GENOME WATCH

Stronger together

John A. Lees, Gerry Tonkin-Hill and Stephen D. Bentley

This month's Genome Watch discusses how genome-wide epistasis analysis and genome-to-genome analyses are providing insights into the relevance of interactions between and within the genomes of pathogens and their hosts.

Susceptibility to infectious diseases and disease severity may be affected by interactions between the genomes of the host and pathogen. Although the effect of variations in gene sequence on disease has been studied for several pathogens, interactions between the genomes of pathogens and their hosts have been less appreciated, but are now under active investigation.

In a pioneering study, Bartha *et al.* used genotype data from 1,071 individuals infected with HIV-1 and the genome sequence of their respective infecting strain to carry out a genome-wide association study (GWAS) that correlated host DNA polymorphisms and HIV-1 sequence variation with plasma viral load¹. In addition to the GWAS study, the authors then carried out a global genome-to-genome analysis, looking for host SNPs that correlated with amino acids changes in the HIV-1 sequence. They found that SNPs in the major histocompatibility complex were associated with amino acid variants in the Gag and Nef viral proteins,

which suggests that HIV-1 infection exerts selection on these cytotoxic T cell epitopes.

A second GWAS compared paired host and viral genome sequences from 542 individuals who were chronically infected with hepatitis C virus (HCV)². The genome-to-genome analysis identified a correlation between host variants in interferon- λ 4 (*IFNL4*) loci and a specific amino acid polymorphism in the viral NS5A protein, and found that it was associated with a reduced viral load. The authors hypothesized that in individuals in whom *IFNL4* is not expressed, innate immunity against HCV is reduced, whereas in patients who express *IFNL4*, an antiviral innate immune response to the virus is maintained, which leads to lower viral loads.

The availability of large collections of bacterial population genomic datasets has also facilitated the analysis of interactions between genetic loci in the same genome. Interactions between loci in the same genome that have a fitness effect are known as epistatic interactions. Preliminary data from theoretical modelling studies suggest that such epistatic interactions may drive adaptation in bacteria that undergo recombination³; however, only relatively few examples of epistatic interactions in bacteria are known.

The potential for many different combinations of interactions between genetic loci makes analysing such interactions a challenge. The combinatorial nature of the problem results in a large computational burden and requires suitable statistical tests to cope with multiple comparisons. However, recent advances in statistical methodologies and larger sample populations have helped to increase the statistical power of analyses to overcome these issues.

Direct coupling analysis (DCA) can be used to computationally assess the strength of interactions between two loci

from a multiple sequence alignment (MSA). Previously, DCA was used to identify correlations between amino acids in protein alignments that represented real contacts in the folded structure. Skwark *et al.* applied DCA to whole-genome alignments of two populations of *Streptococcus pneumoniae* and *Streptococcus pyogenes* to identify co-evolving loci⁴. To reduce the high computational burden, they sampled random pairs of loci from distant locations across the genome. Using DCA to identify significant associations between loci in a population of *S. pneumoniae*, the authors found that epistatic interactions between three penicillin-binding proteins were the strongest, confirming previous experimental results.

DCA is a promising approach that can be used in combination with traditional GWAS analyses to identify epistatic interactions. As larger datasets from host and pathogen cohorts are collected, our ability to identify interactions within genomes and between genomes will increase. The analysis and discovery of these genomic interactions may provide important insights into antimicrobial resistance, as well as into host and pathogen evolution and biology.

John A. Lees, Gerry Tonkin-Hill and Stephen D. Bentley are at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro.2017.95

1. Bartha, I. *et al.* A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* **2**, e01123 (2013).
2. Azim Ansari, M. *et al.* Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat. Genet.* **49**, 666–673 (2017).
3. Arnold, B. J. *et al.* Weak epistasis may drive adaptation in recombining bacteria. Preprint at *bioRxiv* <http://dx.doi.org/10.1101/119958> (2017).
4. Skwark, M. J. *et al.* Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis. *PLoS Genet.* **13**, e1006508 (2017).

Competing interests statement

The author declares no competing interests.

Philip Patenali/Macmillan Publishers Limited