populations, and it is therefore an ideal tool for probing the genetic make-up of epidemics and the genetic basis of their success and behaviour.

In a recent paper, Beres *et al.*¹ describe how they used a large-scale pathogenomic approach to investigate the evolution of successive epidemics of a single serotype of group A *Streptococcus* (GAS; also known as *Streptococcus pyogenes*) from the same geographical area. In this unparalleled study, they examined GAS isolates that caused invasive disease in three successive epidemics over a period of 16 years in Ontario, Canada. GAS is a human pathogen that is associated with a wide range of diseases, from minor

infections such as pharyngitis to more severe, life-threatening conditions such as necrotizing fasciitis and toxic shock syndrome. Classically, GAS has been typed according to the serology of the cell surface M protein, with certain M protein types being more frequently associated with disease than others. In the case of the Ontario epidemics, all of the isolates belong to the M3 serotype. This serotype causes a disproportionate number of cases of invasive disease and is associated with a higher rate of lethal infection than other serotypes.

Using a combination of approaches, Beres *et al.* generated a high-resolution genetic structure for 344 GAS isolates from the 3 epidemics. Second-generation sequencing techniques were used to sequence the entire genomes of 87 isolates, thus providing a snapshot of the genomic diversity involved. This revealed that no 2 isolates were identical, with each isolate differing from a reference strain by 49 single-nucleotide polymorphisms (SNPs) and 11 indels (insertions and deletions) in the core genome, on average. Beres and colleagues were able to extend the scale of the genotyping to encompass additional epidemic isolates using a method based on mass spectroscopy to assay 280 of the SNP loci that had been identified by sequencing. Phylogenetic analysis of the results showed that each epidemic was dominated by micro- and macro-bursts of multiple emerging clones and each involved isolates that were genetically distinct from isolates in the preceding epidemic, indicating that none represented a re-emergence. The authors also looked at the phylogeography of the isolates and found that there was a correlation across large numbers of strains between genetic relatedness and spatial proximity, suggesting the existence of some underlying geographical structuring.

The heterogeneity of these epidemics also provided insights into the impact of micro-evolutionary changes on GAS pathogenesis. Analysis of the associations between strain genotypes and patient phenotypes revealed

non-random distributions, suggesting that there were differences in the potential of the isolates to cause invasive disease. Notably, related isolates containing modest genetic differences showed marked differences in their transcriptomes. It had been shown previously that subtle genetic differences are important for the pathogenic potentials of the isolates from these epidemics^{2,3}. In the latest study, the authors identified 22 genes that had a statistically significant over-abundance of SNPs, including several genes that encode virulence determinants and regulators. In particular, an elevated level of non-synonymous substitution was observed in *ropB*, which encodes a regulator of the cysteine protease virulence factor, streptococcal pyrogenic exotoxin B (SpeB). Examination of this locus in invasive isolates identified further non-synonymous mutations, which the authors suggest is evidence of positive selection driven by interaction with the host.

This study provides an elegant illustration of the potential applications of genomics, and second-generation techniques in particular, in analysing the biology and genetics of bacterial epidemics, and it will undoubtedly fuel future studies.

Matthew Holden is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

e-mail: microbes@sanger.ac.uk

doi10.1038/nrmicro2342

1. Beres, S. B. *et al.* Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc. Natl Acad. Sci. USA* **107**, 4371–4376 (2010).
2. Beres, S. B. *et al.* Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. *Proc. Natl Acad. Sci. USA* **103**, 7059–7064 (2006).
3. Olsen, R. J. *et al.* Decreased necrotizing fasciitis capacity caused by a single nucleotide mutation that alters a multiple gene virulence axis. *Proc. Natl Acad. Sci. USA* **107**, 888–893 (2010).

Competing interests statement

The author declares no competing financial interests.